Selective Data Acquisition in Learning and Decision Making Problems

Yining Wang

Machine Learning Department
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Aarti Singh, Chair
Sivaraman Balakrishnan
Larry Wasserman
Robert Nowak (University of Wisconsin at Madison)

*Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.*

Abstract

Classical statistics and machine learning posit that data are passively collected, usually assumed to be independently and identically distributed. In modern data science applications, however, many times a data analyst has control over how data are acquired or selected. For example, in simulation/hyper-parameter optimization the input parameter configurations can be adaptively chosen to obtain data resulting from the carefully chosen input parameters. In sequential decision making problems, data such as feedback or utility depend on the particular decisions which can be adaptively and selectively made.

The main topic of this thesis is to study how *selective data acquisition* techniques can be applied in estimation, optimization and/or decision making problems. Three representative problems are studied, as we explain in more details below:

1. **Computationally tractable experimental design**, which studies the classical question of (optimal) experimental design in linear and generalized linear models from a *computational* perspective. We design polynomial-time algorithms with rigorous approximation guarantees in terms of optimality criteria, and show an application to a 3D lightweight structure optimization problem.

2. **Sample-efficient query regimes for nonparametric optimization**, which tries to understand the most sample efficient regimes to make adaptive queries to a nonparametric function for optimization purposes. We consider three different settings of nonparametric optimization: smooth non-convex functions in low dimensions, high-dimensional convex functions with sparsity structures, and convex function sequences that evolve slowly over time.

3. **Dynamic assortment optimization**, which studies the classical assortment optimization problem in revenue management from a dynamic perspective, by combining statistical estimation of customers' utility models and optimization of assortments based on estimated utilities into a unified theoretical framework.

We characterize through statistical minimaxity the fundamental information-theoretic limits of these problems as well as notions of optimality of our proposed methodologies. On the practical side, we demonstrate industrial engineering and/or operations management applications such as lightweight structural design, dynamic pricing and assortment planning.

Contents

List of Figures

List of Tables

List of Algorithms

Chapter 1

Introduction

Classical statistics and machine learning posit that data are passively collected, usually assumed to be independently and identically distributed. In modern data science applications, however, many times a data analyst has control over how data are acquired or selected. One particular interesting data collection scheme is *selective/active* data acquisition, in which a data analyst is capable of selecting which data should be collected prior to experiments/measurements, or updating his/her data collection as the experiments and data analysis are undergoing.

In this thesis, we concentrate on theoretical and practical aspects of selective/active data acquisition for a wide range of estimation and optimization problems. The problems we considered and results we obtained are summarized in the rest of this chapter.

1.1 Computationally tractable experimental design methods

Given a large pool of candidate design points (unlabeled data points), the problem of *experimental design* is to select, prior to any actual experimental/labeling procedures, a small subset of design points on which measurements or labels are to be collected, in order to maximize statistical efficiency and minimize measurement/labeling efforts. More specifically, given a pool of data/design points $\mathcal{X}$, the objective is to select aq subset $X \subseteq \mathcal{X}$ under resource constraints such as $|X| \leqslant k$ that is optimal for a given learning objective.

The experimental design problem has two aspects: the statistical question focuses on which candidate set has the maximal statistical efficiency, and the computational question studies how to find a good candidate set in a computationally tractable manner. While the statistical question has been mostly well understood (at least for linear models and their variants) (Fedorov, 1972; Pukelsheim, 2006), the computational aspect is less investigated, and computationally tractable (polynomial running-time) methods with rigorous approximation statistical efficiency guarantees are particularly rare.

My thesis work on the computationally tractable experimental design problem is presented in Chapter 2. The algorithmic framework is a *continuous relaxation* of the discrete combinatorial experimental design problem, which is easy to solve using conventional convex continuous optimization methods (Boyd & Vandenberghe, 2004; Yudin & Nemirovskii, 1983). After the continuous relaxation, sampling based or greedy techniques are applied to "round/sparsify" con-

tinuous solutions into discrete subsets of selected candidate design points.

Rigorous approximation guarantees are established for the proposed computationally tractable algorithms. More specifically, we show for linear regression models that if k (the number of selected candidate points) is at least $\Omega(p/\varepsilon^2)$ where p is the problem dimension (i.e., number of variables), then the proposed algorithm achieves a $(1 + \varepsilon)$ relative approximation of the optimal statistical efficiency. We also demonstrate our methods on a real-world application of 3D lightweight structural design as an application of our proposed methods.

1.2 Selective queries in nonparametric optimization

Many practical questions can be cast as *optimization* (e.g., finding the minimum value) of an *unknown* function $f : \mathcal{X} \to \mathbb{R}$ on a known domain $\mathcal{X} \subseteq \mathbb{R}^d$, through *selective queries* of noisy function values $f(x_t)$ at carefully selected query points $\{x_t\}_t$. The examples of hyper-parameter estimation and some experiment/simulation optimization problems fall into this framework, by abstracting the mapping from hyper-parameter or simulation settings to performance of an algorithm/experimental protocols as the unknown function f to be optimized.

Chapter 3 presents my thesis work on the nonparametric optimization question, underly three different settings. Sec. 3.1 considers the case where domain dimension d is very small, and derives tight *local* minimax rates for optimizing a smooth nonparametric function with certain level set growth conditions. Sec. 3.2, on the other hand, focuses on settings where domain dimension d is very large, maybe far exceeding the number of queries/experiments n allowed. By imposing sparsity and convexity assumptions on f, the number of queries n needs only to depend on the "intrinsic" dimension s and the *logarithm* of the ambient dimension d. In Sec. 3.3, we consider a *non-stationary* version of the nonparametric optimization problem in which the function to be optimized is allowed to slightly change over time. Such non-stationary settings are useful in operations/revenue management applications such as dynamic pricing.

1.3 Dynamic assortment planning

Consider N items for sale, each associate with a known "revenue parameter" $r_i \in [0, 1]$ indicating the amount of revenue collected once a customer purchases the ith item. At each time epoch t, the retailer provides an *assortment* of items $S_t \subseteq [N]$ to an incoming customer, and observes a *purchasing action* $i_t \in S_t \cup \{0\}$, indicating which item in the assortment S_t is purchased by the customer (if $i_t \in S_t$) or no purchase is made ($i_t = 0$), in the case that none of the items in S_t is satisfactory to the customer.

Usually, the customer's purchasing choice i_t is governed by a probabilisitic model

$$i_t \ \sim \ p_{\theta_0}(\cdot|S_t),$$

where θ_0 is an underlying parameter characterizing the customer's *preferences* of items. Examples include independent preference parameters v_i for each $i \in [N]$, or contextual models $v_i = \exp\{x_i^\top \theta_0\}$. Unlike *stationary* settings where θ_0 is perfectly known and the assortment planning problem is merely a combinatorial optimization one (see, e.g., Anderson et al. (1992);

2

Kök et al. (2008)), under dynamic settings the preference parameters θ_0 are *unknown* and have to be inferred or estimated on the fly from customers' purchasing actions $\{i_t\}_t$ (Agrawal et al., 2017a; Rusmevichientong & Topaloglu, 2012; Saure & Zeevi, 2013).

Chapter 4 describes my thesis work on the dynamic assortment planning problem. Sec. 4.1 considers the plain multinomial logit choice model and drives a surprising N-independent regret bound based on a novel trisection based algorithm. Sec. 4.2 studies the more complex *nested* logit choice model, and finally in Sec. 4.3 a discrete choice model with contextual information of items is studied. For all variants of discrete choice models (and the dynamic assortment optimization problems they give rise to), rigorous regret upper bounds are proved for the policies, and regret lower bounds are proved whenever possible to show the optimality of our proposed methods.

Chapter 2

Computationally tractable experimental design methods

(Optimal) design of experiments is a classical topic in statistics research (Pukelsheim, 2006). Given a large collection of *design points* $\mathcal{X} = \{x_1, x_2, \cdots, x_n\} \in \mathbb{R}^p$, the objective of experimental design is to select a small subset $\{z_1, \cdots, z_k\} \subseteq \mathcal{X}$ with $k \ll n$ such that regression over the selected subset of design points achieves the optimal statistical efficiency. The experimental design problem is particularly important in several scientific and engineering fields, where experiments are expensive and time-consuming to carry out, and a careful experimental design strategy is mandatory.

In this chapter, we concentrate on the *computational* aspects of experimental design. More specifically, we design algorithms that are computationally tractable for very large pool of design points (large n) while still maintaining the near-optimal statistical efficiency of the selected design subset $\{z_1, \cdots, z_k\}$. Apart from rigorous approximation guarantees theoretically, we also consider a real-world application of 3D lightweight structure design (Ulu et al., 2017) and show significant improvement over existing methodologies. Finally, we study several extension of our proposed methodologies, including the application to quantized linear regression, transfer learning and generalized linear models.

2.1 Backgrounds and optimality criteria

Consider a linear model

$$y = Z\beta_0 + \xi, \tag{2.1}$$

where $Z = (z_1, \cdots, z_k) \in \mathbb{R}^{k \times p}$ is the stacked design matrix consisting of selected design points $\{z_1, \cdots, z_k\}$, $\beta_0 \in \mathbb{R}^p$ is an unknown p-dimensional regression model to be estimated, and $\xi \sim \mathcal{N}_k(0, \sigma^2 I)$ is a centered Gaussian noise random vector.

The standard estimator of β_0 is the *ordinary least squares (OLS)* estimator, of the form $\widehat{\beta} = (Z^\top Z)^{-1} Z^\top y$. By simple algebra it is easy to check that the estimation error $\widehat{\beta} - \beta_0$ is a centered

Gaussian random vector:

$$\widehat{\beta} - \beta_0 \ \sim \ \mathcal{N}_p(0, \sigma^2 \Sigma) \quad \text{where} \quad \Sigma := (Z^\top Z)^{-1} = \left(\sum_{i=1}^{k} z_i z_i^\top \right)^{-1}. \tag{2.2}$$

If the *variance* of $\widehat{\beta} - \beta$ is to be minimized (i.e., minimizing $\mathbb{E}\|\widehat{\beta} - \beta\|_2^2$), one should seek design subset Z with the smallest $\mathrm{tr}(\Sigma^{-1})$. Another popular objective is to maximize the *determinant* of Σ, which has the advantage of being invariant with respect to unit of measurements. We use the following definition of *optimality criteria* to abstract all criteria that reflect certain aspects of desirable statistical efficiency:

Definition 1 (Optimality criteria). *An* optimality criterion *is a function $f : \mathbb{S}_p^+ \to \mathbb{R}^+$ that maps a p-dimensional positive definite matrix Σ to a positive real number $f(\Sigma)$. A smaller $f(\Sigma)$ indicates better statistical efficiency of the corresponding selected design set Z.*

Below we list several popular choices of the objective functions f:

- A-optimality (**Average**): $f_A(\Sigma) = \frac{1}{p}\mathrm{tr}(\Sigma^{-1})$;

- D-optimality (**Determinant**): $f_D(\Sigma) = (\det |\Sigma|)^{-1/p}$;

- T-optimality (**Trace**): $f_T(\Sigma) = p/\mathrm{tr}(\Sigma)$;

- E-optimality (**Eigenvalue**): $f_E(\Sigma) = \|\Sigma^{-1}\|_{\mathrm{op}} = \lambda_{\max}(\Sigma^{-1})$;

- V-optimality (**Variance**): $f_V(\Sigma) = \frac{1}{n} \sum_{i=1}^{n} x_i^\top \Sigma^{-1} x_i$;

- G-optimality: $f_G(\Sigma) = \max_i x_i^\top \Sigma^{-1} x_i$.

We refer the readers to (Pukelsheim, 2006) for a complete list and discussion of various optimality criteria used in the experimental design literature.

With the formal definition of an optimality criterion f, the experimental design problem can then be formulated as a *combinatorial optimization* problem

$$\min_s F(s) = \min_s f\left(\sum_{i=1}^{n} s_i x_i x_i^\top \right) \quad s.t. \ \ s_i \in \mathbb{N}, \ \ 0 \leqslant s_i \leqslant b, \ \ \sum_{i=1}^{n} s_i \leqslant k. \tag{2.3}$$

Here, when $b = 1$ we operate under the "without replacement" setting in which each design point x_i in pool $\mathcal{X}$ can be selected *at most once*. On the other hand, $b = k$ is referred to as the "with replacement" setting in which there is no limits on the number of times each point is selected. It is a simple exercise that the $b = k$ problem can be reduced to the $b = 1$ problem by replicating each x_i for k times and is therefore easier. Very often, an algorithm that solves the $b = 1$ instance can be easily altered to handle the $b = k$ case without explicit replication of design points or an increase of running time.

Instead of considering each optimality criterion separately, in our work we adopt a unified approach that applies to a wide range of optimality criteria satisfying minimal regularity conditions. More specifically, we define "regular" criteria as follows:

Definition 2 (Regular criteria). *An optimality criterion $f : \mathbb{S}_p^+ \to \mathbb{R}$ is* regular *if it satisfies the following properties:*

1. Convexity: [1] $f(\lambda A + (1-\lambda)B) \leqslant \lambda f(A) + (1-\lambda)f(B)$ *for all $\lambda \in [0,1]$ and $A, B \in \mathbb{S}_p^+$;*

---

[1]This property could be relaxed to allow a proxy function $g : \mathbb{S}_p^+ \to \mathbb{R}$ being convex, where $g(A) \leqslant g(B) \Leftrightarrow f(A) \leqslant f(B)$.

Algorithm 1 The projected entropic mirror descent algorithm for solving Eq. (2.5).

Require: function $\min_\omega \widetilde{F}(\omega)$ defined in Eq (2.5); T number of iterations, step size rules $\{\eta_t\}$
  1: $\omega^{(0)} = (1/n, \cdots, 1/n)$; $\triangleright$ initialization
  2: **for** $t \leftarrow 0$ **to** $T-1$ **do**
  3:     Compute subgradient $g^{(t)} \in \partial \widetilde{F}(\omega^{(t)})$;
  4:     Update: $\omega_i^{(t+1/2)} \propto \omega_i^{(t)} \exp\{-\eta_t g_i^{(t)}\}$, normalized so that $\sum_{i=1}^n \omega_i^{(t+1/2)} = 1$;
  5:     Projection: $\omega^{(t+1)} \leftarrow \text{BOXSIMPLEXPROJECT}(\omega^{(t+1/2)}, b/k)$;
  6: **end for**
  7: **return** $\widehat{\omega} := \frac{1}{T} \sum_{t=0}^{T-1} \omega^{(t)}$.
---

    2. *Monotonicity: If $A \preceq B$ then $f(A) \geqslant f(B)$;*
    3. *Reciprocal multiplicity: $f(tA) = t^{-1} f(A)$ for all $t > 0$ and $A \in \mathbb{S}_p^+$.*

    It can be verified that all A, D, T, E, V and G-optimality are regular. Note that for D-optimality the proxy function $g_D(\Sigma) = -\log \det(\Sigma)$ is considered to satisfy the convexity property.

2.2 The continuous relaxation framework

A straightforward solution to the combinatorial optimization problem mentioned in Eq. (2.3) is to enumerate in brute force all $\binom{n}{k}$ possible solutions of $s \in \{0,1\}^k$. Such an approach would be however too computationally expensive and inadequate for even moderately sized design pools.

    An alternative approach, which is adopted as the main algorithmic framework in this chapter, is to consider a *continuous relaxation* to the discrete optimization problem in Eq. (2.3):

$$\pi^* \in \arg\min_\pi F(\pi) = \arg \min_{\pi_1, \cdots, \pi_n} \ f\left(\sum_{i=1}^n \pi_i x_i x_i^\top\right) \quad s.t. \ 0 \leqslant \pi_i \leqslant b, \ \sum_{i=1}^n \pi_i \leqslant k. \quad (2.4)$$

    It should be noted that any feasible solution to the original discrete optimization problem (2.3) is also feasible for the continuously relaxed program in Eq. (2.4), and hence $F(\pi^*) \leqslant F(s^*)$ always holds. In addition, thanks to the first property in Definition 2, any regular optimality criterion f leads to a continuous program in Eq. (2.4) with a *convex* objective and a *convex and compact* feasibility region. In the rest of this section we present two efficient approaches to optimize Eq. (2.4): the first approach formulates Eq. (2.4) as a *semi-definite programming* (SDP), which can be provably solved in polynomial time (Vandenberghe & Boyd, 1996); the second approach uses *entropic mirror descent* (Beck & Teboulle, 2003) to solve Eq. (2.4), which is practically efficient.

2.2.1 The semidefinite programming (SDP) approach

For $\pi \in \mathbb{R}^n$ define $A(\pi) = \sum_{i=1}^n \pi_i x_i x_i^\top$, which is a $p \times p$ positive semidefinite matrix. By definition, $f(\pi; X) = \sum_{j=1}^p e_j^\top A(\pi)^{-1} e_j$, where e_j is the p-dimensional vector with only pth

coordinate being 1. Subsequently, Eq. (2.4) is equivalent to the following SDP problem:

$$\min_{\pi, t} \sum_{j=1}^{p} t_j \quad \text{subject to} \quad 0 \leqslant \pi_i \leqslant b, \ \sum_{i=1}^{n} \pi_i \leqslant k, \ \text{diag}(B_1, \cdots, B_p) \succeq 0,$$

where

$$B_j = \begin{bmatrix} A(\pi) & e_j \\ e_j^\top & t_j \end{bmatrix}, \qquad j = 1, \cdots, p.$$

Global optimal solution of an SDP can be computed in polynomial time (Vandenberghe & Boyd, 1996). However, this formulation is not intended for practical computation because of the large number of variables in the SDP system. First-order methods such as projected gradient descent is a more appropriate choice for practical computation.

2.2.2 The entropic mirror descent approach

We first note that Eq. (2.4) can be re-formulated as

$$\min_{\omega} \widetilde{F}(\omega) := \min_{\omega} f\left(\sum_{i=1}^{n} \omega_i x_i x_i^\top\right) \quad s.t. \ 0 \leqslant \omega_i \leqslant b/k, \ \sum_{i=1}^{k} \omega_i = 1, \tag{2.5}$$

by the change of variables $\omega_i = \pi_i/k$ and noting that the $\sum_{i=1}^{n} \pi_i \leqslant k$ constraint in Eq. (2.4) must bind, meaning that the optimal solution π^* must satisfy $\sum_{i=1}^{n} \pi_i^* = k$.

The *entropic mirror descent* (Beck & Teboulle, 2003) is a classical algorithm that takes into account the geometry of high-dimensional probabilistic simplex to efficiently solve constrained convex optimization problems. At a high level, entropic mirror descent uses the Kullbeck-Leibler (KL) divergence $\sum_i x_i \log(x_i/y_i)$ as the Bregman divergence, whose proximal operator can be evaluated in closed form as multiplicative weight updates.

We describe in Algorithm 1 how (projected) entropic mirror descent is applied to solve program 2.5. As our problem has an extra box constraint $\omega_i \leqslant 1/k$, we present in Algorithm 2 a simple algorithm that computes such projection in $O(n \log n)$ time and the KL divergence. The projection algorithm is (in principle similar to but) much simpler than existing algorithms that compute projections onto simplex or L_1 balls (Condat, 2015; Duchi et al., 2008).

2.3 Rounding techniques

In the previous section we described a continuous relaxation of the original discrete optimization problem, and briefly explained how the continuous relaxation can be solved efficiently to obtain a (near) optimal *continuous* solution $\pi^* \in [0, 1]^n$. In this section we describe various strategies of "rounding" the continuous solution π^* to a discrete one s, and prove rigorous approximation guarantees for these rounding techniques.

Algorithm 2 Projection onto the probabilistic simplex with box constraint

---

Require: $\omega \in \Delta_n$, parameter $b \in [1/n, 1]$. $\rhd \Delta_n = \{x \in \mathbb{R}^n : x_i \geqslant 0, \sum_i x_i = 1\}$
Ensure: an output $\omega' \in \Delta_n$ such that $\|\omega'\|_\infty \leqslant b$. $\rhd \omega' = \arg\min_{y \in \Delta_n, y_i \leqslant b} \mathrm{KL}(y\|\omega)$
                                              $\rhd$ where $\mathrm{KL}(y\|\omega) := \sum_i y_i \log \frac{y_i}{\omega_i}$

1: Sort ω in descending order: $\omega_1 \geqslant \omega_2 \geqslant \cdots \geqslant \omega_n > 0$;
2: **if** $\omega_1 \leqslant b$ **then return** ω;
3: $\mathrm{KL}_1 \leftarrow b\log(b/\omega_1)$, $Z_1 \leftarrow 1 - \omega_1$, $\mathrm{KL}_{\mathrm{opt}} \leftarrow \infty$, $\eta_{\mathrm{opt}} \leftarrow 0$,   and $C_{\mathrm{opt}} \leftarrow 0$;
4: **for** $q \leftarrow 2$ to n **do**
5:      $C \leftarrow (1 - b(q-1))/Z_{q-1}$;
6:      **if** $C > 0$ and $Cw_q \leqslant b$ and $\big(\mathrm{KL}_{q-1} + C\log(C) \cdot Z_{q-1} \leqslant \mathrm{KL}_{\mathrm{opt}}\big)$ **then**
7:          $\mathrm{KL}_{\mathrm{opt}} \leftarrow \mathrm{KL}_{q-1} + C\log(C) \cdot Z_{q-1}$, $\eta_{\mathrm{opt}} \leftarrow \omega_q$, $C_{\mathrm{opt}} \leftarrow C$;
8:      **end if**
9:      $\mathrm{KL}_q \leftarrow b\log(b/\omega_q)$, $Z_q \leftarrow Z_{q-1} - \omega_q$;
10: **end for**
11: Set $\omega'_i \leftarrow b$ if $\omega_i \geqslant \eta_{\mathrm{opt}}$, and $\omega'_i \leftarrow C_{\mathrm{opt}}\omega_i$ if $\omega_i < \eta_{\mathrm{opt}}$;
12: **return** ω'.

Algorithm 3 Sampling based experiment selection.

    Input: $X \in \mathbb{R}^{n \times p}$, optimal solution π^*, target subset size k.
    Output: $\widehat{S} \subseteq [n]$, a selected subset of size at most k.
    Initialization: $t = 0$, $S_0 = \varnothing$, $R_0 = \varnothing$.
    1. *With replacement*: sample $i_t \sim P^{(1)}$; set $w_t = \lceil \pi^*_{i_t}/(kp^{(1)}_{i_t}) \rceil$;
       Without replacement: pick random $i_t \notin R_{t-1}$; sample $w_t \sim \mathrm{Bernoulli}(kp^{(2)}_{i_t})$.
    2. *Update*: $S_t = S_{t-1} \cup \{w_t \text{ repetitions of } i_t\}$, $R_t = R_{t-1} \cup \{i_t\}$.
    3. Repeat steps 1 and 2 until at some $t = T$, $|S_{T+1}| > k$ or $R_{T+1} = [n]$. Output $\widehat{S} = S_T$.

2.3.1 Sampling based techniques

Perhaps the simplest idea of rounding π^* into a discrete solution is to *sample* according to an appropriately normalized categorical distribution based on π^*. More specifically, define $\Sigma_* := X^\top \mathrm{diag}(\pi^*)X = \sum_{i=1}^n \pi^*_i x_i x_i^\top$ and

$$P^{(1)}: \quad p^{(1)}_j = \pi^*_j x_j^\top \Sigma_*^{-1} x_j/p, \qquad\qquad \text{with replacement } (b = k);$$
$$P^{(2)}: \quad p^{(2)}_j = \pi^*_j/k, \qquad\qquad\qquad\quad \text{without replacement } (b = 1).$$

Note that both $\{p^{(1)}_j\}_{j=1}^n$ and $\{p^{(2)}_j\}_{j=1}^n$ sum to one because $\sum_{j=1}^n \pi^*_j = k$ and $\sum_{j=1}^n \pi^*_j x_j^\top \Sigma_*^{-1} x_j = \mathrm{tr}((\sum_{j=1}^n \pi^*_j x_j x_j^\top)\Sigma_*^{-1}) = \mathrm{tr}(\Sigma_* \Sigma_*^{-1}) = p$.

     Pseudo-codes of the sampling procedure are given in Algorithm 3. The sampling procedure is easy to understand in an asymptotic sense: it is easy to verify that $\mathbb{E}X_{i_t}^\top X_{i_t} = X^\top \mathrm{diag}(\pi^*/k)X$ and $\mathbb{E}[w_{i_t}] = 1$, for both with and without replacement settings. Note that $\|p^{(2)}_{i_t}\|_\infty \leqslant 1/k$ by feasibility constraints and hence $\mathrm{Bernoulli}(kp^{(2)}_{i_t})$ is a valid distribution for all $i_t \in [n]$. For the with replacement setting, by weak law of large numbers, $X_{\widehat{S}}^\top X_{\widehat{S}} \xrightarrow{p} X^\top \mathrm{diag}(\pi^*)X$ as $k \to \infty$

Algorithm 4 Greedy experiment selection.

    Input: $X \in \mathbb{R}^{n \times p}$, Initial subset $S_0 \subseteq [n]$, target size $k \leqslant |S_0|$.
    Output: $\widehat{S} \subseteq [n]$, a selected subset of size k.
    Initialization: $t = 0$.
    1. Find $j^* \in S_t$ such that $\mathrm{tr}[(X_{S_t \setminus \{j^*\}}^\top X_{S_t \setminus \{j^*\}})^{-1}]$ is minimized.
    2. Remove j^* from S_t: $S_{t+1} = S_t \setminus \{j^*\}$.
    3. Repeat steps 1 and 2 until $|S_t| = k$. Output $\widehat{S} = S_t$.
---

and hence $\mathrm{tr}[(X_S^\top X_S)^{-1}] \xrightarrow{p} f(\pi^*; X)$ by continuous mapping theorem.

The following theorem provides a rigorous, finite-sample statement of the above intuition:

Theorem 1. *Suppose the following conditions hold:*

$$\textit{with replacement}: \quad p \log k / k = O(1);$$
$$\textit{without replacement}: \quad \|\Sigma_*^{-1}\|_2 \kappa(\Sigma_*) \|X\|_\infty^2 \log p = O(1).$$

Then with probability at least 0.8 the subset $\widehat{S}$ output by Algorithm 3 satisfies

$$F(\widehat{S}) \leqslant O(1) \cdot \min_{\|s\|_1 \leqslant k, \|s\|_\infty \leqslant b, s \in \mathbb{N}^n} F(s), \qquad b \in \{1, k\}$$

for all regular criteria F.

The $O(1)$ approximation ratio in Theorem 1 can be improved to an $(1 + \varepsilon)$-approximation if "oversampling" is allowed. Interested readers are referred to Sec. 3.2 of Wang et al. (2017) for details of such an improvement.

2.3.2   Greedy removal techniques for A-optimality

In this section we consider a greedy removal procedure outlined in Algorithm 4. The procedure is specifically for the A-optimality criterion $f_A(\Sigma) = \mathrm{tr}(\Sigma^{-1})$ and the without replacement ($b = 1$) setting.

Built upon a result from Avron & Boutsidis (2013) and an observation on the Karush-Kuhn-Tucker (KKT) conditions of the optimization problem in Eq. (2.4), we have the following result:

Theorem 2. *Let $\widehat{S}$ be the output of Algorithm 4 with initial subset $S_0 = \{i \in [n] : \pi_i^* > 0\}$. If $k > p$ and $\{x_i\}_{i=1}^n$ are independently sampled from continuous densities, then with probability 1*

$$F_A(\widehat{S}) \leqslant \left(1 + \frac{p(p+1)}{2(k-p+1)}\right) \cdot \min_{\|s\|_1 \leqslant k, \|s\|_\infty \leqslant 1, s \in \mathbb{N}^n} F_A(s).$$

Under a slightly stronger condition that $k > 2p$, the approximation ratio $1 + \frac{p(p+1)}{2(k-p+1)}$ can be simplified to $1 + O(p^2/k)$. In addition, the approximation ratio is $1 + o(1)$ if $p^2/k \to 0$, meaning that near-optimal experiment selection is achievable with computationally tractable methods if $O(p^2)$ design points are allowed in the selected subset.

2.3.3 Greedy swapping techniques

In this section we introduce a general-purpose greedy swapping technique that applies to all regular optimality criteria.

Pre-processing The first step of our technique is a pre-processing "whitening" step. Define $W := \sum_{i=1}^{n} \pi_i^* x_i x_i^\top$. The pre-processing step is to *whiten* the design points $\{x_i\}_{i=1}^{n}$ by taking

$$\widetilde{x}_i := W^{-1/2} x_i. \tag{2.6}$$

Note that after pre-processing, the transformed data points $\{\widetilde{x}_i\}_{i=1}^{n}$ satisfies $\sum_{i=1}^{n} \pi_i^* \widetilde{x}_i \widetilde{x}_i^\top = I_{p \times p}$, meaning that they are "whitened" to have identity sample covariance, and hence the name.

Potential functions The main component of our proposed rounding algorithm is a carefully designed *potential* function $\phi(u, v; Z)$, which measures contributions to the least eigenvalue of a p-dimensional positive semi-definite matrix Z by *swapping* design points u and v.

Fix hyper-parameter $\alpha > 0$, whose values will be discussed in the next section. For any p-dimensional positive semi-definite matrix Z define $A_Z := (cI_{p \times p} + \alpha Z)^{-2}$ where $c \in \mathbb{R}$ is the unique real number such that $\text{tr}(A_Z) = 1$. The potential function $\phi(u, v; Z)$ for any pairs of p-dimensional vectors $u, v \in \mathbb{R}^d$ are then defined as

$$\phi(u, v; Z) = \phi_+(u; Z) - \phi_-(v; Z) \tag{2.7}$$

where

$$\phi_+(u; Z) = \frac{u^\top A_Z u}{1 + 2\alpha u^\top A_Z^{1/2} u} \quad \text{and} \quad \phi_-(v; Z) = \frac{v^\top A_Z v}{1 - 2\alpha v^\top A_Z^{1/2} v}. \tag{2.8}$$

While the definitions of the potential function ϕ seems arbitrary, its form has deep roots in online matrix games. More specifically, the form of the intermediate matrix $A_Z = (cI + \alpha Z)^{-2}$, $\text{tr}(A_Z) = 1$ corresponds to updates rules in a *Follow-The-Regularized-Leader (FTRL)* (McMahan, 2011) with the matrix $\ell_{1/2}$-regularizer $\psi(A) = -2\text{tr}(A^{1/2})$, first considered by Allen-Zhu et al. (2015) for a related spectral sparsification problem. The potential function ϕ then falls naturally from a regret analysis of FTRL type policies in online matrix games, summarized in the following lemma:

Lemma 1. *For any p-dimensional vectors $\{u_t, v_t\}_{t=1}^{T}$ and fixed positive-semidefinite matrix Z_0, define $Z_t := Z_0 + \sum_{t'=1}^{t} u_{t'} u_{t'}^\top - v_{t'} v_{t'}^\top$. If further $v_t^\top A_{Z_{t-1}}^{1/2} v_t < 1/2\alpha$ holds for all t, then*

$$\lambda_{\min}(Z_T) \geqslant \sum_{t=1}^{T} \phi(u_t, v_t; Z_{t-1}) - \frac{2\sqrt{p}}{\alpha}. \tag{2.9}$$

Algorithm and approximation guarantees The lower bound of least eigenvalues in Lemma 1 immediately suggests a greedy swapping algorithm, which starts with an arbitrary subset $S_0 \subseteq [n]$ of size K and repeatedly find $i \in S_0$, $j \notin S_0$ for "swapping" so as to maximize $\phi(\widetilde{x}_j, \widetilde{x}_i; Z)$, where $Z = \sum_{\ell \in S} \widetilde{x}_\ell \widetilde{x}_\ell^\top$.

Algorithm 5 A swapping algorithm for rounding

1: **Input**: design points $\{x_i\}_{i=1}^n$, optimal continuous solution π^*, budget k, desired accuracy ε.
2: $\alpha \leftarrow \sqrt{p}/\varepsilon$; $\triangleright$ configuration of hyper-parameters
3: Compute $\widetilde{x}_i = W^{-1/2}x_i$ where $W = \sum_{j=1}^n \pi_j^* x_j x_j^\top$; $\triangleright$ the whitening step
4: $S_0 \subseteq [n]$ an arbitrary subset of size k and $t \leftarrow 1$; $\triangleright$ initialization
5: **while** $\lambda_{\min}(\sum_{i \in S_{t-1}} \widetilde{x}_i \widetilde{x}_i^\top) \leqslant 1 - 3\varepsilon$ **do**
6:      Compute $A_{Z_{t-1}} = (c_t I + \alpha Z_{t-1})^{-2}$ such that $\mathrm{tr}(A_{Z_{t-1}}) = 1$, where $Z_{t-1} = \sum_{i \in S_{t-1}} \widetilde{x}_i \widetilde{x}_i^\top$;
7:      Find $i_t \in S_{t-1}, \widetilde{x}_{i_t}^\top A_{Z_{t-1}}^{1/2} \widetilde{x}_{i_t} < 1/2\alpha$ that minimizes $\phi_-(\widetilde{x}_{i_t}; Z_{t-1})$;
8:      Find $j_t \notin S_{t-1}$ that maximizes $\phi_+(\widetilde{x}_{j_t}; Z_{t-1})$;
9:      Swapping update: $S_t = S_{t-1} \cup \{j_t\} \backslash \{i_t\}$, and $t \leftarrow t + 1$;
10: **end while**
11: **return** $\widehat{s} \in \{0,1\}^n$ where $\widehat{s}_i = 1$ iff $i \in S_T$.

Detailed pseudo-codes of this greedy swapping procedure is given in Algorithm 5. Note that the pair $i_t \in S_{t-1}$, $j_t \notin S_{t-1}$ that maximizes $\phi(\widetilde{x}_{j_t}, \widetilde{x}_{i_t}; Z_{t-1})$ can be found in $O(n+k)$ instead of $O(nk)$ time by separately maximizing and minimizing $\phi_+(\widetilde{x}_{j_t}; Z_{t-1})$ and $\phi_-(\widetilde{x}_{i_t}; Z_{t-1})$ as shown in Steps 6 and 7 in Algorithm 5, because the potential ϕ decomposes additively. In step 5 of Algorithm 5, the unique real number $c_t \in \mathbb{R}$ such that $\mathrm{tr}(A_{Z_{t-1}}) = 1$ can be found by a binary search, because $\mathrm{tr}[(c_t I + \alpha Z_{t-1})^{-2}]$ is a monotonically decreasing function in c_t.

The following theorem gives approximation guarantees of Algorithm 5 when k is not too small compared with p.

Theorem 3. *Suppose $k \geqslant 5p/\varepsilon^2$ for some $\varepsilon \in (0, 1/6]$. Then for any regular f, $\widehat{s} \in \{0,1\}^n$ output by Algorithm 5 has size $\sum_{i=1}^n \widehat{s}_i \leqslant k$ and satisfies*

$$F(\widehat{s}) \leqslant (1 + 6\varepsilon)F(s^*) = (1 + 6\varepsilon) \max_{s \in \{0,1\}^n, \|s\|_1 \leqslant k} F(s).$$

2.4 Application: 3D lightweight structure optimization

We consider an application of our method to a 3D lightweight structure design problem. Most results in this section appeared in (Wang et al., 2018) with more details.

2.4.1 Background

3D lightweight structure design is the question of carefully distributing material mass in complicated 3D structures so that the resulting object has sufficient strength to withstand everyday use. An important task is then to quantify the structural performance of an object under the external forces it may experience during its use. Figure 2.1 from (Ulu et al., 2017) gives an intuitive illustration of the performance of structures under external forces applied at different locations, measured by stress distributed on the rest of the structure among which the *maximum* stress defines the performance of structures under given external forces.
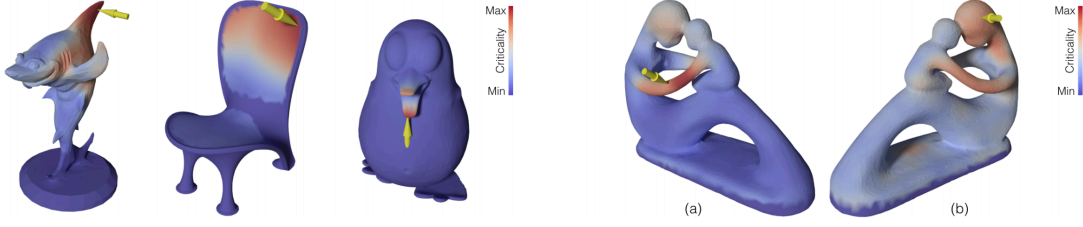
Figure 2.1: Stress distributions on designed structures under external forces.

Suppose external forces can be applied on n possible locations for a specific structure. For each location $i \in [n]$, the stress distribution as well as the maximum stress suffered by an unit amount of external force can be computed by an accurate yet time consuming *finite element analysis (FEA)* method. As each external force location $i \in [n]$ requires an independent FEA run, it is very desirable to select a few "representative" locations $S \subseteq [n]$, $|S| \leqslant k \ll n$ and *estimate* the maximum stress outcomes of the other force locations not selected in S.

This "location selection" problem fits well within the experimental design framework considered in this paper, and in the next two paragraphs we explain how to apply our developed algorithm as well as its experimental performances.

2.4.2 Method

Let G be a graph with n vertices, representing the spatial affinity of the n possible force locations on a structure surface. The readers are referred to (Ulu et al., 2017; Wang et al., 2018) for details of the construction of G. Let L be the graph Laplacian matrix of G, and $X \in \mathbb{R}^{n \times p}$ be the top-p eigenvectors of the graph Laplacian L. A linear regression model is used to model the maximum stress y_i induced by an unit external force applied at location $i \in [n]$ (corresponding to $x_i \in \mathbb{R}^d$ in the top eigenvectors matrix X), as

$$y_i = x_i^\top \beta_0 + \xi_i, \tag{2.10}$$

where β_0 is a p-dimensional unknown regression model and $\{\xi_i\}_{i=1}^n$ are noise variables.

To select a subset $S \subseteq [n]$, $|S| \leqslant k$ of locations, we use the algorithm proposed in the previous sections to solve the discrete optimization problem in Eq. (2.4), restated below:

$$\min_s f\left(\sum_{i=1}^n s_i x_i x_i^\top\right) \quad s.t. \ s_i \in \{0,1\}, \sum_{i=1}^n s_i \leqslant k.$$

The selected subset S is then chosen as all locations with $s_i = 1$, and FEA analysis on these force locations is carried out to obtain their corresponding induced maximum stress y_i. The regression model β_0 is then estimated by ordinary least squares $\widehat{\beta} = (\sum_{i \in S} x_i x_i^\top)^{-1}(\sum_{i \in S} y_i x_i)$, and predictions on the other external force locations are produced by $\widehat{y_i} = x_i^\top \widehat{\beta}$ for $i \notin S$. The force locations $i \in [n]$ are then ranked in descending order according to $\{\widehat{y_i}\}_{i=1}^n$, and FEA analysis is computed again on the top ranked force locations to determine the final location $i^* \in [n]$ that yields the largest stress response y_{i*}. More details of our algorithmic pipeline is given in (Wang et al., 2018).

13

<div align="center">(a) Fertility   (b) Rocking Chair   (c) Shark</div>
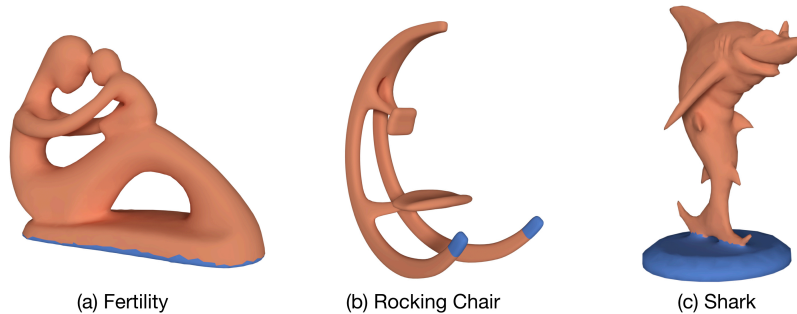
Figure 2.2: Example test structures with complex geometries. Fixed boundary conditions and contact regions are indicated in blue and red, respectively.

2.4.3 Experimental settings

We evaluate the performance of our algorithm on three test structures (FERTILITY, ROCK-INGCHAIR and SHARK) illustrated in Fig. 2.2. Descriptions and some basic statistics of the considered structures are given in (Wang et al., 2018).

In our experiments, we consider 5 methods to sample the force locations subset $S \subseteq [n]$, $|S| \leqslant k$. We compare our proposed algorithm (abbreviated as GREEDY) with baseline methods UNIFORM and LEVSCORE, as well as the previous work K-MEANS (Ulu et al., 2017) and SAMPLING (Wang et al., 2017).

The performance of an algorithm is evaluated by the smallest integer m required so that the top-m ranked lists according to $\{\widehat{y}_i\}_{i=1}^n$ include the external force location $i^* \in [n]$ that actually leads to the maximum stress a structure suffers.

2.4.4 Results and discussion

Table 2.1 reports the performance (m needed to cover $i^* \in [n]$ leading to maximum stress) of our algorithm and its competitors under variance k settings for all three different structures. In Fig. 2.4, we plot the sub-sampled force locations (i.e., S) of our proposed algorithm for $k = 200$ point. We provide the samples obtained by the K-MEANS algorithm in Ulu et al. (2017) for comparison. The difference in the sampling patterns between GREEDY and K-MEANS are quite obvious from the figures.

2.5 Extensions

Our proposed methodologies can be further extended, to applications on generalized linear models, transfer learning and quantized linear regression.

2.5.1 Generalized linear models

In a generalized linear model $\mu(x) = \mathbb{E}[Y|x]$ satisfies $g(\mu(x)) = \eta = x^\top \beta_0$ for some known link function $g : \mathbb{R} \to \mathbb{R}$. Under regularity conditions (Van der Vaart, 1998), the maximum-likelihood

Table 2.1: Results for the three test structures. Numbers in each cell are the smallest m such that the top ranked m force locations include the actual location $i^* \in [n]$ leading to maximal stress. Randomized algorithms (UNIFORM, LEVSCORE and SAMPLING) are run for 10 independent trials and the median performance is reported. Best performing settings are indicated in bold.

	$k =$	25	50	100	150	200	250	300	Total FEAs $(k + m)$
Fertility	UNIFORM	316.5	149	78.5	37.5	98.5	42.5	39	178.5 $(k = 100)$
	LEVSCORE	252.5	54.5	73.5	68.5	42.5	31	13.5	104.5 $(k = 50)$
	K-MEANS	237	**25**	61	82	57	**17**	**16**	75 $(k = 50)$
	SAMPLING	210.5	148.5	51	30	35.5	34	26.5	151 $(k = 100)$
	GREEDY	**12**	26	**13**	**7**	**11**	25	33	**37** $(k = 25)$
RockingChair	UNIFORM	716	857	385.5	42	135.5	269.5	36	192 $(n_{FL} = 150)$
	LEVSCORE	764.5	208.5	**36**	**36**	**36**	**36**	**36**	136 $(k = 100)$
	K-MEANS	4013	4400	4573	4301	4320	4620	4757	4038 $(k = 25)$
	SAMPLING	672.5	282	38.5	38	38	**36**	**36**	138.5 $(k = 100)$
	GREEDY	**36**	**35**	208	35	**36**	**36**	**36**	**61** $(k = 25)$
Shark	UNIFORM	585	384	141.5	208.5	20	**9**	9.5	220 $(n_{FL} = 200)$
	LEVSCORE	478.5	**9**	**9**	**9**	**9**	**9**	**9**	59 $(n_{FL} = 50)$
	K-MEANS	133	102	**9**	**9**	**9**	**9**	**9**	109 $(n_{FL} = 100)$
	SAMPLING	963.5	87	**9**	**9**	**9**	**9**	**9**	109 $(n_{FL} = 100)$
	GREEDY	**9**	171	**9**	**9**	**9**	**9**	**9**	**34** $(n_{FL} = 25)$

estimator $\widehat{\beta}_n \in \mathrm{argmax}_\beta\{\sum_{i=1}^n \log p(y_i|x_i; \beta)\}$ satisfies $\mathbb{E}\|\widehat{\beta}_n - \beta_0\|_2^2 = (1 + o(1))\mathrm{tr}(I(X, \beta_0)^{-1})$, where $I(X, \beta_0)$ is the Fisher's information matrix:

$$I(X, \beta_0) = -\sum_{i=1}^n \mathbb{E}\frac{\partial^2 \log p(y_i|x_i; \beta_0)}{\partial\beta\partial\beta^\top} = -\sum_{i=1}^n \left(\mathbb{E}\frac{\partial^2 \log p(y_i; \eta_i)}{\partial\eta_i^2}\right) x_i x_i^\top. \tag{2.11}$$

Here both expectations are taken over y conditioned on X and the last equality is due to the sufficiency of $\eta_i = x_i^\top \beta_0$. The experiment selection problem is then formulated to select a subset $S \subseteq [n]$ of size k, either with or without duplicates, that minimizes $\mathrm{tr}(I(X_S, \beta_0)^{-1})$.

It is clear from Eq. (2.11) that the optimal subset S^* depends on the unknown parameter β_0, which itself is to be estimated. This issue is known as the *design dependence* problem for generalized linear models (Khuri et al., 2006). One approach is to consider *locally optimal designs* (Chernoff, 1953; Khuri et al., 2006), where a consistent estimate $\breve{\beta}$ of β_0 is first obtained on an initial design subset [2] and then $\breve{\eta}_i = x_i^\top \breve{\beta}$ is supplied to compute a more refined design subset to get the final estimate $\widehat{\beta}$. With the initial estimate $\breve{\beta}$ available, one may apply transform $x_i \mapsto \widetilde{x}_i$ defined as

$$\widetilde{x}_i = \sqrt{-\mathbb{E}\frac{\partial^2 \log p(y_i; \breve{\eta}_i)}{\partial\eta^2}} x_i.$$

Note that under regularity conditions $-\mathbb{E}\frac{\partial^2 \log p(y_i; \breve{\eta}_i)}{\partial\eta_i^2} = \mathbb{E}\left(\frac{\partial \log(y_i; \widetilde{x}_i)}{\partial\eta_i}\right)^2$ is non-negative and hence the square-root is well-defined. All our results are valid with $X = [x_1, \cdots, x_n]^\top$ replaced by

---

[2]Notice that a consistent estimate can be obtained using much fewer points than an estimate with finite approximation guarantee.
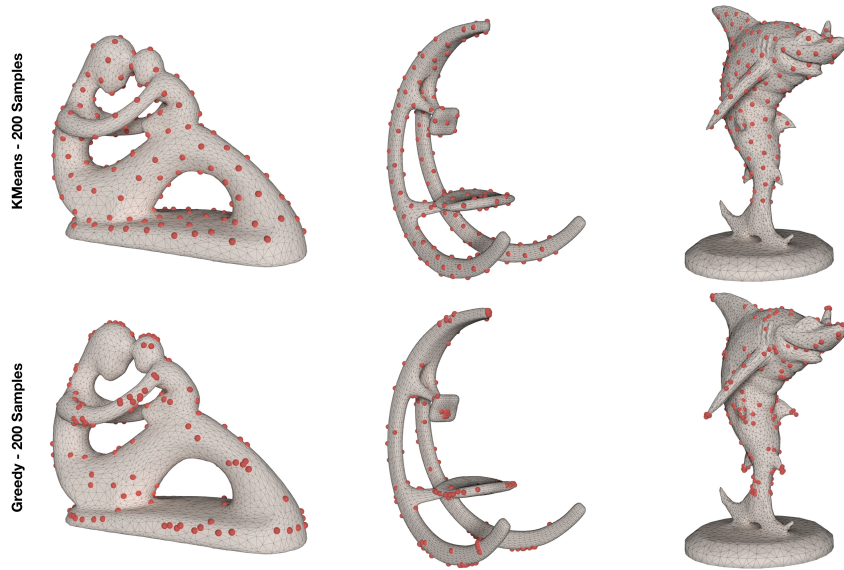
Figure 2.3: Sampled force locations (S) using the K-MEANS algorithm (top row) versus our proposed algorithm (bottom row), for $k = 100$.

$\widetilde{X} = [\widetilde{x}_1, \cdots, \widetilde{x}_n]^\top$ for generalized linear models. Below we consider two generalized linear model examples and derive explicit forms of $\widetilde{X}$.

Example 1: Logistic regression   In a logistic regression model responses $y_i \in \{0, 1\}$ are binary and the likelihood model is

$$p(y_i; \eta_i) = \psi(\eta_i)^{y_i}(1 - \psi(\eta_i))^{1-y_i}, \qquad \text{where} \quad \psi(\eta_i) = \frac{e^{\eta_i}}{1 + e^{\eta_i}}.$$

Simple algebra yields

$$\widetilde{x}_i = \sqrt{\frac{e^{\check{\eta}_i}}{(1 + e^{\check{\eta}_i})^2}} x_i,$$

where $\check{\eta}_i = x_i^\top \check{\beta}$.

Example 2: Poisson count model   In a Poisson count model the response variable y_i takes values of non-negative integers and follows a Poisson distribution with parameter $\lambda = e^{\eta_i} = e^{x_i^\top \beta_0}$. The likelihood model is formally defined as

$$p(y_i = r; \eta_i) = \frac{e^{\eta_i r} e^{-e^{\eta_i}}}{r!}, \qquad r = 0, 1, 2, \cdots.$$

Simple algebra yields

$$\widetilde{x}_i = \sqrt{e^{\check{\eta}_i}} x_i,$$

where $\check{\eta}_i = x_i^\top \check{\beta}$.

16
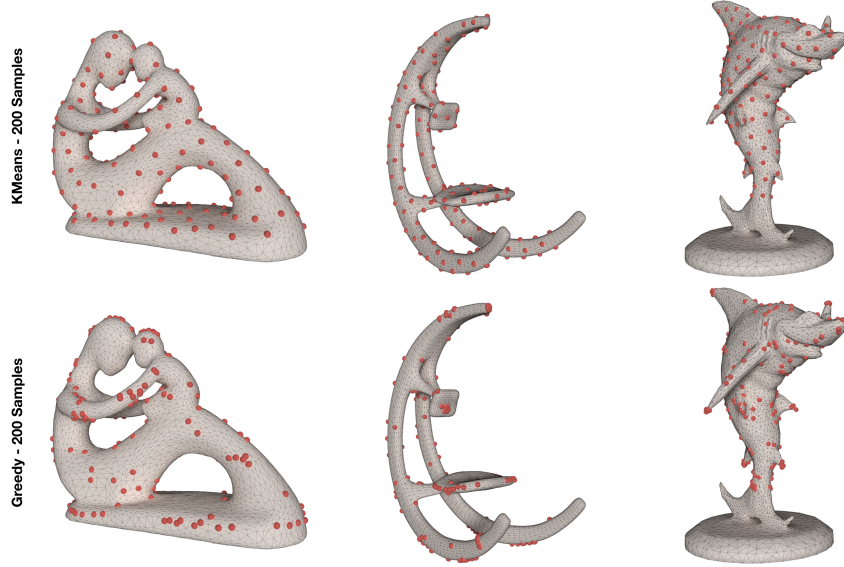
Figure 2.4: Sampled force locations (S) using the K-MEANS algorithm (top row) versus our proposed algorithm (bottom row), for $k = 200$.

2.5.2 Transfer learning and Delta's method

Suppose $g(\beta_0)$ is the quantity of interest, where $\beta_0 \in \mathbb{R}^p$ is the parameter in a linear regression model and $g : \mathbb{R}^p \to \mathbb{R}^m$ is some known function. Let $\widehat{\beta}_n = (X^\top X)^{-1} X^\top y$ be the OLS estimate of β_0. If ∇g is continuously differentiable and $\widehat{\beta}_n$ is consistent, then by the classical delta's method (Van der Vaart, 1998) $\mathbb{E}\|g(\widehat{\beta}_n) - g(\beta_0)\|_2^2 = (1+o(1))\sigma^2 \mathrm{tr}(\nabla g(\beta_0)(X^\top X)^{-1} \nabla g(\beta_0)^\top) = (1 + o(1))\sigma^2 \mathrm{tr}(G_0(X^\top X)^{-1})$, where $G_0 = \nabla g(\beta_0)^\top \nabla g(\beta_0)$. If G_0 depends on the unknown parameter β_0 then the design dependence problem again exists, and a locally optimal solution can be obtained by replacing G_0 in the objective function with $\check{G} = \nabla g(\check{\beta})^\top \nabla g(\check{\beta})$ for some initial estimate $\check{\beta}$ of β_0.

If $\check{G}$ is invertible, then there exists invertible $p \times p$ matrix $\check{P}$ such that $\check{G} = \check{P}\check{P}^\top$ because $\check{G}$ is positive definite. Applying the linear transform

$$x_i \mapsto \widetilde{x}_i = \check{P}^{-1} x_i$$

we have that $\mathrm{tr}[G_0(X^\top X)^{-1}] = \mathrm{tr}[(\widetilde{X}^\top \widetilde{X})^{-1}]$, where $\widetilde{X} = [\widetilde{x}_1, \cdots, \widetilde{x}_n]^\top$. Our results remain valid by operating on the transformed matrix $\widetilde{X} = X\check{P}^{-\top}$.

Example: prediction error. In some application scenarios the *prediction error* $\|Z\widehat{\beta} - Z\beta_0\|_2^2$ rather than the estimation error $\|\widehat{\beta} - \beta_0\|_2^2$ is of interesting, either because the linear model is used mostly for prediction or component of the underlying model β_0 lack physical interpretations. Another interesting application is the *transfer learning* (Pan & Yang, 2010), in which the training and testing data have different designs (e.g., Z instead of X) but share the same conditional distribution of labels, parameterized by the linear model β_0.

Suppose $Z \in \mathbb{R}^{m \times p}$ is a known full-rank data matrix upon which predictions are seeked, and define $\widehat{\Sigma}_Z = \frac{1}{m} Z^\top Z > 0$ to be the sample covariance of Z. Our algorithmic framework

as well as its corresponding analysis remain valid for such prediction problems with transform $x_i \mapsto \widehat{\Sigma}_Z^{-1/2} x_i$. In particular, the guarantees for the greedy algorithm and the with replacement sampling algorithm remain unchanged, and the guarantee for the without replacement sampling algorithm is valid as well, except that the $\|\Sigma_*^{-1}\|_2$ and $\kappa(\Sigma_*)$ terms have to be replaced by the (relaxed) optimal sample covariance after the linear transform $x_i \mapsto \widehat{\Sigma}_Z^{-1/2} x_i$.

2.5.3   Quantized linear regression

Consider the noiseless linear signal model

$$y = X\beta_0 \tag{2.12}$$

where $X \in \mathbb{R}^{n \times p}$ is an exactly known design matrix, typically generated from certain physical procedures, and $\beta_0 \in \mathbb{R}^p$ is an unknown p-dimensional signal to be recovered. We restrict ourselves to the "low-dimensional" setting $p < n$. Unlike the classical linear regression model ubiquitous in the statistics literature, the model in Eq. (2.12) is assumed to be *noiseless* as no noise variables are included in the measurement model $y = X\beta_0$. Such a model arises in various scenarios where the signal can be expressed or well-approximated by a small number of basis elements. We mention one specific example from the framework of signal processing on graphs (Sandryhaila & Moura, 2014; Shuman et al., 2013), which studies signals with an underlying complex structure that is modeled by a graph such as measurements at nodes of a network. The band-limited model for graph signals is a linear model in which the network node measurements y are well represented by a linear model where the features are the eigenvectors of the graph Laplacian or adjacency matrix corresponding to the smallest/largest eigenvalues, respectively.

The measurements of y, however, can only be made up to a total of k binary bits and hence cannot be perfectly accurate. Such measurement-constrained settings are ubiquitous in statistical signal processing and machine learning applications, such as brain signal sensing (Lebedev & Nicolelis, 2006), Internet of Things (Zhou et al., 2013) and electric power grids (Nabaee & Labeau, 2012). It is therefore important to design intelligent bit allocation algorithms such that the recovery of signal β_0 is the most accurate possible subject to given bit measurement constraints.

The bit allocation problem in quantized linear regression can be formulated as follows:

Problem 1 (passive bit allocation). *Given exactly measured design $X \in \mathbb{R}^{n \times p}$ and a bit budget $k \in \mathbb{N}$, $k \geqslant n$, find a bit allocation $k = (k_1, \cdots, k_n) \in \mathbb{N}^n$, $k_1 + \cdots + k_n \leqslant k$ such that the mean square error between the recovered signal $\widehat{\beta}_k$ and the true signal β_0 is mimimized.*

Suppose a bit allocation strategy $k = (k_1, \cdots, k_n) \in \mathbb{N}_+^n$ is given, such that $\sum_{i=1}^n k_i \leqslant n$. Let $\mathrm{round}(\cdot)$ be the rounding operator towards the closest integer and $U[a, b]$ be the uniform distribution on interval $[a, b]$. The observed quantized value of $y_i = x_i^\top \beta_0$ with $k_i \in \mathbb{N}_+$ binary bits of measurement can then be expressed as

$$\widetilde{y}_i = 2^{-(k_i-1)} \cdot \mathrm{round}\left[2^{k_i-1} \left(\frac{y_i}{M} + \delta_i \right) \right] \tag{2.13}$$

where $M := \max_{1 \leqslant i \leqslant n} |x_i^\top \beta_0|$ is a known bounded constant and $\delta \sim U[-2^{-k_i}M, 2^{-k_i}M]$ is a *dithering* variable that introduces additional stochasticity to the deterministic model (2.12). The

dithering step de-couples the statistical dependency in the quantized error and is an important concept in the signal processing literature (Schuchman, 1964). Note also that the most significant bit in $\widetilde{y}_i$ indicates the sign of y_i, and hence only $(k_i - 1)$ bits are available to measure the absolute value of y_i.

As the number of measure bits differ for different design points x_i, the rounding (quantization) error of each $\widetilde{y}_i$ also differs, making the quantized linear model (2.13) similar to a linear regression model with heteroscedastic noise. Because the noise levels are known (controlled by the bit allocation strategy k directly), a *weighted* Ordinary Least Squares (OLS) estimator is reasonable for the recovery of β_0 which we define as follows:

$$\widehat{\beta}_k \in \text{argmin}_{\beta \in \mathbb{R}^p} \sum_{i:k_i > 0} 4^{k_i+1}(\widetilde{y}_i - x_i^\top \beta)^2. \tag{2.14}$$

The following lemma upper bounds the mean square error of $\widehat{\beta}_k$. Its proof is a standard analysis of weighted OLS estimators for heteroscedastic linear models.

Lemma 2. *The weighted OLS estimator $\widehat{\beta}_k$ satisfies*

$$\mathbb{E}\|\widehat{\beta}_k - \beta_0\|_2^2 \leqslant M^2 \cdot \text{tr}\left[\left(\sum_{i:k_i>0} 4^{k_i+1} x_i x_i^\top\right)^{-1}\right]. \tag{2.15}$$

With Lemma 2, we can solve the following continuous optimization problem:

$$\min_{\pi = (\pi_1, \cdots, \pi_n) \in \mathbb{R}^n} \text{tr}\left[\left(\sum_{i=1}^n (4^{\pi_i} - 1) x_i x_i^\top\right)^{-1}\right] \quad \pi_i \geqslant 0, \;\; \|\pi\|_1 \leqslant k_0. \tag{2.16}$$

After the continuous optimization is solved, with (approximately) optimal solution π, a *leverage score sampling* algorithm with careful rounding procedures can be used to obtain integer-valued bit allocation. The pseudocode of the algorithm is listed in Algorithm 6, and some basic properties of Algorithm 6 are described in (Wang & Singh, 2018).

2.6 Summary and related works

The works presented in this section can be best summarized as *computational* aspects of experimental design. The main motivations behind the presented work is the (theoretical) *computational intractability* of many (optimal) experimental design problems, including even the very basic ones such as A-optimal or D-optimal designs of discrete design point sets. Instead of studying heuristic algorithms, our results fall under the framework of polynomial-time computability in theory of computation, with rigorous approximation guarantees in terms of optimal design objective values.

Experimental design is an old topic and we do not intend to provide a comprehensive literature review at this point. Interested readers should consult the classical references of (Chaloner & Verdinelli, 1995; Fedorov, 1972; Pukelsheim, 2006). Related works reviewed in this section focus primarily on computational aspects of the experimental design problem with rigorous theoretical/approximation guarantees.

Algorithm 6 Bit allocation algorithm by leverage score sampling

---

1: **Input**: $X \in \mathbb{R}^{n \times p}$, quantization budget k, support size $s < k - p$, number of repetitions B.
2: **Output**: $\widehat{k} \in \mathbb{N}^n$ satisfying $k_1 + \cdots + k_n \leqslant k$.
3: Continuous optimization: solve for π^*, the optimal solution of Eq. (2.16).
4: Pre-conditioning: $\Sigma_* = \sum_{i=1}^n 4^{\pi_i^* + 1} x_i x_i^\top$; leverage scores $\ell_i = x_i^\top \Sigma_*^{-1} x_i$.
5: **for** $b \in \{1, \cdots, B\}$ **do**
6:     Initialization: $\{w_i^{(b)}\}_{i=1}^n = 0$.
7:     Repeat for s times: sample $i_t \in [n]$ from the categorical distribution $\Pr[i_t = i] = p_i \propto 4^{\pi_i^* + 1} \ell_i$ and update $w_{i_t}^{(b)} \leftarrow w_{i_t}^{(b)} + 4^{\pi_i^* + 1}/p_{i_t}$.
8:     Define allocation $\widehat{k}^{(b)}$: $\widehat{k}_i^{(b)} = \left\lceil \frac{(k-s)\log(1+w_i^{(b)})}{\sum_{j: w_j^{(b)} > 0} \log(1+w_j^{(b)})} \right\rceil$ and $\widehat{k}_i^{(b)} = 0$ if $w_i^{(b)} = 0$.
9: **end for**
10: Output $\widehat{k}$ in $\{\widehat{k}^{(b)}\}_{b=1}^B$ with the smallest objective $F(\widehat{k}; X)$.

D-optimal designs Perhaps the most well-studied optimality criterion is D-optimality $f_D(\Sigma) = \det(\Sigma)^{1/p}$,' whose negative logarithm (i.e., $\log \det \Sigma$) is *submodular*, a property that sometimes gives rises to $1 - 1/e$ approximation ratio using pipage rounding (Ageev & Sviridenko, 2004). Unfortunately, $\log \det \Sigma$ can be negative and thus pipage rounding could fail to provide a constant relative approximation ratio with respect to $\det(\Sigma)$ or $\det(\Sigma)^{1/p}$. In (Bouhtou et al., 2010), Bouhtou et al. proposed to maximize a function $h(\Sigma) := \frac{1}{p}\text{tr}(\Sigma^q)$ for $q \in (0, 1]$, and it satisfies $\lim_{q \to 0}(h(\Sigma))^{-1/q} = f_D(\Sigma)$. They showed that $h(\Sigma)$ is submodular and gave a $(1 - 1/e)$ approximation to $h(\Sigma)$ for every $q \in (0, 1]$ using pipage rounding. This does not translate to any bounded approximation ratio for $f_D(\Sigma)$ because $(1 - 1/e)^{-1/q}$ is unbounded when q approaches zero.

Summa et al. (2015) gave a polynomial-time algorithm for a related maximum volume simplex (MVS) problem in computational geometry with an $O(\log p)$ approximation ratio, which was later improved to $O(1)$ by Nikolov (2015); Nikolov & Singh (2016). Their results imply an e approximation ratio in the special case of $k = p$. On the other hand, Summa et al. (2015) showed that there exists a constant $c > 1$ such that polynomial-time c-approximation of the D-optimality is impossible for the $p = k$ case, unless $\mathsf{P} = \mathsf{NP}$. Therefore, additional assumptions on k are necessary for the $(1 + \varepsilon)$-approximation regime we consider in this paper.

Concurrent and independent of our work, the works of (Singh & Xie, 2017) achieved $(1+\varepsilon)$-relative approximation for the D-optimality criteria under the weaker condition that $k = \Omega(p/\varepsilon + \log(\varepsilon^{-1})/\varepsilon^2)$. Their techniques are based on volume sampling of symmetric elementary functions of matrix eigenvalues, and are less likely to be extendable to general optimality criteria objectives. Indeed, Nikolov et al. (2018) proved a negative result showing that no continuous-relaxation based method can possibly attain $(1 + \varepsilon)$ approximation for the E-optimaliy unless $k = \Omega(p/\varepsilon^2)$, essentially showing our theoretical analysis is the best possible one can hope for.

A/V-optimal designs For the A-optimality criterion, Avron & Boutsidis (2013) proposed a greedy algorithm with an approximation ratio $O(n/k)$ with respect to $f(\sum_{i=1}^n x_i x_i^\top)$. This ratio is

tight for their algorithm in the worst case,[3] Li et al. (2017a) further computationally accelerated this greedy algorithm, and achieved similar approximation guarantees. Nikolov et al. (2018) showed that the $k = \Omega(p/\varepsilon + \log(\varepsilon^{-1}/\varepsilon^2)$ condition is sufficient for $(1 + \varepsilon)$-approximation of A/V-optimality as well.

Fast and subsampling least squares solvers   There has been an increasing amount of work on fast solvers for the general least-square problem $\min_\beta \|y - X\beta\|_2^2$. Most of existing work along this direction (Dhillon et al., 2013; Drineas et al., 2011; Raskutti & Mahoney, 2015; Woodruff, 2014) focuses solely on the computational aspects and do not consider statistical constraints such as limited measurements of y. A convex optimization formulation was proposed in Davenport et al. (2015) for a constrained adaptive sensing problem, which is a special case of our setting, but without finite sample guarantees with respect to the combinatorial problem.

Popular subsampling techniques such as leverage score sampling (Drineas et al., 2008) were studied in least square and linear regression problems (Chen et al., 2015a; Ma et al., 2015). None of these methods achieve near minimax optimal statistical efficiency in terms of estimating the underlying linear model β_0, since the methods can be worse than uniform sampling which has a fairly large approximation constant for general X.

Active regression   Another related area is *active learning* (Chaudhuri et al., 2015; Hazan & Karnin, 2015; Sabato & Munos, 2014), which is a stronger setting where feedback from prior measurements can be used to guide subsequent data selection. Chaudhuri et al. (2015) analyzes an SDP relaxation in the context of active maximum likelihood estimation.

2.7   Proofs

2.7.1   Proof of Theorem 1

The following lemma is key to the proof of Theorem 1.

Lemma 3. *Define $\widehat{\Sigma}_{\widehat{S}} = X_{\widehat{S}}^\top X_{\widehat{S}}$. Suppose the following conditions hold:*

$$
\begin{aligned}
\textit{With replacement}: &\quad p\log T/T = O(1); \\
\textit{Without replacement}: &\quad \|\Sigma_*^{-1}\|_2 \kappa(\Sigma_*)\|X\|_\infty^2 \log p = O(T/n).
\end{aligned}
$$

Then with probability at least 0.9 the following holds:

$$
z^\top \widehat{\Sigma}_{\widehat{S}} z \geqslant K_T z^\top \Sigma_* z, \qquad \forall z \in \mathbb{R}^p, \tag{2.17}
$$

where $K_T = \Omega(T/k)$ for with replacement and $K_T = \Omega(T/n)$ for without replacement.

We also need to relate conditions on T in Lemma 3 to interpretable conditions on subset budget k:

---

[3]In the worst case, even the exact minimum $\min_{|S|\leqslant k} f(\sum_{i\in S} x_i x_i^\top)$ can be indeed $O(n/k)$ times larger than $f(\sum_{i=1}^n x_i x_i^\top)$ (Avron & Boutsidis, 2013) different from the subset selection objective in Eq. (2.3). This worst-case scenario may not always happen, but to the best our knowledge, their proof is tight in this worst case.

Lemma 4. *Let $\delta > 0$ be an arbitrarily small fixed failure probability. The with probability at least $1 - \delta$ we have that $T \geqslant \delta k$ for with replacement and $T \geqslant \delta n$ for without replacement.*

Proof. For with replacement we have $\mathbb{E}[\sum_{t=1}^{T} w_t] = T$ and for without replacement we have $\mathbb{E}[\sum_{t=1}^{T} w_t] = Tk/n$. Applying Markov's inequality on $\Pr[\sum_{t=1}^{T} w_t > k]$ for $T = \delta k$ and/or $T = \delta n$ we complete the proof of Lemma 4. $\qquad\square$

Combining Lemmas 3 and 4 with $\delta = 0.1$ and note that $T \leqslant k$ almost surely (because $w_t \geqslant 1$), we prove Theorem 1.

The rest of this section is devoted to proving Lemma 3. We treat the with replacement and without replacement settings separately.

Proof of the with replacement setting For the with replacement setting, we adopt the proof strategy of Spielman & Srivastava (2011). Define $\Phi = \mathrm{diag}(\pi^*)$ and $\Pi = \Phi^{1/2} X \Sigma_*^{-1} X^\top \Phi^{1/2} \in \mathbb{R}^{n \times n}$. The following proposition lists properties of Π:

Proposition 1 (Properties of projection matrix)**.** *The following properties for Π hold:*

1. *Π is a projection matrix. That is, $\Pi^2 = \Pi$.*
2. $\mathrm{Range}(\Pi) = \mathrm{Range}(\Phi^{1/2}X)$.
3. *The eigenvalues of Π are 1 with multiplicity p and 0 with multiplicity $n - p$.*
4. $\Pi_{ii} = \|\Pi_{i,\cdot}\|_2^2 = \pi_i^* x_i^\top \Sigma_*^{-1} x_i$.

Proof. Proof of 1: By definition, $\Sigma_* = X^\top \Phi X$ and subsequently

$$
\begin{aligned}
\Pi^2 &= \Phi^{1/2} X \Sigma_*^{-1} X^\top \Phi^{1/2} \Phi^{1/2} X \Sigma_*^{-1} X^\top \Phi^{1/2} \\
&= \Phi^{1/2} X (X^\top \Phi X)^{-1} X^\top \Phi X (X^\top \Phi X)^{-1} \Phi^{1/2} \\
&= \Phi^{1/2} X (X^\top \Phi X)^{-1} X^\top \Phi^{1/2} = \Pi.
\end{aligned}
$$

Proof of 2: First note that $\mathrm{Range}(\Pi) = \mathrm{Range}(\Phi^{1/2} X \Sigma_*^{-1} X^\top \Phi^{1/2}) \subseteq \mathrm{Range}(\Phi^{1/2} X)$. For the other direction, take arbirary $u \in \mathrm{Range}(\Phi^{1/2} X)$ and express u as $u = \Phi^{1/2} X v$ for some $v \in \mathbb{R}^p$. We then have

$$
\begin{aligned}
\Pi u &= \Phi^{1/2} X \Sigma_*^{-1} X^\top \Phi^{1/2} u \\
&= \Phi^{1/2} X (X^\top \Phi X)^{-1} X^\top \Phi^{1/2} \Phi^{1/2} X v \\
&= \Phi^{1/2} X v = u
\end{aligned}
$$

and hence $u \in \mathrm{Range}(\Pi)$.

Proof of 3: Because $\Sigma_* = X^\top \Phi X$ is invertible, the $n \times p$ matrix $\Phi^{1/2} X$ must have full column rank and hence $\ker(\Phi^{1/2} X) = \{0\}$. Consequently, $\dim(\mathrm{Range}(\Pi)) = \dim(\mathrm{Range}(\Phi^{1/2} X)) = p - \dim(\ker(\Phi^{1/2} X)) = p$. On the other hand, the eigenvalues of Π must be either 0 or 1 because Π is a projection matrix. So the eigenvalues of Π are 1 with multiplicity p and 0 with multiplicity $n - p$.

Proof of 4: By definition,

$$
\Pi_{ii} = \sqrt{\pi_i^*} x_i^\top \Sigma_*^{-1} x_i \sqrt{\pi_i^*} = \pi_i^* x_i^\top \Sigma_*^{-1} x_i.
$$

22

In addition, Π is a symmetric projection matrix. Therefore,

$$\Pi_{ii} = [\Pi^2]_{ii} = \Pi_{i,\cdot}^\top \Pi_{i,\cdot} = \|\Pi_{i,\cdot}\|_2^2.$$

$\square$

The following lemma shows that a spectral norm bound over deviation of the projection matrix implies spectral approximation of the underlying (weighted) covariance matrix.

Lemma 5 (Spielman & Srivastava (2011), Lemma 4). *Let $\Pi = \Phi^{1/2} X \Sigma_*^{-1} X^\top \Phi^{1/2}$ and W be an $n \times n$ non-negative diagonal matrix. If $\|\Pi W \Pi - \Pi\|_2 \leqslant \epsilon$ for some $\epsilon \in (0, 1/2)$ then*

$$(1 - \epsilon) u^\top \Sigma_* u \leqslant u^\top \widetilde{\Sigma}_* u \leqslant (1 + \epsilon) u^\top \Sigma_* u, \quad \forall u \in \mathbb{R}^p,$$

where $\Sigma_* = X^\top \Phi X$ and $\widetilde{\Sigma}_* = X^\top W^{1/2} \Phi W^{1/2} X$.

We next proceed to find an appropriate diagonal matrix W and validate Lemma 5. Define $\widetilde{w}_j^* = k/T \cdot w_j^* = \pi_j^*/(T p_j^{(1)})$ and let $\widehat{\Sigma}_T = \sum_{t=1}^T \widetilde{w}_{i_t}^* x_{i_t} x_{i_t}^\top$. Because $\widehat{\Sigma}_T = \frac{k}{T} \widehat{\Sigma}_{\widehat{\Sigma}_W} \preceq \widehat{\Sigma}_{\widehat{S}}$, we have that $\widehat{\Sigma}_{\widehat{S}} \succeq \frac{T}{k} \widehat{\Sigma}_T$ and hence $z^\top \widehat{\Sigma}_{\widehat{S}} z \geqslant \frac{T}{k} z^\top \widehat{\Sigma}_T z$ for all $z \in \mathbb{R}^p$. Therefore, to lower bound the spectrum of $\widehat{\Sigma}_{\widehat{S}}$ it suffices to lower bound the spectrum of $\widehat{\Sigma}_T$.

Define diagonal matrix $\widetilde{W}$ as

$$\widetilde{W}_{jj} = \frac{\sum_{t=1}^T \widetilde{w}_{i_t}^* \mathbb{I}[i_t = j]}{\pi_j^*}, \qquad j = 1, \cdots, n.$$

We have that $\widetilde{\Sigma}_* = \widehat{\Sigma}_T$ for this particular choice of W.

Lemma 6. *For any $\epsilon > 0$,*

$$\Pr\left[\|\Pi \widetilde{W} \Pi - \Pi\|_2 > \epsilon \right] \leqslant 2 \exp\left\{ -C \cdot \frac{T \epsilon^2}{p \log T} \right\}, \tag{2.18}$$

where $C > 0$ is an absolute constant.

Proof. Define n-dimensional random vector v as [4]

$$\Pr\left[v = \sqrt{\frac{T \widetilde{w}_j^*}{\pi_j^*}} \Pi_{j\cdot} \right] = p_j^{(1)}, \qquad j = 1, \cdots, n.$$

Let $v_1, \cdots, v_k$ be i.i.d. copies of v and define $A_t = v_t v_t^\top$. By definition, $\Pi W^{(1)} \Pi$ is equally distributed with $\frac{1}{k} \sum_{t=1}^k A_t$. In addition,

$$\mathbb{E} A_t = \sum_{j=1}^n \frac{T w_j^*}{\pi_j^*} p_j^{(1)} \Pi_{j\cdot} \Pi_{j\cdot}^\top = \Pi^2 = \Pi,$$

---

[4]For those j with $\pi_j^* = 0$, we have by definition that $p_j^{(1)} = 0$.

which satisfies $\|\mathbb{E}A_t\|_2 = 1$, and

$$\|A_t\|_2 = \|v_t\|_2^2 \leqslant \sum_{1 \leqslant j \leqslant n} \frac{T \widetilde{w}_j^*}{\pi_j^*} \|\Pi_{j\cdot}\|_2^2 = \sum_{1 \leqslant j \leqslant n} \frac{1}{p_j^{(1)}} \|\Pi_{j\cdot}\|_2^2 \leqslant \sup_{1 \leqslant j \leqslant n} \frac{p}{x_j^\top \Sigma_*^{-1} x_j} \cdot x_j^\top \Sigma_*^{-1} x_j = p.$$

Applying Lemma 95 we have that

$$\Pr\left[\|\Pi \widetilde{W} \Pi - \Pi\|_2 > \epsilon\right] \leqslant 2 \exp\left\{-C \cdot \frac{T\epsilon^2}{p \log T}\right\}.$$

$\square$

Set $t = O(1)$ and equate the right-hand side of Eq. (2.18) with $O(1)$. We then have

$$\Pr\left[\widehat{\Sigma}_T \succeq \Omega(1) \cdot \widehat{\Sigma}_*\right] = \Omega(1) \quad \text{if } p \log T/T = O(1).$$

Subsequently, under the condition that $p \log T/T = O(1)$, with probability at least 0.9 it holds that

$$\widehat{\Sigma}_{\widehat{S}} \succeq \Omega(T/k) \cdot \Sigma_*,$$

which completes the proof of Lemma 3 for the with replacement setting.

Proof of the without replacement setting Define $\widehat{\Sigma}_{\widehat{R}} = X_{R_T}^\top X_{R_T} = \sum_{t=1}^T \pi_{i_t}^* x_{i_t} x_{i_t}^\top$. Conditioned on R_T, the subset $\widehat{S} = S_T$ is selected randomly as a subset of R_T. We can then use matrix concentration inequalities to upper bound the discrepancy between $\widehat{\Sigma}_{\widehat{S}}$ and $\widehat{\Sigma}_{\widehat{R}}$.

More specifically, define independently distributed random matrices $A_1, \cdots, A_T$ as

$$A_t = (w_t - \pi_{i_t}^*) x_{i_t} x_{i_t}^\top, \quad t = 1, \cdots, T.$$

Note that w_t is a random Bernoulli variable with $\Pr[w_t = 1] = k p_{i_t}^{(2)} = \pi_{i_t}^*$. Therefore, $\mathbb{E}A_t = 0$. In addition,

$$\sup_{1 \leqslant t \leqslant T} \|A_t\|_2 \leqslant \sup_{1 \leqslant j \leqslant n} \|x_j\|_2^2 \leqslant \|X\|_\infty^2 \quad a.s.,$$

and

$$\left\|\sum_{t=1}^T \mathbb{E}A_t^2\right\|_2 = \left\|\sum_{t=1}^T \pi_{i_t}^*(1 - \pi_{i_t}^*) \|x_{i_t}\|_2^2 x_{i_t} x_{i_t}^\top\right\|_2 \leqslant \|X\|_\infty^2 \|\widehat{\Sigma}_{\widehat{R}}\|_2.$$

Noting that $\sum_{t=1}^T A_t = \widehat{\Sigma}_{\widehat{S}} - \widehat{\Sigma}_{\widehat{R}}$ and invoking Lemma 96, we have that

$$\Pr\left[\|\widehat{\Sigma}_{\widehat{S}} - \widehat{\Sigma}_T\|_2 > t \big| R_T\right] \leqslant 2p \cdot \exp\left\{-\frac{t^2}{3\|\widehat{\Sigma}_T\|_2\|X\|_\infty^2 + 2\|X\|_\infty^2 t}\right\}.$$

Setting $t = O(1) \cdot \lambda_{\min}(\Sigma_T)$ we have that, if $\|\widehat{\Sigma}_T^{-1}\|_2 \kappa(\widehat{\Sigma}_T) \|X\|_\infty^2 \log p = O(1)$, then with probability at least 0.95 conditioned on $\widehat{\Sigma}_T$

$$\Omega(1) \cdot \widehat{\Sigma}_T \preceq \widehat{\Sigma}_{\widehat{S}} \preceq O(1) \cdot \widehat{\Sigma}_T. \tag{2.19}$$

24

It remains to establish spectral similarity between $\widehat{\Sigma}_T$ and $\frac{T}{n}\Sigma_*$, a scaled version of Σ_*. Define deterministic matrices $A_1, \cdots, A_n$ as

$$A_j = \pi_j^* x_j x_j^\top - \frac{1}{n}\Sigma_*, \qquad j = 1, \cdots, n.$$

By definition, $\sum_{j=1}^n A_j = 0$ and $\sum_{t=1}^T A_{\sigma(t)} = \widehat{\Sigma}_T - \frac{T}{n}\Sigma_*$, where σ is a random permutation from $[n]$ to $[n]$. In addition,

$$\sup_{1 \leqslant j \leqslant n} \|A_j\|_2 \leqslant \frac{1}{n}\|\Sigma_*\|_2 + \sup_{1 \leqslant j \leqslant n} \|x_j\|_2^2 \leqslant 2\|X\|_\infty^2$$

and

$$\begin{aligned}
\frac{T}{n}\left\|\sum_{j=1}^n A_j^2\right\|_2 &\leqslant \frac{2T}{n}\left(\left\|\sum_{j=1}^n (\pi_j^*)^2\|x_i\|_2^2 x_i x_i^\top\right\|_2 + \frac{1}{n^2}\|\Sigma_*\|_2^2\right) \\
&\leqslant \frac{2T}{n}\left(\|X\|_\infty^2 \left\|\sum_{j=1}^n \pi_j^* x_i x_i^\top\right\|_2 + \frac{1}{n^2}\|\Sigma_*\|_2^2\right) \\
&\leqslant \frac{2T}{n}\left(\|X\|_\infty^2\|\Sigma_*\|_2 + \frac{1}{n^2}\|\Sigma_*\|_2^2\right) \\
&\leqslant \frac{4T}{n}\|X\|_\infty^2\|\Sigma_*\|_2.
\end{aligned}$$

Invoking Lemma 97, we have that

$$\Pr\left[\left\|\widehat{\Sigma}_T - \frac{T}{n}\Sigma_*\right\|_2 > t\right] \leqslant p \exp\left\{-t^2\left[\frac{48T}{n}\|X\|_\infty^2\|\Sigma_*\|_2 + 8\sqrt{2}\|X\|_\infty^2 t\right]^{-1}\right\}.$$

Set $t = O(T/n) \cdot \lambda_{\min}(\Sigma_*)$. We then have that, if $\|\Sigma_*^{-1}\kappa(\Sigma_*)\|X\|_\infty^2 \log p = O(T/n)$ holds, then with probability at least 0.95

$$\Omega(T/n) \cdot \Sigma_* \preceq \widehat{\Sigma}_T \preceq O(T/n) \cdot \Sigma_*. \tag{2.20}$$

Combining Eqs. (2.19,2.20) and noting that $\|\widehat{\Sigma}_T^{-1}\|_2 \leqslant O(\frac{n}{T})\|\Sigma_*^{-1}\|_2$, $\kappa(\Sigma_T) \leqslant O(1)\kappa(\Sigma_*)$, we complete the proof of Lemma 3 under the without replacement setting.

2.7.2 Proof of Theorem 2

We first extract the following lemma from (Avron & Boutsidis, 2013), which analyzes the performance of the greedy removal algorithm used as a sub-routine in our Algorithm 4 when starting from an arbitrary set S_0.

Lemma 7. *Suppose $\widehat{S} \subseteq [n]$ of size k is obtained by running algorithm in Algorithm 4 with an initial subset $S_0 \subseteq [n]$, $|S_0| \geqslant k$. Both $\widehat{S}$ and S_0 are standard sets (i.e., without replacement). Then*

$$\operatorname{tr}\left[(X_{\widehat{S}}^\top X_{\widehat{S}})^{-1}\right] \leqslant \frac{|S_0| - p + 1}{k - p + 1}\operatorname{tr}\left[(X_{S_0}^\top X_{S_0})^{-1}\right].$$

In (Avron & Boutsidis, 2013) the greedy removal procedure in Algorithm 4 is applied to the entire design set $S_0 = [n]$, which gives approximation guarantee $\text{tr}[(X_{\hat{S}}^\top X_{\hat{S}})^{-1}] \leqslant \frac{n-p+1}{k-p+1}\text{tr}[(X^\top X)^{-1}]$. This results in an approximation ratio of $\frac{n-p+1}{k-p+1}$ by applying the trivial bound $\text{tr}[(X^\top X)^{-1}] \leqslant \min_{s_i \in \{0,1\}, \sum_i s_i \leqslant k} F(s)$, which is tight for a design that has exactly k non-zero rows.

To further improve the approximation ratio, we consider applying the greedy removal procedure with S_0 equal to the support of π^*; that is, $S_0 = \{j \in [n] : \pi_j^* > 0\}$. Because $\|\pi^*\|_\infty \leqslant 1$ under the without replacement setting, we have the following corollary:

Corollary 1. *Let S_0 be the support of π^* and suppose $\|\pi^*\|_\infty \leqslant 1$. Then*

$$\text{tr}[(X_{\hat{S}}^\top X_{\hat{S}})^{-1}] \leqslant \frac{\|\pi^*\|_0 - p + 1}{k - p + 1} F(\pi^*) \leqslant \frac{\|\pi^*\|_0 - p + 1}{k - p + 1} \cdot \min_{s_i \in \{0,1\}, \sum_i s_i \leqslant k} F(s).$$

It is thus important to upper bound the support size $\|\pi^*\|_0$. With the trivial bound of $\|\pi^*\|_0 \leqslant n$ we recover the $\frac{n-p+1}{k-p+1}$ approximation ratio by applying Figure 4 to $S_0 = [n]$. In order to bound $\|\pi^*\|_0$ away from n, we consider the following assumption imposed on X:

Assumption 1. *Define mapping $\phi : \mathbb{R}^p \to \mathbb{R}^{\frac{p(p+1)}{2}}$ as $\phi(x) = (\xi_{ij} x(i) x(j))_{1 \leqslant i \leqslant j \leqslant p}$, where $x(i)$ denotes the ith coordinate of a p-dimensional vector x and $\xi_{ij} = 1$ if $i = j$ and $\xi_{ij} = 2$ otherwise. Denote $\widetilde{\phi}(x) = (\phi(x), 1) \in \mathbb{R}^{\frac{p(p+1)}{2}+1}$ as the affine version of $\phi(x)$. For any $\frac{p(p+1)}{2} + 1$ distinct rows of X, their mappings under $\widetilde{\phi}$ are linear independent.*

Assumption 1 is essentially a general-position assumption, which assumes that no $\frac{p(p+1)}{2} + 1$ design points in X lie on a degenerate affine subspace after a specific quadratic mapping. Like other similar assumptions in the literature (Tibshirani, 2013), Assumption 1 is very mild and almost always satisfied in practice, for example, if each row of X is independently sampled from absolutely continuous distributions.

We are now ready to state the main lemma bounding the support size of π^*.

Lemma 8. $\|\pi^*\|_0 \leqslant k + \frac{p(p+1)}{2}$ *if Assumption 1 holds.*

The proof of Lemma 8 is based on an interesting observation into the properties of Karush-Kuhn-Tucker (KKT) conditions of the optimization problem Eq. (2.4), which involves a linear system with $\frac{p(p+1)}{2} + 1$ variables. To contrast the results in Lemma 8 with classical rank/support bounds in SDP and/or linear programming (e.g. the Pataki's bound (Pataki, 1998)), note that the number of constraints in the SDP formulation of Eq. (2.4) (see also Sec. 2.2.1) is linear in n, and hence analysis similar to (Pataki, 1998) would result in an upper bound of $\|\pi^*\|_0$ that scales with n, which is less useful for our analytical purpose.

Let $f(\pi; \lambda, \widetilde{\lambda}, \mu)$ be the Lagrangian muliplier function of the without replacement formulation ($b = 1$) of Eq. (2.4):

$$f(\pi; \lambda, \widetilde{\lambda}, \mu) = \sum_{i=1}^n \pi_i x_i x_i^\top - \sum_{i=1}^n \lambda_i \pi_i + \sum_{i=1}^n \widetilde{\lambda}_i \left(\pi_i - \frac{1}{k} \right) + \mu \left(\sum_{i=1}^n \pi_i - 1 \right).$$

Here $\{\lambda_i\}_{i=1}^n \geqslant 0$, $\{\widetilde{\lambda}_i\}_{i=1}^n \geqslant 0$ and $\mu \geqslant 0$ are Lagrangian multipliers for constraints $\pi_i \geqslant 0$, $\pi_i \leqslant 1$ and $\sum_i \pi_i \leqslant k$, respectively. By KKT condition, $\frac{\partial f}{\partial \pi_i}\big|_{\pi^*} = 0$ and hence

$$-\frac{\partial f}{\partial \pi_i}\bigg|_{\pi^*} = x_i^\top \Sigma_*^{-2} x_i = \widetilde{\lambda}_i - \lambda_i + \mu, \quad i = 1, \cdots, n,$$

where $\Sigma_* = X^\top \mathrm{diag}(\pi^*) X$ is a $p \times p$ positive definite matrix.

Split the index set $[n]$ into three disjoint sets defined as $A = \{i \in [n] : \pi_i^* = 1\}$, $B = \{i \in [n] : 0 < \pi_i^* < 1\}$ and $C = \{i \in [n] : \pi_i^* = 0\}$. Note that $\|\pi^*\|_0 = |A| + |B|$ and $|A| \leqslant k$. Therefore, to upper bound $\|\pi^*\|_0$ it suffices to upper bound $|B|$. By complementary slackness, for all $i \in B$ we have that $\widetilde{\lambda}_i = \lambda_i = 0$; that is,

$$x_i^\top \Sigma_*^{-2} x_i = \langle \phi(x_i), \psi(\Sigma_*^{-2}) \rangle = \mu, \qquad \forall i \in B, \tag{2.21}$$

where $\phi : \mathbb{R}^p \to \mathbb{R}^{p(p+1)/2}$ is the mapping defined in Assumption 1 and $\psi(\cdot)$ takes the upper triangle of a symmetric matrix and vectorizes it into a $\frac{p(p-1)}{2}$-dimensional vector. Assume by way of contradiction that $|B| > p(p+1)/2$ and let $x_1, \cdots, x_{p(p+1)/2+1}$ be arbitrary distinct $\frac{p(p+1)}{2} + 1$ rows whose indices belong to B. Eq. (2.21) can then be cast as a homogenous linear system with $\frac{p(p+1)}{2} + 1$ variables and equations as follows:

$$\begin{bmatrix} \widetilde{\phi}(x_1) \\ \widetilde{\phi}(x_2) \\ \vdots \\ \widetilde{\phi}(x_{p(p+1)/2+1}) \end{bmatrix} \begin{bmatrix} \psi(\Sigma_*^{-2}) \\ -\mu \end{bmatrix} = 0.$$

Under Assumption 1, $\widetilde{\Phi} = [\widetilde{\phi}(x_1); \cdots ; \widetilde{\phi}(x_{p(p+1)/2+1})]^\top$ is invertible and hence both $\psi(\Sigma_*^{-2})$ and μ must be zero. This contradicts the fact that Σ_*^{-2} is positive definite.

2.7.3 Proof of Lemma 1

To prove Lemma 1 we consider the following online matrix game: let $\Delta_{p \times p} = \{A \in \mathbb{R}^{p \times p} : A \succeq 0, \mathrm{tr}(A) = 1\}$ be an *action space* that consists of PSD matrices of unit trace (a.k.a. density matrices). Consider an iterative game for T iterations. At iteration t, the player chooses an action $A_t \in \Delta_{p \times p}$; afterwards, a loss matrix F_t is revealed and the player suffers loss $\langle F_t, A_t \rangle = \mathrm{tr}(F_t^\top A_t)$. The goal of the player is to minimize his/her *regret*:

$$\mathrm{Regret}(\{A_t\}_{t=0}^{T-1}) := \sum_{t=0}^{T-1} \langle F_t, A_t \rangle - \min_{U \in \Delta_{p \times p}} \sum_{t=0}^{T-1} \langle F_t, U \rangle, \tag{2.22}$$

which is the "excess loss" of $\{A_t\}_{t=0}^{T-1}$ compared to the single optimal action $U \in \Delta_{p \times p}$ in "hindsight" (knowing all the loss matrices $\{F_t\}_{t=0}^{T-1}$).

We immediately observe that the second term $\min_{U \in \Delta_{p \times p}} \sum_{t=0}^{T-1} \langle F_t, U \rangle$ in (2.22) is precisely the minimum eigenvalue of $\sum_{t=0}^{T-1} F_t$. Hence, the task of lower bounding $\lambda_{\min}(\sum_{t=0}^{T-1} F_t)$ can be reduced to upper bounding the regret in Eq. (2.22).

A popular strategy to minimize regret for the player is *Follow-The-Regularized-Leader (FTRL)*, also known to be equivalent to *Mirror Descent (MD)* (McMahan, 2011). It specifies strategy A_t for player at each round $t = 0, 1, \ldots, T - 1$ as follows:

$$A_t = \arg \min_{A \in \Delta_{p \times p}} \{\Delta_\psi(A_{t-1}, A) + \alpha \langle F_{t-1}, A \rangle\} . \tag{2.23}$$

Above, $\alpha > 0$ is the learning rate, $\psi : \mathbb{R}^{p \times p} \to \mathbb{R}$ is some differentiable *regularizer* function, and $\Delta_\psi(A, B) = \psi(B) - \psi(A) - \langle \nabla \psi(A), B - A \rangle$ is the so-called Bregman divergence function associated with ψ.

Perhaps the most famous choice of ψ is the matrix entropy $\psi(A) = \langle A, \log A - I \rangle$, and the resulting MD strategy is referred to as matrix multiplicative weight updates (Arora & Kale, 2007). In this paper, to achieve better regret, we adopt the less famous $\ell_{1/2}$-regularizer $\psi(A) = -2\mathrm{tr}(A^{1/2})$ introduced in (Allen-Zhu et al., 2015), and call the resulting MD strategy the $\ell_{1/2}$ strategy.

Remark 1. The vector version of this $\ell_{1/2}$ strategy was first introduced in (Audibert et al., 2011) to obtain optimum regret for combinatorial prediction games. The matrix generalization of this $\ell_{1/2}$ strategy is non-trivial, and leads to optimum regret for problems related to graph sparsification (Allen-Zhu et al., 2015), and faster algorithms for online eigenvector (Allen-Zhu & Li, 2017).

The following proposition gives an alternative closed form for the $\ell_{1/2}$ strategy. Its proof is by careful manipulations of the definition of A_t, and has implicitly appeared in Allen-Zhu et al. (2015). We include its proof for the sake of completeness, later in this section.

Proposition 2 (closed form $\ell_{1/2}$ strategy). *Assume without loss of generality that $A_0 = (c_0 I + \alpha Z_0)^{-2}$ for some $c_0 \in \mathbb{R}$ and symmetric matrix Z_0 such that $c_0 I + \alpha Z_0 > 0$. Then,*

$$A_t = \left(c_t I + \alpha Z_0 + \alpha \sum_{\ell=0}^{t-1} F_\ell \right)^{-2}, \qquad t = 1, 2, \ldots, \tag{2.24}$$

where $c_t \in \mathbb{R}$ is the unique constant that ensures $c_t I + \alpha Z_0 + \alpha \sum_{\ell=0}^{t-1} F_\ell > 0$ and $\mathrm{tr}(A_t) = 1$.

At a high level, if $Z_0 = 0$ were the zero matrix, then $A_0 = \frac{I}{\sqrt{p}}$ would be a multiple of identity. This corresponds to the standard way to initialize the player's strategy in online learning, and was used in Allen-Zhu et al. (2015). In this paper, we need this more general Z_0 to support our proposed swapping algorithm.

If each loss matrix F_t can be rank-2 decomposed as $F_t = u_t u_t^\top - v_t v_t^\top$, then we prove the following lemma which upper bounds the total regret of the $\ell_{1/2}$ strategy:

Lemma 9 (main regret lemma). *Suppose $F_t = u_t u_t^\top - v_t v_t^\top$ for vectors $u_t, v_t \in \mathbb{R}^p$, and $A_0, \ldots, A_{T-1} \in \Delta_{p \times p}$ are defined according to the $\ell_{1/2}$ strategy with some learning rate $\alpha > 0$. Then, as long as $\alpha \langle A_t^{1/2}, v_t v_t^\top \rangle < 1/2$ for all t, we have for any $U \in \Delta_{p \times p}$,*

$$-\sum_{t=0}^{T-1} \langle F_t, U \rangle \leqslant \sum_{t=0}^{T-1} \left(-\frac{\langle A_t, u_t u_t^\top \rangle}{1 + 2\alpha \langle A_t^{1/2}, u_t u_t^\top \rangle} + \frac{\langle A_t, v_t v_t^\top \rangle}{1 - 2\alpha \langle A_t^{1/2}, v_t v_t^\top \rangle} \right) + \frac{\Delta_\psi(A_0, U)}{\alpha}. \tag{2.25}$$

$$= \sum_{t=0}^{T-1} -\phi_+(u_t; Z_t) + \phi_-(v_t; Z_t) + \frac{\Delta_\psi(A_0, U)}{\alpha}. \tag{2.26}$$

Remark 2. To better see why Lemma 9 is a bound on regret (2.22), we rearrange the two sides:

$$\sum_{t=0}^{T-1} \langle F_t, A_t - U \rangle \leqslant 2\alpha \sum_{t=0}^{T-1} \left(\frac{\langle A_t, u_t u_t^\top \rangle \cdot \langle A_t^{1/2}, u_t u_t^\top \rangle}{1 + 2\alpha \langle A_t^{1/2}, u_t u_t^\top \rangle} + \frac{\langle A_t, v_t v_t^\top \rangle \cdot \langle A_t^{1/2}, v_t v_t^\top \rangle}{1 - 2\alpha \langle A_t^{1/2}, v_t v_t^\top \rangle} \right) + \frac{\Delta_\psi(A_0, U)}{\alpha}.$$

28

Our proof of Lemma 9 involves a non-classical regret analysis designed for the matrix $\ell_{1/2}$ strategy. It is based on the closed-form expressions in Eq. (2.24). Note that a variant of Lemma 9, but only for matrices $F_t = u_t u_t^\top$ (thus of rank 1) was originally presented in Theorem 3.2 of (Allen-Zhu et al., 2015). The involvement of the extra $-v_t v_t^\top$ components is, however, a non-trivial extension and brings in extra technical difficulties.

Proof of Lemma 9. To prove this lemma we consider an equivalent "2-step" description of the mirror descent procedure:

$$\widetilde{A}_t = \arg\min_{A \geq 0} \{\Delta_\psi(A_{t-1}, A) + \alpha\langle F_{t-1}, A\rangle\}; \qquad A_t = \arg\min_{A \in \Delta_{p \times p}} \Delta_\psi(\widetilde{A}_t, A).$$

By the so-called "tweaked analysis" of mirror descent (Rakhlin, 2009; Zinkevich, 2003), the matrix A_t defined above is identical to its original definition of $\arg\min_{A \in \Delta_{p \times p}} \{\Delta_\psi(A_{t-1}, A) + \alpha\langle F_{t-1}, A\rangle\}$. This can also be verified by writing $\widetilde{A}_t$ explicitly using the following claim, and verifying that A_t (in its closed form by Proposition 2) is indeed a minimizer of $\Delta_\psi(\widetilde{A}_t, A)$ over $A \in \Delta_{p \times p}$ by taking its gradient.

Proposition 3. *We have $\widetilde{A}_t = (A_{t-1}^{-1/2} + \alpha F_{t-1})^{-2}$.*

The proof of Proposition 3 is given later. Since $\nabla\psi(\widetilde{A}_t) - \nabla\psi(A_{t-1}) + \alpha F_{t-1} = 0$ as shown in the proof above, we have (by defining $\widetilde{A}_0 = A_0$)

$$
\begin{aligned}
\langle \alpha F_{t-1}, A_{t-1} - U\rangle &= \langle \nabla\psi(A_{t-1}) - \nabla\psi(\widetilde{A}_t), A_{t-1} - U\rangle \\
&= \Delta_\psi(A_{t-1}, U) - \Delta_\psi(\widetilde{A}_t, U) + \Delta_\psi(\widetilde{A}_t, A_{t-1}) \\
&\leqslant \Delta_\psi(\widetilde{A}_{t-1}, U) - \Delta_\psi(\widetilde{A}_t, U) + \Delta_\psi(\widetilde{A}_t, A_{t-1}).
\end{aligned}
\tag{2.27}
$$

Above, the second equality and the last inequality follow from the "three-point" equality and the generalized Pythagorean theorem of Bregman divergence (see for example, Lemma 2.1 of Allen-Zhu et al. (2015)). Expanding $\Delta_\psi(\widetilde{A}_t, A_{t-1})$ by its definition gives

$$
\begin{aligned}
\Delta_\psi(\widetilde{A}_t, A_{t-1}) &= \psi(A_{t-1}) - \psi(\widetilde{A}_t) - \langle \nabla\psi(\widetilde{A}_t), A_{t-1} - \widetilde{A}_t\rangle \\
&= -2\mathrm{tr}(A_{t-1}^{1/2}) + 2\mathrm{tr}(\widetilde{A}_t^{1/2}) + \langle \widetilde{A}_t^{-1/2}, A_{t-1} - \widetilde{A}_t\rangle \\
&= \langle \widetilde{A}_t^{-1/2}, A_{t-1}\rangle + \mathrm{tr}(\widetilde{A}_t^{1/2}) - 2\mathrm{tr}(A_{t-1}^{1/2}) \\
&= \langle A_{t-1}^{-1/2} + \alpha F_{t-1}, A_{t-1}\rangle + \mathrm{tr}(\widetilde{A}_t^{1/2}) - 2\mathrm{tr}(A_{t-1}^{1/2}) \\
&= \alpha\langle F_{t-1}, A_{t-1}\rangle + \mathrm{tr}(\widetilde{A}_t^{1/2}) - \mathrm{tr}(A_{t-1}^{1/2}).
\end{aligned}
\tag{2.28}
$$

Combining Eqs. (2.27, 2.28) and telescoping from $t = 1$ to $t = T$ we obtain

$$
\begin{aligned}
-\alpha\sum_{t=0}^{T-1} \langle F_t, U\rangle &\leqslant \Delta_\psi(A_0, U) - \Delta_\psi(\widetilde{A}_T, U) + \sum_{t=0}^{T-1} \mathrm{tr}(\widetilde{A}_{t+1}^{1/2}) - \mathrm{tr}(A_t^{1/2}) \\
&\leqslant \Delta_\psi(A_0, U) + \sum_{t=0}^{T-1} \mathrm{tr}(\widetilde{A}_{t+1}^{1/2}) - \mathrm{tr}(A_t^{1/2}),
\end{aligned}
\tag{2.29}
$$

29

where the second inequality holds because Bregman divergence $\Delta_\psi(\widetilde{A}_T, U)$ is always non-negative.

It remains to upper bound the "consecutive difference" $\mathrm{tr}(\widetilde{A}_{t+1}^{1/2}) - \mathrm{tr}(A_t^{1/2})$. Let $P_t = \sqrt{\alpha}[u_t \ v_t] \in \mathbb{R}^{p\times 2}$ and $J = \mathrm{diag}(1, -1) \in \mathbb{R}^{2\times 2}$, so we have $\alpha F_t = P_t J P_t^\top$. By the definition of $\widetilde{A}_{t+1}^{1/2}$ and the Woodbury formula[5],

$$\mathrm{tr}(\widetilde{A}_{t+1}^{1/2}) = \mathrm{tr}\left[(A_t^{-1/2} + P_t J P_t^\top)^{-1}\right] = \mathrm{tr}\left[A_t^{1/2} - A_t^{1/2}P_t(J + P_t^\top A_t^{1/2}P_t)^{-1}P_t^\top A_t^{1/2}\right]. \quad (2.30)$$

It is crucial to spectrally lower bound the core 2×2 matrix $(J + P_t^\top A_t^{1/2}P_t)^{-1/2}$ in the middle of Eq. (2.30). For this purpose, we claim that

Proposition 4. *Suppose $P_t^\top A_t^{1/2}P_t = [b \ d; d \ c] \in \mathbb{R}^{2\times 2}$ and $2\alpha\langle A_t^{1/2}, v_t v_t^\top\rangle < 1$. Then*

$$(J + P_t^\top A_t^{1/2}P_t)^{-1} = \left(J + \begin{bmatrix} b & d \\ d & c \end{bmatrix}\right)^{-1} \succeq \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1}.$$

Proposition 4 is trivially true if $J \succeq 0$, but becomes less obvious when J has negative eigenvalues. In fact, Proposition 4 is not universally true for any matrices of the form $PAP^\top$, and specifically requires the condition that $2\alpha\langle A_t^{1/2}, v_t v_t^\top\rangle < 1$. We defer the proof of Proposition 4 later.

With Proposition 4, the consecutive gap $\mathrm{tr}(\widetilde{A}_{t+1}^{1/2}) - \mathrm{tr}(A_t^{1/2})$ can be bounded as

$$\mathrm{tr}(\widetilde{A}_{t+1}^{1/2}) - \mathrm{tr}(A_t^{1/2}) = -\mathrm{tr}\left[-A_t^{1/2}P_t(J + P_t^\top A_t^{1/2}P_t)^{-1}P_t^\top A_t^{1/2}\right]$$

$$\leqslant -\mathrm{tr}\left[-A_t^{1/2}P_t\left(J + \begin{bmatrix} 2\alpha u_t^\top A_t^{1/2}u_t & 0 \\ 0 & 2\alpha v_t^\top A_t^{1/2}v_t \end{bmatrix}\right)^{-1}P_t^\top A_t^{1/2}\right]$$

$$= -\frac{\alpha\langle A_t, u_t u_t^\top\rangle}{1 + 2\alpha\langle A_t^{1/2}, u_t u_t^\top\rangle} + \frac{\alpha\langle A_t, v_t v_t^\top\rangle}{1 - 2\alpha\langle A_t^{1/2}, v_t v_t^\top\rangle}. \quad (2.31)$$

Plugging Eq. (2.31) into Eq. (2.29) we complete the proof of Lemma 9. $\qquad\square$

To establish Lemma 1 from Lemma 9, we also need the following lemma to bound the Bregman divergence term $\Delta_\psi(A_0, U)$:

Lemma 10. *Suppose $A_0 = (c_0 I + \alpha Z_0)^{-2}$ as in Proposition 2, then for any $U \in \Delta_{p\times p}$:*

$$\Delta_\psi(A_0, U) \leqslant 2\sqrt{p} + \alpha\langle Z_0, U\rangle .$$

Proof. By definition of Δ_ψ, ψ and A_0, we have

$$\Delta_\psi(A_0, U) = \langle A_0^{-1/2}, U\rangle + \mathrm{tr}(A_0^{1/2}) - 2\mathrm{tr}(U^{1/2}) \leqslant \langle c_0 I + \alpha Z_0, U\rangle + \sqrt{p},$$

[5]$(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1}$, provided that all inverses exist.

where the last inequality holds because $\mathrm{tr}(U^{1/2}) \geqslant 0$ and $\mathrm{tr}(A_0^{1/2}) \leqslant \sqrt{p \cdot \mathrm{tr}(A_0)} = \sqrt{p}$. Note also that $\langle I, U \rangle = \mathrm{tr}(U) = 1$. Therefore,

$$\Delta_\psi(A_0, U) \leqslant \alpha\langle Z_0, U\rangle + c_0 + \sqrt{p}.$$

Because $\mathrm{tr}(A_0) = 1$, the constant c_0 (if positive) must be upper bounded by $\sqrt{p}$ because otherwise $\mathrm{tr}(A_0) \leqslant \mathrm{tr}((c_0 I)^{-2}) = p \cdot c_0^{-2} < 1$. Therefore, it is proved that $\Delta_\psi(A_0, U) \leqslant \alpha\langle Z_0, U\rangle + 2\sqrt{p}$. $\qquad\square$

Combining Lemmas 9 and 10, we complete the proof of Lemma 1.

In the rest of this section we state proofs of technical propositions that are omitted above.

Proof of Proposition 2 We first show that for any symmetric matrix $Z \in \mathbb{R}^{p \times p}$, there exists unique $c \in \mathbb{R}$ such that $\alpha Z + cI > 0$ and $\mathrm{tr}[(\alpha Z + cI)^{-2}] = 1$. By simple asymptotic analysis, $\lim_{c \to (-\alpha\lambda_{\min}(Z))^+} \mathrm{tr}[(\alpha Z + cI)^{-2}] = +\infty$ and $\lim_{c \to +\infty} \mathrm{tr}[(\alpha Z + cI)^{-2}] = 0$. Because $\mathrm{tr}[(\alpha Z + cI)^{-2}]$ is a continuous and strictly decreasing function in c on the open interval $(-\alpha\lambda_{\min}(Z), +\infty)$, we conclude that there must exist a unique c such that $\mathrm{tr}[(\alpha Z + cI)^{-2}] = 1$. The range of c also ensures that $\alpha Z + cI > 0$.

We now use induction to prove this proposition. For $t = 0$ the proposition is obviously correct. We shall now assume that the proposition holds true for A_{t-1} (i.e., $A_{t-1} = (c_{t-1}I + \alpha Z_0 + \sum_{\ell=0}^{t-2} F_\ell)^{-2}$) for some $t \geqslant 1$, and try to prove the same for A_t.

The KKT condition of the optimization problem and the gradients of the Bregman divergence Δ_ψ yields

$$\nabla\psi(A_t) - \nabla\psi(A_{t-1}) + \alpha F_{t-1} - d_t I = 0, \qquad (2.32)$$

where the $d_t I$ term arises from the Lagrangian multiplier and $d_t \in \mathbb{R}$ is the unique number that makes $-\nabla\psi(A_{t-1}) + \alpha F_t - d_t I \preceq 0$ (because $\nabla\psi(A_t) \succeq 0$) and $\mathrm{tr}(A_t) = \mathrm{tr}((\nabla\psi)^{-1}(\nabla\psi(A_{t-1}) + d_t I - \alpha F_{t-1})) = 1$. Re-organizing terms in Eq. (3.110) and invoking the induction hypothesis we have

$$A_t = (\nabla\psi)^{-1}\left(\nabla\psi(A_{t-1}) + d_t I - \alpha F_{t-1}\right)$$
$$= (\nabla\psi)^{-1}\left(-c_{t-1}I - \alpha Z_0 - \alpha\sum_{\ell=0}^{t-1} F_\ell + d_t I\right).$$

Because d_t is the unique number that ensures $A_t \succeq 0$ and $\mathrm{tr}(A_t) = 1$, and $Z_0 + \sum_{\ell=0}^{t-1} F_\ell \succeq 0$, it must hold that $-c_{t+1} + d_t = c_t$. Subsequently, $\nabla\psi(A_t) = -(A_t^{-1/2}) = -c_t I - \alpha Z - \alpha\sum_{\ell=0}^{t-1} F_\ell$. The claim is thus proved by raising both sides of the identity to the power of -2.

Proof of Proposition 3 We first show $(A_{t-1}^{-1/2} + \alpha F_{t-1})^{-2}$ is well defined. By assumption $\alpha\langle A_{t-1}^{1/2}, v_{t-1}v_{t-1}^\top\rangle < 1$, and hence $-\alpha F_{t-1} \preceq \alpha v_{t-1}v_{t-1}^\top < A_{t-1}^{-1/2}$. This is because for any matrices $A > 0$ and $B \succeq 0$, we have $\langle A, B \rangle = \mathrm{tr}(A^\top B) < 1 \implies A^\top B \preceq I \implies B \preceq A^{-1}$. Consequently, we have $A_{t-1}^{-1/2} + \alpha F_{t-1} > 0$ and its inverse exists.

31

Next, to prove $\widetilde{A}_t = (A_{t-1}^{-1/2} + \alpha F_{t-1})^{-2}$ is a minimizer of the convex function $\{\Delta_\psi(A_{t-1}, A) + \alpha\langle F_{t-1}, A\rangle\}$ over all positive semi-definite matrices A, we show its gradient evaluated at $\widetilde{A}_t$ is zero.[6] Indeed,

$$
\begin{aligned}
\nabla\left(\Delta_\psi(A_{t-1}, \widetilde{A}_t) + \alpha\langle F_{t-1}, \widetilde{A}_t\rangle\right) &= \nabla\psi(\widetilde{A}_t) - \nabla\psi(A_{t-1}) + \alpha F_{t-1} \\
&= -\widetilde{A}_t^{-1/2} + A_{t-1}^{-1/2} + \alpha F_{t-1} = 0 \ .
\end{aligned}
$$

Proof of Proposition 4 Define $R = \begin{bmatrix} b & -d \\ -d & c \end{bmatrix}$. Because $P_t^\top A_t^{1/2} P_t = \begin{bmatrix} b & d \\ d & c \end{bmatrix}$ is positive semi-definite, we conclude that R is also positive semi-definite and hence can be written as $R = QQ^\top$. To prove Proposition 4, we only need to establish the positive semi-definiteness of the following difference matrix:

$$
\begin{aligned}
&\left(J + \begin{bmatrix} b & d \\ d & c \end{bmatrix}\right)^{-1} - \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1} \\
&= \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix} - \begin{bmatrix} b & -d \\ -d & c \end{bmatrix}\right)^{-1} - \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1} \\
&= \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1} Q \left(I - Q^\top \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1} Q\right)^{-1} Q^\top \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1} .
\end{aligned}
$$

Here in the last equality we again use the Woodbury matrix identity. It is clear that to prove the positive semi-definiteness right-hand side of the above equality, It suffices to show $Q^\top(J + \mathrm{diag}(2b, 2c))^{-1}Q \prec I$. By standard matrix analysis and the fact that $J = \mathrm{diag}(1, -1)$,

$$
\begin{aligned}
Q^\top \left(J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}\right)^{-1} Q &= Q^\top \begin{bmatrix} (1 + 2b)^{-1} & 0 \\ 0 & -(1 - 2c)^{-1} \end{bmatrix} Q \\
&\overset{(a)}{\preceq} Q^\top \begin{bmatrix} (1 + 2b)^{-1} & 0 \\ 0 & 0 \end{bmatrix} Q \preceq \frac{\|QQ^\top\|_{\mathrm{op}}}{1 + 2b} \cdot I \\
&\overset{(b)}{\preceq} \frac{\max\{2b, 2c\}}{1 + 2b} \cdot I \overset{(c)}{\prec} I.
\end{aligned}
$$

Some steps in the above derivation require additional explanation. In (a), we use the fact that $2c = 2\alpha\langle A_t^{1/2}, v_t v_t^\top\rangle < 1$, and hence $(1 - 2c)^{-1} > 0$; in (b), we use the fact that $QQ^\top = \begin{bmatrix} b & -d \\ -d & c \end{bmatrix} \preceq \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix}$; finally, (c) holds because $b = 2\alpha\langle A_t^{1/2}, u_t u_t^\top\rangle \geqslant 0$, $\frac{2b}{1+2b} < 1$ and $2c < 1$. The proof of Proposition 4 is thus completed.

2.7.4 Proof of Theorem 3

The following lemma is the key result in our proof of Theorem 3:

---

[6]The convexity of this objective follows from Lieb's concavity theorem (Bhatia, 1997; Lieb, 1973), and is already a known fact in matrix regret minimization literatures (Allen-Zhu et al., 2015).

Lemma 11 (main averaging lemma). *For every $\varepsilon > 0$ and subset $\Lambda \subseteq [n]$ of cardinality k, suppose $\lambda_{\min}(\sum_{i\in\Lambda} x_i x_i^\top) \leqslant 1 - 15\varepsilon$ and $A = (cI + \alpha \sum_{i\in\Lambda} x_i x_i^\top)^{-2}$, where $c \in \mathbb{R}$ is the unique number such that $A \succeq 0$ and $\mathrm{tr}(A) = 1$. Then, the following statements are true:*

$$\nu := \min_{i\in\Lambda, 2\alpha\langle A^{1/2}, x_i x_i^\top\rangle < 1} \frac{\langle A, x_i x_i^\top\rangle}{1 - 2\alpha\langle A^{1/2}, x_i x_i^\top\rangle} \leqslant \frac{1-\varepsilon}{k}; \tag{2.33}$$

$$\max_{j\in[n]\backslash\Lambda} \frac{\langle A, x_j x_j^\top\rangle}{1 + 2\alpha\langle A^{1/2}, x_j x_j^\top\rangle} \geqslant \nu + \frac{\varepsilon}{k}. \tag{2.34}$$

Furthermore, if $\alpha = \sqrt{p}/\varepsilon$ and $k \geqslant 6p/\varepsilon^2$ for some $\varepsilon \in (0, 1/15)$, then there always exists $i \in \Lambda$ such that $2\alpha\langle A^{1/2}, x_i x_i^\top\rangle < 1$.

In other words, Lemma 11 suggests that, as long as $\lambda_{\min}(\sum_{i\in\Lambda} x_i x_i^\top) \leqslant 1 - 15\varepsilon$, we can simply choose i_t to be the index $i \in \Lambda_t$ which minimizes $\frac{\langle A_t, x_i x_i^\top\rangle}{1 - 2\alpha\langle A_t^{1/2}, x_i x_i^\top\rangle}$, and j_t to be the index $j \notin \Lambda_t$ which maximizes $\frac{\langle A_t, x_j x_j^\top\rangle}{1 + 2\alpha\langle A_t^{1/2}, x_j x_j^\top\rangle}$. Eqs. (2.33) and (2.34) together imply that

$$\left(-\frac{\langle A_t, x_{j_t} x_{j_t}^\top\rangle}{1 + 2\alpha\langle A_t^{1/2}, x_{j_t} x_{j_t}^\top\rangle} + \frac{\langle A_t, x_{i_t} x_{i_t}^\top\rangle}{1 - 2\alpha\langle A_t^{1/2}, x_{i_t} x_{i_t}^\top\rangle} \right) \leqslant -\frac{\varepsilon}{k}. \tag{2.35}$$

In sum, either there exists some index $t = 0, 1, \ldots, T - 1$ such that $\lambda_{\min}(\sum_{i\in\Lambda_t} x_i x_i^\top) > 1 - 15\varepsilon$ is satisfied, or we can always find pairs (i_t, j_t) satisfying Eq. (2.35), which implies

$$-\lambda_{\min}\left(\sum_{j\in\Lambda_T} x_j x_j^\top \right) \leqslant \sum_{t=0}^{T-1} -\frac{\varepsilon}{k} + \frac{2\sqrt{p}}{\alpha} = -\frac{T\varepsilon}{k} + 2\varepsilon.$$

Here in the last inequality we apply the choice $\alpha = \sqrt{p}/\varepsilon$. In other words, as long as $T \geqslant k/\varepsilon$, it must satisfy $\lambda_{\min}(\sum_{i\in\Lambda_T} x_i x_i^\top) \geqslant 1 - 2\varepsilon$, and subsequently by the reciprocal multiplicity property we complete the proof of Theorem 3.

Proof of Lemma 11   We first state a technical proposition as follows:

Proposition 5. *Suppose $Z \succeq 0$ is a p-dimensional PSD matrix with $\lambda_{\min}(Z) \leqslant 1$. Let $A = (\alpha Z + cI)^{-2}$, where $c \in \mathbb{R}$ is the unique real number such that $A \succeq 0$ and $\mathrm{tr}(A) = 1$. Then*
1. *$\alpha\langle A^{1/2}, Z\rangle \leqslant p + \alpha\sqrt{p}$;*
2. *$\langle A, Z\rangle \leqslant \sqrt{n}/\alpha + \lambda_{\min}(Z)$.*

Proof. For any orthogonal matrix U, the transform $Z \mapsto UZU^\top$ leads to $X \mapsto UXU^\top$ and $X^{1/2} \mapsto UX^{1/2}U^\top$; thus both inner products are invariant to orthogonal transform of Z. Therefore, we may assume without loss of generality that $Z = \mathrm{diag}(\sigma_1, \ldots, \sigma_p)$ for $\lambda_1 \geqslant \cdots \lambda_p \geqslant 0$, because $Z \succeq 0$. Subsequently,

$$\alpha\langle A^{1/2}, Z\rangle = \sum_{i=1}^{p} \frac{\alpha\lambda_i}{\alpha\lambda_i + c} = p - c \cdot \sum_{i=1}^{p} \frac{1}{\alpha\lambda_i + c}.$$

33

If $c \geqslant 0$, then $\alpha\langle A^{1/2}, Z\rangle \leqslant p$ and the first property is clearly true. For the case of $c < 0$, note that c must be strictly larger than $-\alpha\lambda_p$, as we established in Proposition 2. Subsequently, by the Cauchy-Schwarz inequality,

$$\alpha\langle A^{1/2}, Z\rangle = p - c \cdot \sum_{i=1}^{p} \frac{1}{\alpha\lambda_i + c} \leqslant p - c \cdot \sqrt{p} \cdot \sqrt{\sum_{i=1}^{p} \frac{1}{(\alpha\lambda_i + c)^2}}.$$

Because $\lambda_p = \lambda_{\min}(Z) \leqslant 1$ and $\mathrm{tr}(A) = \mathrm{tr}[(\alpha Z + cI)^{-2}] = 1$, we have that $c \geqslant -\alpha$ and $\sqrt{\sum_{i=1}^{p}(\alpha\lambda_i + c)^{-2}} = 1$. Therefore, $\alpha\langle A^{1/2}, Z\rangle \leqslant p + \alpha\sqrt{p}$, which establishes the first property in Proposition 5.

We next turn to the second property. Using similar analysis, we have

$$\alpha\langle Z, A\rangle = \sum_{i=1}^{p} \frac{\alpha\lambda_i}{(\alpha\lambda_i + c)^2} = \sum_{i=1}^{p} \frac{1}{\alpha\lambda_i + c} - c \cdot \sum_{i=1}^{p} \frac{1}{(\alpha\lambda_i + c)^2}$$

$$\leqslant \sqrt{p} \cdot \sqrt{\sum_{i=1}^{p} \frac{1}{(\alpha\lambda_i + c)^2}} - c \cdot \sum_{i=1}^{p} \frac{1}{(\alpha\lambda_i + c)^2} \leqslant \sqrt{p} - c.$$

Property 2 is then proved by noting that $c > -\lambda_{\min}(Z)$. $\qquad\square$

Back to the proof of Lemma 11, we first show the existence of (at least one) $i \in \Lambda$ such that $2\alpha\langle A^{1/2}, x_i x_i^\top\rangle < 1$. Define $Z = \sum_{i \in \Lambda} x_i x_i^\top$, and by definition $A = (cI + \alpha\sum_{i \in \Lambda} x_i x_i^\top)^{-2} = (\alpha Z + cI)^{-2}$. Assume by way of contradiction that such i does not exist. We then have

$$\sum_{i \in \Lambda} 2\alpha\langle A^{1/2}, x_i x_i^\top\rangle = 2\alpha\langle A^{1/2}, Z\rangle \geqslant |\Lambda| = k. \tag{2.36}$$

On the other hand, because $Z \succeq 0$ and $\lambda_{\min}(Z) < 1$, invoking Proposition 5 we get

$$2\alpha\langle A^{1/2}, Z\rangle \leqslant 2p + 2\alpha\sqrt{p}$$

which contradicts Eq. (2.36) provided that $\alpha = \sqrt{p}/\varepsilon$ and $k > 4p/\varepsilon$. Thus, there must exist $i \in \Lambda$ such that $2\alpha\langle A^{1/2}, x_i x_i^\top\rangle < 1$. In fact, a stronger result $\sum_{i \in \Lambda}(1 - 2\alpha\langle A^{1/2}, x_i x_i^\top\rangle) > 0$ can be established following the above arguments.

We next proceed to prove Eq. (2.33). By definition of ν, we must have that $(1 - 2\alpha\langle A^{1/2}, x_i x_i^\top\rangle)\nu \leqslant \langle A, x_i x_i^\top\rangle$ for all $i \in \Lambda$, because if $2\alpha\langle A^{1/2}, x_i x_i^\top\rangle \geqslant 1$ the left-hand side is non-positive while the right-hand side is always non-negative, thanks to the positive semi-definiteness of A. Subsequently,

$$\nu \leqslant \frac{\sum_{i \in \Lambda}\langle A, x_i x_i^\top\rangle}{\sum_{i \in \Lambda}(1 - 2\alpha\langle A^{1/2}, x_i x_i^\top\rangle)} \leqslant \frac{\sqrt{p}/\alpha + \lambda_{\min}(\sum_{i \in \Lambda} x_i x_i^\top)}{k - 2p - 2\alpha\sqrt{p}} \leqslant \frac{\varepsilon + 1 - 15\varepsilon}{k(1 - 13\varepsilon)} \leqslant \frac{1 - \varepsilon}{k},$$

where the first inequality holds because the denominator is strictly positive, and in the second inequality we invoke Proposition 5. We have thus proved that $\nu \leqslant (1 - \varepsilon)/k$.

Finally we prove Eq. (2.34). Define $t = \nu + \varepsilon/k \leqslant 1/k$. To prove Eq. (2.34) it suffices to show that $\sum_{j \in [n] \setminus \Lambda} \pi_j \langle A, x_j x_j^\top \rangle \geqslant t \sum_{j \in [n] \setminus \Lambda} \pi_j (1 + 2\alpha \langle A^{1/2}, x_j x_j^\top \rangle)$, because $\pi_j \geqslant 0$ for all j. Recall that $\sum_{j=1}^n \pi_j = k$, $\sum_{j=1}^n \pi_j x_j x_j^\top = I$. We then have

$$
\begin{aligned}
\sum_{j \in [n] \setminus \Lambda} \pi_j (1 + 2\alpha \langle A^{1/2}, x_j x_j^\top \rangle) &= \left(k - \sum_{j \in \Lambda} \pi_j \right) + 2\alpha \cdot \sum_{j \in [n] \setminus \Lambda} \pi_j \langle A^{1/2}, x_j x_j^\top \rangle \\
&\leqslant \left(k - \sum_{j \in \Lambda} \pi_j \right) + 2\alpha \cdot \sum_{j=1}^n \pi_j \langle A^{1/2}, x_j x_j^\top \rangle \\
&= k - \sum_{j \in \Lambda} \pi_j + 2\alpha \langle I, A^{1/2} \rangle = k - \sum_{j \in \Lambda} \pi_j + 2\alpha \mathrm{tr}(A^{1/2}).
\end{aligned}
$$

Similarly,

$$
\sum_{j \in [n] \setminus \Lambda} \pi_j \langle A, x_j x_j^\top \rangle = \left\langle I - \sum_{j \in \Lambda} \pi_j x_i x_i^\top, X \right\rangle = \mathrm{tr}(A) - \sum_{j \in \Lambda} \pi_j \langle A, x_j x_j^\top \rangle.
$$

Note that for any $p \times p$ positive semi-definite matrix $Z \succeq 0$, $\mathrm{tr}(Z^{1/2}) \leqslant \sqrt{p \cdot \mathrm{tr}(Z)}$ thanks to the Hölder's inequality [7] applied to the non-negative spectrum of $Z^{1/2}$, and that $\mathrm{tr}(A) = 1$ by definition. Subsequently,

$$
\begin{aligned}
&\sum_{j \in [n] \setminus \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \cdot \sum_{j \in [n] \setminus \Lambda} \pi_j (1 + 2\alpha \langle A^{1/2}, x_j x_j^\top \rangle) \\
&\geqslant \mathrm{tr}(A) - \sum_{j \in \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \left(k - \sum_{j \in \Lambda} \pi_j \right) - 2\alpha t \cdot \mathrm{tr}(A^{1/2}) \\
&\geqslant 1 - \sum_{j \in \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \left(k - \sum_{j \in \Lambda} \pi_j \right) - 2\alpha t \sqrt{p} \\
&= 1 - tk - 2t\alpha \sqrt{p} - \sum_{j \in \Lambda} \pi_j (\langle A, x_j x_j^\top \rangle - t) \\
&\geqslant 1 - tk - 2t\alpha \sqrt{p} - \sum_{j \in \Lambda} \max\{\langle A, x_j x_j^\top \rangle - t, 0\} \quad (2.37) \\
&\geqslant 1 - tk - 2t\alpha \sqrt{p} - \sum_{j \in \Lambda} (\langle A, x_j x_j^\top \rangle - t) - \sum_{j \in \Lambda} \max\{(t - \langle A, x_j x_j^\top \rangle), 0\} \\
&\geqslant 1 - 2t\alpha \sqrt{p} - \sqrt{p}/\alpha - \lambda_{\min} \left(\sum_{j \in \Lambda} x_j x_j^\top \right) - \sum_{j \in \Lambda} \max\{(t - \langle A, x_j x_j^\top \rangle), 0\}. \quad (2.38)
\end{aligned}
$$

Here Eq. (2.37) holds because $\pi_j \leqslant 1$ for all j, and in Eq. (2.38) we apply the $\sum_{j \in \Lambda} \langle A, x_j x_j^\top \rangle \leqslant \sqrt{p}/\alpha + \lambda_{\min}(\sum_{j \in \Lambda} x_j x_j^\top)$ property as established in Proposition 5. By the conditions that $\alpha =$

---

[7] $|x_1| + \cdots + |x_d| \leqslant \sqrt{d} \cdot \sqrt{x_1^2 + \cdots + x_d^2}$ for any sequences of d real numbers $x_1, \ldots, x_n$.

$\sqrt{p}/\varepsilon$, $t \leqslant 1/k$ and $\lambda_{\min}(\sum_{j\in\Lambda} x_j x_j^\top) \leqslant 1 - 3\varepsilon$, the right-hand side of Eq. (2.38) can be lower bounded by the simplified form of

$$\sum_{j\in[n]\backslash\Lambda} \pi_j \left\{ \langle A, x_j x_j^\top \rangle - t(1 + 2\alpha\langle A^{1/2}, x_j x_j^\top \rangle) \right\} \geqslant 2\varepsilon - \frac{2p}{\varepsilon k} - \sum_{j\in\Lambda} \max\{t - \langle A, x_j x_j^\top \rangle, 0\}. \quad (2.39)$$

Furthermore, because $(1 - 2\alpha\langle A^{1/2}, x_i x_i^\top \rangle)\nu \leqslant \langle A, x_i x_i^\top \rangle$ for all $i \in \Lambda$, invoking Proposition 5 we have

$$\sum_{i\in\Lambda'} (\nu - \langle A, x_i x_i^\top \rangle) \leqslant \sum_{i\in\Lambda'} 2\nu\alpha\langle A^{1/2}, x_j x_j^\top \rangle \leqslant 2\nu(p + \alpha\sqrt{p}) \qquad \text{for all } \Lambda' \subseteq \Lambda.$$

Consider $\Lambda = \{i \in \Lambda : t - \langle A, x_i x_i^\top \rangle \geqslant 0\}$. We then have

$$\sum_{j\in\Lambda} \max\{t - \langle A, x_j x_j^\top \rangle, 0\} = \sum_{j\in\Lambda} (t - \langle A, x_j x_j^\top \rangle) = (t - \nu)|\Lambda| + \sum_{j\in\Lambda} (\nu - \langle A, x_j x_j^\top \rangle)$$

$$\leqslant \varepsilon + \frac{4p/\varepsilon}{k}, \quad (2.40)$$

where the last inequality holds because $t - \nu = \varepsilon/k \geqslant 0$, $|\Lambda| \leqslant k$, $\nu \leqslant 1/k$ and $\alpha = \sqrt{p}/\varepsilon$. Combining Eqs. (2.39) and (2.40) we arrive at

$$\sum_{j\in[n]\backslash\Lambda} \pi_j \left\{ \langle A, x_j x_j^\top \rangle - t(1 + 2\alpha\langle A^{1/2}, x_j x_j^\top \rangle) \right\} \geqslant \varepsilon - \frac{6p}{\varepsilon k}.$$

If $k \geqslant 6p/\varepsilon^2$, the right-hand side of the above inequality is non-negative, which is to be demonstrated.

2.7.5 Proof of Lemma 2

Without loss of generality assume $k_i > 0$ for all i, because for those design points with $k_i = 0$ no information is gained and therefore these points can be excluded from the analysis. Let $w_i = 4^{k_i+1}$ be the weight of design point x_i and define $W := \text{diag}(w_1, \cdots, w_n)$. The weighted OLS estimator $\widehat{\beta}_k$ then admits a closed-form expression

$$\widehat{\beta}_k = (X^\top W X)^{-1} X^\top W \widetilde{y}, \quad (2.41)$$

where $\widetilde{y} = (\widetilde{y}_1, \cdots, \widetilde{y}_n) \in \mathbb{R}^n$. Define $\varepsilon := \widetilde{y} - y$. Using the linear model that $y = X\beta_0$, we have $\widehat{\beta}_k - \beta_0 = (X^\top W X)^{-1} X^\top W \varepsilon$. On the other hand, by the quantized error model Eq. (2.13), it holds that $\mathbb{E}[\varepsilon_i|X] = 0$ and $\mathbb{E}[\varepsilon_i^2|X] \leqslant 4^{-(k_i+1)}M^2 = w_i^{-1}M^2$. Subsequently,

$$\mathbb{E}\|\widehat{\beta}_k - \beta_0\|_2^2$$
$$= \text{tr}\left[(X^\top W X)^{-1} X^\top W \mathbb{E}(\varepsilon\varepsilon^\top) W^\top X (X^\top W X)^{-1} \right]$$
$$\leqslant M^2 \cdot \text{tr}\left[(X^\top W X)^{-1} X^\top W X (X^\top W X)^{-1} \right]$$
$$= M^2 \cdot \text{tr}\left[(X^\top W X)^{-1} \right].$$

Chapter 3

Selective queries in nonparametric optimization

Let $f : \mathcal{X} \to \mathbb{R}$ be an *unknown* function defined on a known d-dimensional domain $\mathcal{X} \subseteq \mathbb{R}^d$ with non-empty interior. The objective of *optimization* is to find the minimizer of f over $\mathcal{X}$, or more specifically

$$\min_{x \in \mathcal{X}} f(x), \tag{3.1}$$

assuming that such a minimizer exists.

Unlike the traditional optimization literature, where the objective function f to be minimized is *known* and first-order or second-order of derivatives of f are readily accessible, in this chapter we consider the case where the function f itself is unknown and therefore no derivatives of f can be accurately evaluated. Instead, an algorithm gains information about f through n rounds of *interactive* queries from a noisy oracle, and produces an estimate $\widehat{x}_n \in \mathcal{X}$ that approximately minimizes the unknown function f.

More specifically, at time $t \in \{1, 2, \cdots, n\}$, an algorithm picks a query point $x_t \in \mathcal{X}$ and observes feedback

$$y_t = f(x_t) + \xi_t, \tag{3.2}$$

where $\xi_t \sim \mathcal{N}(0, \sigma^2)$ is a centered Gaussian random variable reflecting the noise in the evaluation of $f(x_t)$ inherent from the underlying measurement procedures. The optimization error can then be evaluated as $\mathbb{E} f(\widehat{x}_n) - f^*$ where $f^* = \min_{x \in \mathcal{X}} f(x)$, reflecting the gap in function values between the estimated minimizer $\widehat{x}_n$ and the true minimizer $x^* \in \operatorname{argmin}_{x \in \mathcal{X}} f(x)$.

Some constraints such as smoothness or convexity are imposed on the unknown function f (denoted as $f \in \mathcal{F}$) to make the approximate optimization problem information-theoretically feasible. Nevertheless, no strong assumptions such as explicit parametric forms are assumed for f, making the problem essentially *nonparametric*.

The (noisy) nonparametric optimization problem has important applications in machine learning and operations research, under the names of zeroth-order (derivative-free) optimization, black-box optimization, Bayesian optimization, and/or simulation optimization. They can be applied to problems such as hyper-parameter tuning in machine learning systems or search for optimal parameters in experimental or simulation studies (Leeds et al., 2014; Nakamura et al., 2017; Reeja-Jayan et al., 2012; Snoek et al., 2012).

We consider in this chapter three different settings of selective queries in nonparametric optimization. Sec. 3.1 considers very low domain dimension d, but allows the function f to be

37

very flexible. On the other hand, Sec. 3.2 considers the setting of very high domain dimension d (even exceeding the total number of queries n), but places strong assumptions such as convexity and sparsity on f. Finally, Sec. 3.3 studies the *non-stationary* or *dynamic* settings of nonparametric optimization, in which the function f to be optimized may change over time. Such non-stationary optimization questions have important applications in dynamic pricing and other revenue management problems (Besbes et al., 2015).

3.1 Nonparametric optimization: local minimax rates

We consider the question of optimizing a nonparametric $f : \mathcal{X} \rightarrow \mathbb{R}$ over the unit cube $\mathcal{X} = [0,1]^d$. The optimization error of $\widehat{x}_n \in \mathcal{X}$ is evaluated by

$$\mathfrak{L}(\widehat{x}_n; f) := f(\widehat{x}_n) - f^* \qquad \text{where} \quad f^* := \inf_{x \in \mathcal{X}} f(x). \tag{3.3}$$

For simplicity, the variance of the noise variables $\{\xi_t\}_{t=1}^n$ in Eq. (3.2) is set as $\sigma^2 = 1$.

3.1.1 Local minimax rates

We use the classical *local minimax analysis* (Van der Vaart, 1998) to understand the fundamental information-theoretical limits of noisy global optimization of smooth functions. On the upper bound side, we seek (active) estimators $\widehat{x}_n$ such that

$$\sup_{f_0 \in \Theta} \sup_{f \in \Theta', \|f - f_0\|_\infty \leqslant \varepsilon_n(f_0)} \Pr_f \left[\mathfrak{L}(\widehat{x}_n; f) \geqslant C_1 \cdot R_n(f_0) \right] \leqslant 1/4, \tag{3.4}$$

where $C_1 > 0$ is a positive constant. Here $f_0 \in \Theta$ is referred to as the *reference function*, and $f \in \Theta'$ is the true underlying function which is assumed to be "near" f_0. The minimax convergence rate of $\mathfrak{L}(\widehat{x}_n; f)$ is then characterized *locally* by $R_n(f_0)$ which depends on the reference function f_0. The constant of $1/4$ is chosen arbitrarily and any small constant leads to similar conclusions. To establish negative results (i.e., locally minimax lower bounds), in contrast to the upper bound formulation, we assume the potential active optimization estimator $\widehat{x}_n$ has *perfect knowledge* about the reference function $f_0 \in \Theta$. We then prove locally minimax lower bounds of the form

$$\inf_{\widehat{x}_n} \sup_{f \in \Theta', \|f - f_0\|_\infty \leqslant \varepsilon_n(f_0)} \Pr_f \left[\mathfrak{L}(\widehat{x}_n; f) \geqslant C_2 \cdot R_n(f_0) \right] \geqslant 1/3, \tag{3.5}$$

where $C_2 > 0$ is another positive constant and $\varepsilon_n(f_0), R_n(f_0)$ are desired local convergence rates for functions near the reference f_0.

Although in some sense classical, the local minimax definition we propose warrants further discussion. We give some additional remarks on the parameters and the interpretation of Eq. (3.4).

1. **Roles of Θ and Θ':** The reference function f_0 and the true functions f are assumed to belong to different but closely related function classes Θ and Θ'. In particular, in our paper $\Theta \subseteq \Theta'$, meaning that less restrictive assumptions are imposed on the true underlying function f compared to those imposed on the reference function f_0 on which R_n and ε_n are based.

2. **Upper Bounds:** It is worth emphasizing that the estimator $\widehat{x}_n$ has no knowledge of the reference function f_0. From the perspective of upper bounds, we an consider the simpler task of producing f_0-dependent bounds (eliminating the second supremum) to instead study the (already interesting) quantity:

$$\sup_{f_0 \in \Theta} \Pr_{f_0} \left[\mathfrak{L}(\widehat{x}_n; f_0) \geqslant C_1 R_n(f_0) \right] \leqslant 1/4.$$

As indicated above we maintain the double-supremum in the definition because fewer assumptions are imposed directly on the true underlying function f, and further because it allows to more directly compare our upper and lower bounds.

3. **Lower Bounds and the choice of the "localization radius"** $\varepsilon_n(f_0)$**:** Our lower bounds allow the estimator knowledge of the reference function (this makes establishing the lower bound more challenging). Eq. (3.5) implies that no estimator $\widehat{x}_n$ can effectively optimize a function f close to f_0 beyond the convergence rate of $R_n(f_0)$, even if perfect knowledge of the reference function f_0 is available a priori. The $\varepsilon_n(f_0)$ parameter that decides the "range" in which local minimax rates apply is taken to be on the same order as the actual local rate $R_n(f_0)$ in this paper. This is (up to constants) the smallest radius for which we can hope to obtain non-trivial lower-bounds: if we consider a much smaller radius than $R_n(f_0)$ then the trivial estimator which outputs the minimizer of the reference function would achieve a faster rate than $R_n(f_0)$. Selecting the smallest possible radius makes establishing the lower bound most challenging but provides a refined picture of the complexity of zeroth-order optimization.

3.1.2 Assumptions

We state and motivate assumptions that will be used. The first assumption states that f is locally Hölder smooth on its level sets.

(A1) There exist constants $\kappa, \alpha, M > 0$ such that f restricted on $\mathcal{X}_{f,\kappa} := \{x \in \mathcal{X} : f(x) \leqslant f^* + \kappa\}$ belongs to the Hölder class $\Sigma^\alpha(M)$, meaning that f is k-times differentiable on $\mathcal{X}_{f,\kappa}$ and furthermore for any $x, x' \in \mathcal{X}_{f,\kappa}$, [1]

$$\sum_{j=0}^{k} \sum_{\alpha_1 + \ldots + \alpha_d = j} |f^{(\alpha,j)}(x)| + \sum_{\alpha_1 + \ldots + \alpha_d = k} \frac{|f^{(\alpha,k)}(x) - f^{(\alpha,k)}(x')|}{\|x - x'\|_\infty^{\alpha-k}} \leqslant M. \qquad (3.6)$$

Here $k = \lfloor \alpha \rfloor$ is the largest integer lower bounding α and $f^{(\alpha,j)}(x) := \partial^j f(x) / \partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}$. We use $\Sigma_\kappa^\alpha(M)$ to denote the class of all functions satisfying (A1). We remark that (A1) is weaker than the standard assumption that f on its entire domain $\mathcal{X}$ belongs to the Hölder class $\Sigma^\alpha(M)$. This is because places with function values larger than $f^* + \kappa$ can be easily detected and removed by a pre-processing step, as we describe in the next section.

Our next assumption concern the "regularity" of the *level sets* of the "reference" function f_0. Define $L_{f_0}(\epsilon) := \{x \in \mathcal{X} : f_0(x) \leqslant f_0^* + \epsilon\}$ as the ϵ-level set of f_0, and $\mu_{f_0}(\epsilon) := \lambda(L_{f_0}(\epsilon))$ as the Lebesgue measure of $L_{f_0}(\epsilon)$, also known as the *distribution function*. Define also $N(L_{f_0}(\epsilon), \delta)$ as the smallest number of ℓ_2-balls of radius δ that cover $L_{f_0}(\epsilon)$.

---

[1] the particular ℓ_∞ norm is used for convenience only and can be replaced by any equivalent vector norms.
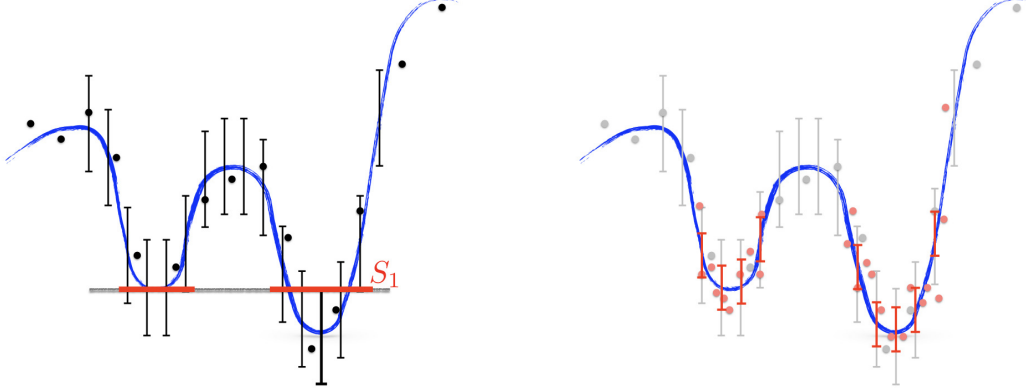
Figure 3.1: Informal illustrations of Algorithm 7. Solid blue curves depict the underlying function f to be optimized, black and red solid dots denote the query points and their responses $\{(x_t, y_t)\}$, and black/red vertical line segments correspond to uniform confidence intervals on function evaluations constructed using current batch of data observed. The left figure illustrates the first epoch of our algorithm, where query points are uniformly sampled from the entire domain $\mathcal{X}$. Afterwards, sub-optimal locations based on constructed confidence intervals are removed, and a shrinkt "candidate set" S_1 is obtained. The algorithm then proceeds to the second epoch, illustrated in the right figure, where query points (in red) are sampled only from the restricted candidate set and shorter confidence intervals (also in red) are constructed and updated. The procedure is repeated until $O(\log n)$ epochs are completed.

(A2) There exist constants $c_0 > 0$ and $C_0 > 0$ such that $N(L_{f_0}(\epsilon), \delta) \leqslant C_0[1 + \mu_{f_0}(\epsilon)\delta^{-d}]$ for all $\epsilon, \delta \in (0, c_0]$.

We use Θ_C to denote all functions that satisfy (A2) with respect to parameters $C = (c_0, C_0)$.

At a higher level, the regularity condition (A2) assumes that the level sets are sufficiently "regular" such that covering them with small-radius balls does not require significantly larger total volumes. For example, consider a perfectly regular case of $L_{f_0}(\epsilon)$ being the d-dimensional ℓ_2 ball of radius r: $L_{f_0}(\epsilon) = \{x \in \mathcal{X} : \|x - x^*\|_2 \leqslant r\}$. Clearly, $\mu_{f_0}(\epsilon) \asymp r^d$. In addition, the δ-covering number in ℓ_2 of $L_{f_0}(\epsilon)$ is on the order of $1 + (r/\delta)^d \asymp 1 + \mu_{f_0}(\epsilon)\delta^{-d}$, which satisfies the scaling in (A2).

When (A2) holds, uniform confidence intervals of f on its level sets are easy to construct because little statistical efficiency is lost by slightly enlarging the level sets so that complete d-dimensional cubes are contained in the enlarged level sets. On the other hand, when regularity of level sets fails to hold such nonparametric estimation can be very difficult or even impossible. As an extreme example, suppose the level set $L_{f_0}(\epsilon)$ consists of $\mathfrak{n}$ standalone and well-spaced points in $\mathcal{X}$: the Lebesgue measure of $L_{f_0}(\epsilon)$ would be zero, but at least $\Omega(\mathfrak{n})$ queries are necessary to construct uniform confidence intervals on $L_{f_0}(\epsilon)$. It is clear that such $L_{f_0}(\epsilon)$ violates (A2), because $N(L_{f_0}(\epsilon), \delta) \geqslant \mathfrak{n}$ as $\delta \to 0^+$ but $\mu_{f_0}(\epsilon) = 0$.

Algorithm 7 Successive elimination for low-dimensional nonparametric optimization
---
 1: **Parameters**: α, M, δ, n
 2: **Output**: $\widehat{x}_n = x_n$, the final prediction
 3: Initialization: $S_0 = G_n$, $\varrho_0(x) \equiv \infty$, $T = \lfloor \log_2 n \rfloor$, $n_0 = \lfloor n/T \rfloor$;
 4: **for** $\tau = 1, 2, \ldots, T$ **do**
 5:     Compute "extended" sample set $S_{\tau-1}^\circ(\varrho_{\tau-1})$ defined in Eq. (3.7);
 6:     **for** $t = (\tau - 1)n_0 + 1$ to τn_0 **do**
 7:         Sample x_t uniformly at random from $S_{\tau-1}^\circ(\varrho_{\tau-1})$ and observe $y_t = f(x_t) + w_t$;
 8:     **end for**
 9:     For every $x \in S_{\tau-1}$, find bandwidth $h_t(x)$ and build CI $[\ell_t(x), u_t(x)]$ in Eq. (3.12);
10:     $S_\tau := \{x \in S_{\tau-1} : \ell_t(x) \leqslant \min_{x' \in S_{\tau-1}} u_t(x')\}$, $\varrho_\tau(x) := \min\{\varrho_{\tau-1}(x), h_t(x)\}$.
11: **end for**

3.1.3   The successive elimination algorithm

Our algorithm is based on the idea of *successive elimination*, which eliminates candidate points that are proven sub-optimal. We start with a cleaner algorithm that operates under the slightly stronger condition that $\kappa = \infty$ in (A1), meaning that f is α-Hölder smooth on the entire domain $\mathcal{X}$. The generalization to $\kappa > 0$ being a constant is given in an additional pre-processing step.

Let $G_n \in \mathcal{X}$ be a *finite* grid of points in $\mathcal{X}$. We assume the finite grid G_n satisfies the following two mild conditions:

(B1) Points in G_n are sampled i.i.d. from an unknown distribution P_X on $\mathcal{X}$; furthermore, the density p_X associated with P_X satisfies $\underline{p}_0 \leqslant p_X(x) \leqslant \overline{p}_0$ for all $x \in \mathcal{X}$, where $0 < \underline{p}_0 \leqslant \overline{p}_0 < \infty$ are uniform constants;

(B2) $|G_n| \gtrsim n^{3d/\min(\alpha,1)}$ and $\log|G_n| = O(\log n)$.

Remark 3. Although typically the choices of the grid points G_n belong to the data analyst, in some applications the choices of design points are not completely free. For example, in material synthesis experiments some environment parameter settings (e.g., temperature and pressure) might not be accessible due to budget or physical constraints. Thus, we choose to consider less restrictive conditions imposed on the design grid G_n, allowing it to be more flexible in real-world applications.

For any subset $S \subseteq G_n$ and a "weight" function $\varrho : G_n \to \mathbb{R}^+$, define the *extension* $S^\circ(\varrho)$ of S with respect to ϱ as

$$S^\circ(\varrho) := \bigcup_{x \in S} B_{\varrho(x)}^\infty(x; G_n) \quad \text{where} \quad B_{\varrho(x)}^\infty(x; G_n) = \{z \in G_n : \|z - x\|_\infty \leqslant \varrho(x)\}. \quad (3.7)$$

The algorithm can then be formulated as two level of iterations, with the outer loop shrinking the "active set" S_τ and the inner loop collecting data that reduce lengths of confidence intervals on the active set. An intuitive illustration of our proposed algorithm is given in Fig. 3.1, and a pseudo-code description is given in Algorithm 7.

Local Polynomial Regression   We use local polynomial regression (Fan & Gijbels, 1996) to obtain the estimate $\widehat{f}(x)$. In particular, for any $x \in G_n$ and a bandwidth parameter $h > 0$,

consider a least square polynomial estimate

$$\widehat{f}_h \in \arg\min_{g \in \mathcal{P}_k} \sum_{t'=1}^{t} \mathbb{I}[x_{t'} \in B_h^\infty(x)] \cdot (y_{t'} - g(x_{t'}))^2, \tag{3.8}$$

where $B_h^\infty(x) := \{x' \in \mathcal{X} : \|x' - x\|_\infty \leqslant h\}$ and $\mathcal{P}_k$ denotes all polynomials of degree k on $\mathcal{X}$.

To analyze the performance of $\widehat{f}_h$ evaluated at a certain point $x \in \mathcal{X}$, define mapping $\psi_{x,h} :$ $z \mapsto (1, \psi_{x,h}^1(z), \ldots, \psi_{x,h}^k(z))$ where $\psi_{x,h}^j : z \mapsto [\prod_{\ell=1}^{j} h^{-1}(z_{i_\ell} - x_{i_\ell})]_{i_1,\ldots,i_j=1}^{d}$ is the degree-j polynomial mapping from $\mathbb{R}^d$ to $\mathbb{R}^{d^j}$. Also define $\Psi_{t,h} := (\psi_{x,h}(x_{t'}))_{1 \leqslant t' \leqslant t, x_{t'} \in B_h(x)}$ as the $m \times D$ aggregated design matrix, where $m = \sum_{t'=1}^{t} \mathbb{I}[x_{t'} \in B_h^\infty(x)]$ and $D = 1 + d + \ldots + d^k$, $k = \lfloor \alpha \rfloor$. The estimate $\widehat{f}_h$ defined in Eq. (3.8) then admits the following closed-form expression:

$$\widehat{f}_h(z) \equiv \psi_{x,h}(z)^\top (\Psi_{t,h}^\top \Psi_{t,h})^\dagger \Psi_{t,h}^\top Y_{t,h}, \tag{3.9}$$

where $Y_{t,h} = (y_{t'})_{1 \leqslant t' \leqslant t, x_{t'} \in B_h^\infty(x)}$ and $A^\dagger$ is the Moore-Penrose pseudo-inverse of A.

The following lemma gives a finite-sample analysis of the error of $\widehat{f}_h(x)$:

Lemma 12. *Suppose f satisfies Eq. (3.6) on $B_h^\infty(x; \mathcal{X})$, $\max_{z \in B_h^\infty(x; \mathcal{X})} \|\psi_{x,h}(z)\|_2 \leqslant b$ and $\frac{1}{m}\Psi_{t,h}^\top \Psi_{t,h} \geq \sigma I_{D \times D}$ for some $\sigma > 0$. Then for any $\delta \in (0, 1/2)$, with probability $1 - \delta$*

$$\left| \widehat{f}_h(x) - f(x) \right| \leqslant \underbrace{\frac{b^2}{\sigma} M d^k h^\alpha}_{\mathfrak{b}_{h,\delta}(x)} + \underbrace{b\sqrt{\frac{5D \ln(1/\delta)}{\sigma m}}}_{\mathfrak{s}_{h,\delta}(x)} =: \eta_{h,\delta}(x). \tag{3.10}$$

Remark 4. $\mathfrak{b}_{h,\delta}(x)$, $\mathfrak{s}_{h,\delta}(x)$ and $\eta_{h,\delta}(x)$ depend on x because σ depends on $\Psi_{t,h}$, which further depends on the sample points in the neighborhood $B_h^\infty(x; \mathcal{X})$ of x.

In the rest of the paper we define $\mathfrak{b}_{h,\delta}(x) := (b^2/\sigma)M d^k h^\alpha$ and $\mathfrak{s}_{h,\delta}(x) := b\sqrt{5D \ln(1/\delta)/\sigma m}$ as the bias and standard deviation terms in the error of $\widehat{f}_h(x)$, respectively. We also denote $\eta_{h,\delta}(x) := \mathfrak{b}_{h,\delta}(x) + \mathfrak{s}_{h,\delta}(x)$ as the overall error in $\widehat{f}_h(x)$.

Notice that when bandwidth h increases, the bias term $\mathfrak{b}_{h,\delta}(x)$ is likely to increase too because of the h^α term; on the other hand, with h increasing the local neighborhood $B_h^\infty(x; \mathcal{X})$ enlarges and would potentially contain more samples, implying a larger m and smaller standard deviation term $\mathfrak{s}_{h,\delta}(x)$. A careful selection of bandwidth h balances $\mathfrak{b}_{h,\delta}(x)$ and $\mathfrak{s}_{h,\delta}(x)$ and yields appropriate confidence intervals on $f(x)$, a topic that is addressed in the next section.

Bandwidth Selection and Confidence Intervals   Given the expressions of bias $\mathfrak{b}_{h,\delta}(x)$ and standard deviation $\mathfrak{s}_{h,\delta}(x)$ in Eq. (3.10), the bandwidth $h_t(x) > 0$ at epoch t and point x is selected as

$$h_t(x) := \frac{j_t(x)}{n^2} \quad \text{where} \quad j_t(x) := \arg\max \left\{ j \in \mathbb{N}, j \leqslant n^2 : \mathfrak{b}_{j/n^2, \delta}(x) \leqslant \mathfrak{s}_{j/n^2, \delta}(x) \right\}. \tag{3.11}$$

More specifically, $h_t(x)$ is the largest positive value in an evenly spaced grid $\{j/n^2\}$ such that the bias of $\widehat{f}_h(x)$ is smaller than its standard deviation. Such bandwidth selection is in principle

similar to the Lepski's method (Lepski et al., 1997), with the exception that an upper bound on the bias for any bandwidth parameter is known and does not need to be estimated from data.

With the selection of bandwidth $h_t(x)$ at epoch t and query point x, a confidence interval on $f(x)$ is constructed as

$$\ell_t(x) := \max_{1 \leq t' \leq t} \left\{ \widehat{f}_{h_{t'}(x)}(x) - \eta_{h_{t'}(x),\delta}(x) \right\} \quad \text{and} \quad u_t(x) := \min_{1 \leq t' \leq t} \left\{ \widehat{f}_{h_{t'}(x)}(x) + \eta_{h_{t'}(x),\delta}(x) \right\}.$$
(3.12)

Note that for any $x \in \mathcal{X}$, the lower confidence edge $\ell_t(x)$ is a non-decreasing function in t and the upper confidence edge $u_t(x)$ is a non-increasing function in t.

Pre-screening We describe a pre-screening procedure that relaxes the smoothness condition from $\kappa = \infty$ to $\kappa = \Omega(1)$, meaning that only local smoothness of f around its minimum values is required. Let $n_0 = \lfloor n/\log n \rfloor$, $x_1, \ldots, x_{n_0}$ be points i.i.d. uniformly sampled from $\mathcal{X}$ and $y_1, \ldots, y_{n_0}$ be their corresponding responses. For every grid point $x \in G_n$, perform the following:

1. Compute $\breve{f}(x)$ as the average of all y_i such that $\|x_i - x\|_\infty \leq n_0^{-1/2d} \log^3 n =: h_0$;

2. Remove all $x \in G_n$ from S_0 if $\breve{f}(x) \geq \min_{z \in G_n} \breve{f}(z) + 1/\log n$.

Remark 5. The $1/\log n$ term in removal condition $\breve{f}(x) \geq \min_{z \in G_n} \breve{f}(z) + 1/\log n$ is not important, and can be replaced with any sequence $\{\omega_n\}$ such that $\lim_{n \to \infty} \omega_n = 0$ and $\lim_{n \to \infty} \omega_n n^t = \infty$ for any $t > 0$. The readers are referred to the proof of Proposition 6 in the appendix for the motivation of this term as well as the selection of the pre-screening bandwidth h_0.

At a high level, the pre-screening step computes local averages of y and remove grid points in $S_0 = G_n$ whose estimated values are larger than the minimum in G_n.

To analyze the pre-screening step, we state the following proposition:

Proposition 6. *Assume $f \in \Sigma_\kappa^\alpha(M)$ and let S_0' be the screened grid after step 2 of the pre-screening procedure. Then for sufficiently large n, with probability $1 - O(n^{-1})$ we have*

$$\min_{x \in S_0'} f(x) = \min_{z \in G_n} f(x) \quad and \quad S_0' \subseteq \bigcup_{x \in L_f(\kappa/2)} B_{h_0}^\infty(x; \mathcal{X}),$$
(3.13)

where $L_f(\kappa/2) = \{x \in \mathcal{X} : f(x) \leq f^* + \kappa/2\}$.

To interpret Proposition 6, note that for sufficiently large n, $f \in \Sigma_\kappa^\alpha(M)$ implies f being α-Hölder smooth (i.e., f satisfies Eq. (3.6)) on $\bigcup_{x \in L_f(\kappa/2)} B_{h_0}^\infty(x; \mathcal{X})$, because $\kappa > 0$ is a constant and $h_0 \to 0$ as $n \to \infty$. Subsequently, the proposition shows that with high probability, the pre-screening step will remove all grid points in G_n in non-smooth regions of f, while maintaining the global optimal solution. This justifies the pre-processing step for $f \in \Sigma_\kappa^\alpha(M)$, because f is smooth on the grid after pre-processing.

The proof of Proposition 6 uses the fact that the local mean estimation is large provided that all data points in the local mean estimator are large, regardless of their underlying smoothness. The complete proof of Proposition 6 is deferred to the appendix. Its proof is simple and is deferred to the appendix.

3.1.4 Locally minimax upper bounds

The following theorem is our main result that upper bounds the local minimax rate of noisy global optimization with active queries.

Theorem 4. *For any* $\alpha, M, \kappa, c_0, C_0 > 0$ *and* $f_0 \in \Sigma_\kappa^\alpha(M) \cap \Theta_C$, *where* $C = (c_0, C_0)$, *define*

$$\varepsilon_n^{\mathsf{U}}(f_0) := \sup \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geqslant n/\log^\omega n \right\}, \qquad (3.14)$$

where $\omega > 5 + d/\alpha$ *is a large constant. Suppose also that* $\varepsilon_n^{\mathsf{U}}(f_0) \to 0$ *as* $n \to \infty$. *Then for sufficiently large* n, *there exists an estimator* $\widehat{x}_n$ *with access to* n *active queries* $x_1, \ldots, x_n \in \mathcal{X}$, *a constant* $C_R > 0$ *depending only on* $\alpha, M, \kappa, c, c_0, C_0$ *and a constant* $\gamma > 0$ *depending only on* α *and* d *such that*

$$\sup_{f_0 \in \Sigma_\kappa^\alpha(M) \cap \Theta_C} \sup_{\substack{f \in \Sigma_\kappa^\alpha(M), \\ \|f - f_0\|_\infty \leqslant \varepsilon_n^{\mathsf{U}}(f_0)}} \Pr_f \left[\mathfrak{L}(\widehat{x}_n, f) > C_R \log^\gamma n \cdot (\varepsilon_n^{\mathsf{U}}(f_0) + n^{-1/2}) \right] \leqslant 1/4. \qquad (3.15)$$

Remark 6. Unlike the (local) smoothness class $\Sigma_\kappa^\alpha(M)$, the additional function class Θ_C that encapsulates (A2) is imposed only on the "reference" function f_0 but not the true function f to be estimated. This makes the assumptions considerably weaker because the true function f may violate either or both (A2) while our results remain valid.

Remark 7. The estimator $\widehat{x}_n$ does *not* require knowledge of parameters κ, c_0, C_0 or $\varepsilon_n^{\mathsf{U}}(f_0)$, and automatically adapts to them, as shown in the next section. While the knowledge of smoothness parameters α and M seems to be necessary, we remark that it is possible to adapt to α and M by running $O(\log^2 n)$ parallel sessions of $\widehat{x}_n$ on $O(\log n)$ grids of α and M values, and then using $\Omega(n/\log^2 n)$ single-point queries to decide on the location with the smallest function value. Such an adaptive strategy was suggested in Grill et al. (2015) to remove an additional condition in Minsker (2013), which also applies to our settings.

Remark 8. When the distribution function $\mu_{f_0}(\epsilon)$ does not change abruptly with ϵ the expression of $\varepsilon_n^{\mathsf{U}}(f_0)$ can be significantly simplified. In particular, if for all $\epsilon \in (0, c_0]$ it holds that

$$\mu_{f_0}(\epsilon/\log n) \geqslant \mu_{f_0}(\epsilon)/[\log n]^{O(1)}, \qquad (3.16)$$

then $\varepsilon_n^{\mathsf{U}}(f_0)$ can be upper bounded as

$$\varepsilon_n^{\mathsf{U}}(f_0) \leqslant [\log n]^{O(1)} \cdot \sup \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geqslant n \right\}. \qquad (3.17)$$

It is also noted that if $\mu_{f_0}(\epsilon)$ has a polynomial behavior of $\mu_{f_0}(\epsilon) \asymp \epsilon^\beta$ for some constant $\beta \geqslant 0$, then Eq. (3.16) is satisfied and so is Eq. (3.17).

The quantity $\varepsilon_n^{\mathsf{U}}(f_0) = \inf\{\varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geqslant n/\log^\omega n\}$ is crucial in determining the convergence rate of optimization error of $\widehat{x}_n$ *locally* around the reference function f_0. While the definition of $\varepsilon_n^{\mathsf{U}}(f_0)$ is mostly implicit and involves solving an inequality concerning the distribution function $\mu_{f_0}(\cdot)$, we remark that it admits a simple form when μ_{f_0} has a polynomial growth rate similar to a local Tsybakov noise condition (Korostelev & Tsybakov, 2012; Tsybakov, 2009), as shown by the following proposition:

Proposition 7. *Suppose $\mu_{f_0}(\epsilon) \lesssim \epsilon^\beta$ for some constant $\beta \in [0, 2 + d/\alpha)$. Then $\varepsilon_n^{\mathsf{U}}(f_0) = \widetilde{O}(n^{-\alpha/(2\alpha+d-\alpha\beta)})$. In addition, if $\beta \in [0, d/\alpha]$ then $\varepsilon_n^{\mathsf{U}}(f_0) + n^{-1/2} \lesssim \varepsilon_n^{\mathsf{U}}(f_0) = \widetilde{O}(n^{-\alpha/(2\alpha+d-\alpha\beta)})$.*

We remark that the condition $\beta \in [0, d/\alpha]$ was also adopted in the previous work (Minsker, 2013, Remark 6). Proposition 7 can be easily verified by solving the system $\varepsilon^{-(2+d/\alpha)}\mu_{f_0}(\varepsilon) \geqslant n/\log^\omega n$ with the condition $\mu_{f_0}(\epsilon) \lesssim \epsilon^\beta$. We therefore omit its proof. The following two examples give some simple reference functions f_0 that satisfy the $\mu_{f_0}(\epsilon) \lesssim \epsilon^\beta$ condition in Proposition 7 with particular values of β.

Example 1. The constant function $f_0 \equiv 0$ satisfies (A1) through (A3) with $\beta = 0$.

Example 2. $f_0 \in \Sigma_\kappa^2(M)$ that is *strongly convex* [2] satisfies (A1) through (A3) with $\beta = d/2$.

Example 1 is simple to verify, as the volume of level sets of the constant function $f_0 \equiv 0$ exhibits a phase transition at $\epsilon = 0$ and $\epsilon > 0$, rendering $\beta = 0$ the only parameter option for which $\mu_{f_0}(\epsilon) \lesssim \epsilon^\beta$. Example 2 is more involved, and holds because the strong convexity of f_0 *lower bounds* the growth rate of f_0 when moving away from its minimum. We give a rigorous proof of Example 2 in the appendix. We also remark that f_0 does *not* need to be exactly strongly convex for $\beta = d/2$ to hold, and the example is valid for, e.g., piecewise strongly convex functions with a constant number of pieces too.

To best interpret the results in Theorem 4 and Proposition 7, it is instructive to compare the "local" rate $n^{-\alpha/(2\alpha+d-\alpha\beta)}$ with the baseline rate $n^{-\alpha/(2\alpha+d)}$, which can be attained by reconstructing f in sup-norm and producing $\widehat{x}_n \in \arg\min_{x \in \mathcal{X}} \widehat{f}(x)$. Since $\beta \geqslant 0$, the local convergence rate established in Theorem 4 is never slower, and the improvement compared to the baseline rate $n^{-\alpha/(2\alpha+d)}$ is dictated by β, which governs the growth rate of volume of level sets of the reference function f_0. In particular, for functions that grows fast when moving away from its minimum, the parameter β is large and therefore the local convergence rate around f_0 could be much faster than $n^{-\alpha/(2\alpha+d)}$. Theorem 4 also implies concrete convergence rates for special functions considered in Examples 1 and 2. For the constant reference function $f_0 \equiv 0$, Example 1 and Theorem 4 yield that $R_n(f_0) \asymp n^{-\alpha/(2\alpha+d)}$, which matches the baseline rate $n^{-\alpha/(2\alpha+d)}$ and suggests that $f_0 \equiv 0$ is the worst-case reference function. This is intuitive, because $f_0 \equiv 0$ has the most drastic level set change at $\epsilon \to 0^+$ and therefore small perturbations anywhere of f_0 result in changes of the optimal locations. On the other hand, if f_0 is strongly smooth and convex as in Example 2, Theorem 4 suggests that $R_n(f_0) \asymp n^{-1/2}$, which is significantly better than the $n^{-2/(4+d)}$ baseline rate [3] and also matches existing works on zeroth-order optimization of convex functions (Agarwal et al., 2010). The faster rate holds intuitively because strongly convex functions grows fast when moving away from the minimum, which implies small level set changes. An active query algorithm could then focus most of its queries onto the small level sets of the underlying function, resulting in more accurate local function reconstructions and faster optimization error rate.

---

[2] A twice differentiable function f_0 is strongly convex if there exists $\sigma > 0$ such that $\nabla^2 f_0(x) \geq \sigma I, \forall x \in \mathcal{X}$.
[3] Note that f_0 being strongly smooth implies $\alpha = 2$ in the local smoothness assumption.

3.1.5 Locally minimax lower bounds

We prove local minimax lower bounds that match the upper bounds in Theorem 4 up to logarithmic terms. As we remarked in Section 3.1.1, in the local minimax lower bound formulation we assume the data analyst has full knowledge of the reference function f_0, which makes the lower bounds stronger as more information is available a priori.

To facilitate such a strong local minimax lower bounds, the following additional condition is imposed on the reference function f_0 of which the data analyst has perfect information.

(A2') There exist constants $c_0', C_0' > 0$ such that $M(L_{f_0}(\epsilon), \delta) \geqslant C_0' \mu_{f_0}(\epsilon) \delta^{-d}$ for all $\epsilon, \delta \in (0, c_0']$, where $M(L_{f_0}(\epsilon), \delta)$ is the maximum number of disjoint ℓ_2 balls of radius δ that can be packed into $L_{f_0}(\epsilon)$.

We denote $\Theta_{C'}'$ as the class of functions that satisfy (A2') with respect to parameters $C' = (c_0', C_0') > 0$. Intuitively, (A2') can be regarded as the "reverse" version of (A2), which basically means that (A2) is "tight".

We are now ready to state our main negative result, which shows, from an information-theoretical perspective, that the upper bound in Theorem 4 is not improvable.

Theorem 5. *Suppose $\alpha, c_0, C_0, c_0', C_0' > 0$ and $\kappa = \infty$. Denote $C = (c_0, C_0)$ and $C' = (c_0', C_0')$. For any $f_0 \in \Theta_C \cap \Theta_{C'}'$, define*

$$\varepsilon_n^{\mathsf{L}}(f_0) := \sup \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geqslant n \right\}. \tag{3.18}$$

Then there exist constant $M > 0$ depending on α, d, C, C' such that, for any $f_0 \in \Sigma_\kappa^\alpha(M/2) \cap \Theta_C \cap \Theta_{C'}$,

$$\inf_{\hat{x}_n} \sup_{\substack{f \in \Sigma_\kappa^\alpha(M), \\ \|f - f_0\|_\infty \leqslant 2\varepsilon_n^{\mathsf{L}}(f_0)}} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geqslant \varepsilon_n^{\mathsf{L}}(f_0) \right] \geqslant \frac{1}{3}. \tag{3.19}$$

Remark 9. For any f_0 and n it always holds that $\varepsilon_n^{\mathsf{L}}(f_0) \leqslant \varepsilon_n^{\mathsf{U}}(f_0)$.

Remark 10. If the distribution function $\mu_{f_0}(\epsilon)$ satisfies Eq. (3.16) in Remark 8, then $\varepsilon_n^{\mathsf{L}}(f_0) \geqslant \varepsilon_n^{\mathsf{U}}(f_0)/[\log n]^{O(1)}$.

Remark 9 shows that there might be a gap between the locally minimax upper and lower bounds in Theorems 4 and 5. Nevertheless, Remark 10 shows that under the mild condition of $\mu_{f_0}(\epsilon)$ does not change too abruptly with ϵ, the gap between $\varepsilon_n^{\mathsf{U}}(f_0)$ and $\varepsilon_n^{\mathsf{L}}(f_0)$ is only a poly-logarithmic term in n. Additionally, the following proposition derives explicit expression of $\varepsilon_n^{\mathsf{L}}(f_0)$ for reference functions whose distribution functions have a polynomial growth, which matches the Proposition 7 up to $\log n$ factors. Its proof is again straightforward.

Proposition 8. *Suppose $\mu_{f_0}(\epsilon) \gtrsim \epsilon^\beta$ for some $\beta \in [0, 2+d/\alpha)$. Then $\varepsilon_n^{\mathsf{L}}(f_0) = \Omega(n^{-\alpha/(2\alpha+d-\alpha\beta)})$.*

The following proposition additionally shows the existence of $f_0 \in \Sigma_\infty^\alpha(M) \cap \Theta_C \cap \Theta_{C'}$ that satisfies $\mu_{f_0}(\epsilon) \asymp \epsilon^\beta$ for any values of $\alpha > 0$ and $\beta \in [0, d/\alpha]$. Its proof is given in the appendix.

Proposition 9. *Fix arbitrary $\alpha, M > 0$ and $\beta \in [0, d/\alpha]$. There exists $f_0 \in \Sigma_\kappa^\alpha(M) \cap \Theta_C \cap \Theta_{C'}$ for $\kappa = \infty$ and constants $C = (c_0, C_0)$, $C' = (c_0', C_0')$ that depend only on α, β, M and d such that $\mu_{f_0}(\epsilon) \asymp \epsilon^\beta$.*

Theorem 5 and Proposition 8 show that the $n^{-\alpha/(2\alpha+d-\alpha\beta)}$ upper bound on local minimax convergence rate established in Theorem 4 is not improvable up to logarithmic factors of n. Such information-theoretical lower bounds on the convergence rates hold *even if the data analyst has perfect information of* f_0, the reference function on which the $n^{-\alpha/(2\alpha+d-\alpha\beta)}$ local rate is based. Our results also imply an $n^{-\alpha/(2\alpha+d)}$ minimax lower bound over all α-Hölder smooth functions, showing that without additional assumptions, noisy optimization of smooth functions is as difficult as reconstructing the unknown function in sup-norm.

Our proof of Theorem 5 also differs from existing minimax lower bound proofs for active nonparametric models (Castro & Nowak, 2008). The classical approach is to invoke Fano's inequality and to upper bound the KL divergence between different underlying functions f and g using $\|f - g\|_\infty$, corresponding to the point $x \in \mathcal{X}$ that leads to the largest KL divergence. Such an approach, however, does not produce tight lower bounds for our problem. To overcome such difficulties, we borrow the lower bound analysis for bandit pure exploration problems in (Bubeck et al., 2009). In particular, our analysis considers the query distribution of any active query algorithm $\mathcal{A} = (\varphi_1, \ldots, \varphi_n, \phi_n)$ under the reference function f_0 and bounds the perturbation in query distributions between f_0 and f using Le Cam's lemma. Afterwards, an adversarial function choice f can be made based on the query distributions of the considered algorithm $\mathcal{A}$.

Theorem 5 applies to any global optimization method that makes *active* queries. The following theorem, on the other hand, shows that for passive algorithms (i.e., $x_1, \cdots, x_n$ drawn independently at uniform from $\mathcal{X}$) the $n^{-\alpha/(2\alpha+d)}$ optimization rate is not improvable even with additional level set assumptions imposed on f_0. This demonstrates an explicit gap between passive and adaptive query models in global optimization problems.

Theorem 6. *Suppose* $\alpha, c_0, C_0, c_0', C_0' > 0$ *and* $\kappa = \infty$. *Denote* $C = (c_0, C_0)$ *and* $C' = (c_0', C_0')$. *Then there exist constant* $M > 0$ *depending on* α, d, C, C' *and* N *depending on* M *such that, for any* $f_0 \in \Sigma_\kappa^\alpha(M/2) \cap \Theta_C \cap \Theta_{C'}$ *satisfying* $\varepsilon_n^{\mathsf{L}}(f_0) \leqslant \tilde{\varepsilon}_n^{\mathsf{L}} =: [\log n/n]^{\alpha/(2\alpha+d)}$,

$$\inf_{\breve{x}_n} \sup_{\substack{f \in \Sigma_\kappa^\alpha(M), \\ \|f-f_0\|_\infty \leqslant 2\tilde{\varepsilon}_n^{\mathsf{L}}}} \Pr_f \left[\mathfrak{L}(\hat{x}_n; f) \geqslant \tilde{\varepsilon}_n^{\mathsf{L}} \right] \geqslant \frac{1}{3} \quad \text{for all } n \geqslant N. \tag{3.20}$$

Intuitively, the apparent gap demonstrated by Theorems 5 and 6 between the active and passive query models stems from the observation that, a passive algorithm $\mathcal{A}$ only has access to uniformly sampled query points $x_1, \ldots, x_n$ and therefore cannot focus on a small level set of f in order to improve query efficiency. In addition, for functions that grow faster when moving away from their minima (implying a larger value of β), the gap between passive and active query models becomes bigger as active queries can more effectively exploit the restricted level sets of such functions.

3.2 High-dimensional derivative-free optimization

Consider optimizing an unknown function $f : \mathbb{R}^d \to \mathbb{R}$ in *very high dimensions*: so high that d even exceeds the number of queries allowed n.

In such settings, additional sparsity type assumptions are mandatory to ensure identifiability, similar to high-dimensional regression problems in statistics (van de Geer, 2000). Such sparsity

assumptions were reflected assumption (A5) in the next section, which we motivate from two real-world example applications of hyper-parameter tuning and visual stimuli optimization.

3.2.1 Assumptions and motivations

We make the following assumptions on the target function $f : \mathcal{X} \to \mathbb{R}$ to be optimized:

(A1) (*Unconstrained convex optimization*): We take $\mathcal{X} = \mathbb{R}^d$ and assume that f is convex, i.e. for all $x, x' \in \mathcal{X}$ and $\lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)x') \leqslant \lambda f(x) + (1 - \lambda)f(x')$.

(A2) (*Minimizer of bounded ℓ_1-norm*): We assume there exists $x^* \in \mathcal{X}$ such that $f(x^*) = f^* = \inf_{x \in \mathcal{X}} f(x)$ and $\|x^*\|_1 \leqslant B$; x^* does not have to be unique.

(A3) (*Sparsity of gradients*): We assume that f is differentiable and that there exist $H > 0$, $s \ll d$ such that $\|\nabla f(x)\|_0 \leqslant s$ and $\|\nabla f(x)\|_1 \leqslant H$ for all $x \in \mathcal{X}$, where $\|z\|_0$ and $\|z\|_1$ are the ℓ_0 and ℓ_1 vector norms; the support of $\nabla f(x)$ could potentially vary with $x \in \mathcal{X}$.

(A4) (*Weak sparsity of Hessians*): We assume that f is twice differentiable and there exists $H > 0$ such that $\|\nabla^2 f(x)\|_1 \leqslant H$ for all where $\|A\|_1 := \sum_{i,j=1}^d |A_{ij}|$ is the entry-wise ℓ_1 norm of matrix A.

(A3) and (A4) are key assumptions in our paper, which assumes the gradients of f are *sparse*, and places a weaker sparsity assumption on the Hessian matrices that constrains their ℓ_1 norm rather than ℓ_0 norm. We also note that, assuming $\|\nabla f(x)\|_\infty$ and $\|\nabla^2 f(x)\|_\infty$ are both bounded, both (A3) and (A4) are implied by the following stronger but more intuitive "function sparsity" assumption:

(A5) (*Function sparsity*): there exists $S \subseteq [d]$, $|S| \leqslant s$ and $f_S : \mathbb{R}^{|S|} \to \mathbb{R}$ such that $f(x) \equiv f_S(x_S)$, where $x_S \in \mathbb{R}^{|S|}$ is the restriction of $x \in \mathbb{R}^d$ on S.

We motivate Assumptions (A3), (A4) and (A5) from both theoretical and practical perspectives. Theoretically, the sparsity assumption allows us to estimate the gradient at a specific point using $n \ll d$ noisy zeroth-order queries. On the other hand, (A5) is at least approximately satisfied in many practical applications of zeroth-order optimization. For example, in hyper-parameter tuning problems of learning systems, it is usually the case that the performance of the system is insensitive to some hyper-parameters, essentially implying the sparsity of the gradients and Hessians. Other examples include the optimization of visual stimuli so that certain types of neural responses are maximized or optimizing experimental parameters (pressure, temperature, etc.) so that the resulting synthesized material has optimal quality (Nakamura et al., 2017; Reeja-Jayan et al., 2012). For the visual stimuli optimization example, it is well known that the hierarchical organization of the human visual system in the brain into regions such as V1, V4, LO, IT etc. is precisely based on the neural response in these regions being sensitive to specific subsets of low-level and higher-level features such as edges and curves. This in turn implies that the underlying function to be optimized satisfies (A5). Finally, we remark that similar sparsity assumptions have been considered in past work (Bandeira et al., 2012; Lei et al., 2017) to obtain improved rates of convergence for optimization methods.

3.2.2 The zeroth-order mirror descent framework

Mirror descent (MD) (Yudin & Nemirovskii, 1983) is a classical method in optimization when smoothness and the domain geometry are measured in (possibly) non-Euclidean metrics. The MD algorithm was applied to stochastic optimization with noisy first-order oracles in the papers (Agarwal et al., 2012; Nemirovski et al., 2009) and was also studied in the work (Lan, 2012) for strongly smooth composite functions with accelerated rates, and in the works (Ghadimi & Lan, 2012, 2013a) for strongly convex composite functions.

Let $\psi : \mathcal{X} \to \mathbb{R}$ be a continuously differentiable, strictly convex function. The *Bregman divergence* $\Delta_\psi : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is defined as

$$\Delta_\psi(x, y) := \psi(y) - \psi(x) - \langle \nabla \psi(x), y - x \rangle. \tag{3.21}$$

Let $\| \cdot \|_\psi$ be a norm and $\| \cdot \|_{\psi*}$ be its dual norm, defined as $\|z\|_{\psi*} := \sup\{z^\top x : \|x\|_\psi \leqslant 1\}$. One important class of Bregman divergences is those that are κ-*strongly convex* with respect to the chosen norm, i.e. they satisfy $\Delta_\psi(x, y) \geqslant \frac{\kappa}{2}\|x - y\|_\psi^2$.

Many choices of ψ lead to a strongly convex Bregman divergence. In this paper we consider the ℓ_a norm as choice of ψ: $\psi_a(x) := \frac{1}{2(a-1)}\|x\|_a^2$ for $1 < a \leqslant 2$. It was proved in (Agarwal et al., 2012; Srebro et al., 2011) that ψ_a leads to a valid Bregman divergence that satisfies 1-strong convexity with respect to $\| \cdot \|_a$. For the $a = 1$ case, we use $\psi_{a'}$ with $a' = \frac{2 \log d}{2 \log d - 1}$ as its potential, which satisfies $\Delta_\psi(x, y) \geqslant \frac{\kappa}{2}\|x - y\|_1^2$ with $\kappa = e$.

With this setup, the MD method iteratively computes

$$x_{t+1} := \arg \min_{x \in \widetilde{\mathcal{X}}} \left\{ \eta_t \nabla f(x_t)^\top (x - x_t) + \Delta_\psi(x, x_t) \right\},$$

where $\{\eta_t\}_{t=1}^T$ is a sequence of step sizes and $\widetilde{\mathcal{X}} \subseteq \mathcal{X}$ is a subset of the domain $\mathcal{X}$ of f.

In zeroth-order optimization settings, the exact gradient $\nabla f(x_t)$ is not available. Instead, we use an *estimated gradient* $\widetilde{g}_t \approx \nabla f(x_t)$ to replace $\nabla f(x_t)$ in the mirror descent update rules. We shall refer to this algorithm as *zeroth-order mirror descent*, whose performance (convergence rates) would depend on the properties of the gradient estimates $\widetilde{g}_t$.

3.2.3 Sparse gradient estimation via the de-biased Lasso

In this section we introduce the Lasso and the de-biased Lasso gradient estimator to estimate sparse gradients. More specifically, for any $x_t \in \mathcal{X}$, the estimator uses $n \ll d$ samples to estimate the unknown gradient $g_t := \nabla f(x_t)$. The high-level idea is to consider $n \ll d$ random samples near the point x_t, and to then formulate the gradient estimation problem as a biased linear regression system. The Lasso procedure (and its de-biased variants) can then be applied to obtain a consistent estimator under certain sparsity assumptions on $\{g_t\}_{t=1}^T$.

Fix an arbitrary $x_t \in \mathcal{X}$ and let $z_1, \ldots, z_m \in \{\pm 1\}^d$ be m samples of i.i.d. binary random vectors such that $\Pr[z_{ij} = 1] = \Pr[z_{ij} = -1] = 1/2$, where $i \in [n]$ and $j \in [d]$. Let $\delta > 0$ be a probing parameter which will be specified later, and $y_1 = f(x_t + \delta z_1) + \xi_1, \ldots, y_n = f(x_t + \delta z_m) + \xi_m$ be m observations. Using first-order Taylor expansions with Lagrangian remainders, the normalized $\widetilde{y}_i := y_i/\delta$ can be written as

$$\widetilde{y}_i = \frac{f(x_t + \delta z_i) + \xi_i}{\delta} = \frac{f(x_t)}{\delta} + g_t^\top z_i + \frac{\delta}{2} z_i^\top H_t(\kappa_i, z_i) z_i + \delta^{-1} \xi_i := \mu_t + g_t^\top z_i + \varepsilon_i, \tag{3.22}$$

where $\mu_t = \delta^{-1} f(x_t)$, $\varepsilon_i = \frac{\delta}{2} z_i^\top H_t(\kappa_i, z_i) z_i + \delta^{-1} \xi_i$ and $H_t(\kappa_i, z_i) = \nabla^2 f(x_t + \kappa_i \delta z_i)$ for some $\kappa_i \in (0, 1)$.

Eq. (3.22) shows that, essentially, the question of estimating $g_t = \nabla f(x_t)$ can be cast as a linear regression model with design $\{z_i\}_{i=1}^m$, unknown parameters $(\mu_t, g_t) \in \mathbb{R}^{d+1}$ and noise variables $\{\varepsilon_i\}_{i=1}^n$ whose bias (i.e., $\mathbb{E}[\varepsilon_i | z_i, x_t]$) goes to 0 as $\delta \to 0$, at the expense of increasing variance. Since g_t is a sparse vector as a consequence of (A3), one can use the Lasso (Tibshirani, 1996) to obtain an estimate of g_t and μ_t:

$$(\widehat{g}_t, \widehat{\mu}_t) = \arg\min_{g \in \mathbb{R}^d, \mu \in \mathbb{R}} \frac{1}{m} \sum_{i=1}^m (\widetilde{y}_i - g^\top z_i - \mu)^2 + \lambda \|g\|_1 + \lambda |\mu|, \tag{3.23}$$

where $\lambda > 0$ is a regularization parameter that will be specified later.

The following lemma shows that with a carefully chosen λ, $\widehat{g}_t$ is a good estimate of g_t in both ℓ_∞ and ℓ_1 norms.

Lemma 13. *Suppose (A1) through (A4) hold. Suppose also that $m = \Omega(s^2 \log d)$, $m \leq d$ and $\lambda \asymp \delta^{-1} \sigma \sqrt{\log d / m} + \delta H$. Then with probability $1 - O(d^{-2})$*

$$\max\{|\widehat{\mu}_t - \mu_t|, \|\widehat{g}_t - g_t\|_\infty\} \lesssim \frac{\sigma}{\delta} \sqrt{\frac{\log d}{m}} + \delta H.$$

Furthermore, with probability $1 - O(d^{-2})$ it holds that $\|\widehat{g}_t - g_t\|_1 \leq 2s \|\widehat{g}_t - g_t\|_\infty$.

Lemma 13 follows by the standard ℓ_1 and ℓ_∞ error bound analyses of the Lasso estimator (Bickel et al., 2009; Lounici, 2008). However, our model has a subtle difference from the standard high-dimensional regression model in that $\mathbb{E}[\varepsilon_i | z_i, x_t]$ are not exactly zero. and we provide a detailed proof in the Appendix.

Remark 11. The penalization of μ in Eq. (3.23) is in general unnecessary as it is a single component; however, we decide to keep this penalization term to simplify our analysis. Neither the estimation error nor the selection of the tuning parameter λ depend on knowledge of μ_t.

Remark 12. Lemma 13 reveals an interesting bias-variance tradeoff controlled by the "probing" parameter $\delta > 0$. When δ is close to 0, the bias (reflected by $\mathbb{E}[\varepsilon_i | z_i, x_t]$) resulting from the second-order Lagrangian remainder term $\frac{\delta}{2} z_i^\top H_t(\kappa_i, z_i) z_i$ is small; however, the variance of $\widehat{g}_t$ is large because the variance of the "stochastic" noise term ξ_i / δ increases as $\delta \to 0$; on the other hand, for large δ the stochastic variance is reduced but the bias from first-order approximation of $f(x_t)$ increases.

We further introduce the de-biased Lasso estimator (Javanmard & Montanari, 2014; Van de Geer et al., 2014; Zhang & Zhang, 2014) to reduce bias of the Lasso estimator for the purpose of constructing confidence intervals for low-dimensional model components. In our application, the bias-reduced gradient estimate allows stochastic noise to concentrate across epochs and leads to improved convergence rates.

Let $\widetilde{Y}_t = (\widetilde{y}_1, \ldots, \widetilde{y}_m) \in \mathbb{R}^n$ and $Z_t = (z_1, \ldots, z_m) \in \mathbb{R}^{n \times d}$ be the vector forms of $\{\widetilde{y}_i\}_{i=1}^m$ and $\{z_i\}_{i=1}^m$. Since the design points z_i are i.i.d. Rademacher variables, the de-biased gradient estimator $\widetilde{g}_t$ takes a particularly simple form:

$$\widetilde{g}_t := \widehat{g}_t + \frac{1}{m} Z_t^\top (\widetilde{Y}_t - Z_t \widehat{g}_t - \widehat{\mu}_t \cdot 1_m). \tag{3.24}$$

Here $(\widehat{g}_t, \widehat{\mu}_t)$ is the Lasso estimator defined in Eq. (3.23) and $1_m = (1, \ldots, 1) \in \mathbb{R}^m$ is the m-dimensional vector of all ones.

Lemma 14. *Suppose $m = \Omega(s^2 \log d)$. With probability $1 - O(d^{-2})$ it holds that*

$$\widetilde{g}_t = g_t + \zeta_t + \gamma_t;$$

where ζ_t is a d-dimensional random vector such that, for any $a \in \mathbb{R}^d$, $\langle \zeta_t, a \rangle$ conditioned on x_t is a centered sub-exponential random variable with parameters $\nu = \sqrt{m/2} \cdot \alpha$ and $\alpha \lesssim \sigma \|a\|_2 / \delta m$; and γ_t is a d-dimensional vector that satisfies

$$\|\gamma_t\|_\infty \lesssim H\delta + \frac{\sigma s \log d}{\delta m} \qquad almost\ surely.$$

Comparing Lemma 14 with the error bound obtained for the Lasso estimator $\widehat{g}_t$ in Lemma 13, it is clear that the entry-wise bias (i.e., $\|\gamma_t\|_\infty$) is reduced from $O(\delta H + \sqrt{\log d / \delta n})$ to $O(\delta H + s \log d / \delta n)$. Such de-biasing is at the cost of inflated stochastic error ζ_t, which means that unlike $\widehat{g}_t$, $\widetilde{g}_t$ is not a good estimator of g_t in the ℓ_1 or ℓ_2 norm.

3.2.4   Rates of convergence

The following theorem is the main result on the rates of convergence of our proposed zeroth-order mirror descent algorithm with sparse gradient estimates.

Theorem 7. *Suppose (A1) through (A4) hold. Suppose also that $n = \Omega(s^3 \log^2 d + s(1+H)^2(1 + B^4 H^4 \log^2 d))$, $n \leqslant d$ and that we choose the parameters $m := \lfloor (1+H)\sqrt{sn} \rfloor$, $\eta_t \equiv B\sqrt{\frac{n \log d}{n}}$, and $\delta_t \equiv \sqrt{s \log d / m}$. Then with probability $1 - O(d^{-1})$*

$$\mathbb{E}f(\widehat{x}_n) - f(x^*) \lesssim \xi_{\sigma,s} B \sqrt{\log d} \left[\frac{(1+H)^2 s}{n} \right]^{1/4} + \widetilde{O}(n^{-1/2}),$$

where $\xi_{\sigma,s} = 1 + \sigma + \sigma^2/s$, and $\widehat{x}_n$ is the average of all $\{x_t\}_{t=1}^{T'}$ with T' being the number of "epochs" in which m design points are constructed. In the $O(\cdot)$ notation we hide polynomial dependency on σ, s, H, B and $\log d$. The $\lesssim$ notation does not hide any dependency on problem dependent constants.

It is possible to further improve the convergence rates in Theorem 7 with additional smoothness conditions on $\nabla^2 f$, with a small loss of computational efficiency. Formally, we assume:

(A6) (*Hessian smoothness*). There exists $L > 0$ such that for all $x, x' \in \mathcal{X}$,

$$\|\nabla^2 f(x) - \nabla^2 f(x')\|_1 \leqslant L \|x - x'\|_\infty$$

Recall that $\|A\|_1 = \sum_{i,j} |A_{ij}|$ denotes the entry-wise ℓ_1 norm of a matrix A.

If f is three-times differentiable, then (A6) is implied by the condition that $\|\nabla^3 f(x)\|_1 \leqslant L$ for all $x \in \mathcal{X}$, where $\|A\|_1 := \sum_{i,j,k} |A_{ijk}|$ is the entry-wise ℓ_1 norm of a third order tensor. However, (A6) in general does not require third-order differentiability of f.

Recall the de-biased Lasso gradient estimator $\widetilde{g}_t(\delta)$ in Eqs. (3.23,3.24) corresponding to a probing step size of δ. Under the additional condition (A6), the analysis in Lemma 14 can be strengthened as below:

Lemma 15. *Suppose (A1) through (A4) and (A6) hold. Suppose also that $n = \Omega(s^2 \log d)$, $n \leqslant d$ and $\lambda \asymp \delta^{-1}\sigma\sqrt{\log d/n} + \delta H$. Then with probability $1 - O(d^{-2})$*

$$\widetilde{g}_t(\delta) = g_t + \frac{\delta}{2}\mathbb{E}\left[(z^\top H_t z)z\right] + \widetilde{\zeta}_t(\delta) + \widetilde{\beta}_t(\delta) + \widetilde{\gamma}_t(\delta),$$

where $g_t = \nabla f(x_t)$, $H_t = \nabla^2 f(x_t)$; for any $a \in \mathbb{R}^d$, $\langle\widetilde{\zeta}_t(\delta), a\rangle$ conditioned on x_t is a centered d-dimensional sub-exponential random variable with parameters $\nu^2 = \sqrt{n/2} \cdot \alpha$ and $\alpha \lesssim \sigma\|a\|_2/\delta n$; $\langle\widetilde{\beta}_t(\delta), a\rangle$ conditioned on x_t is a centered d-dimensional sub-Gaussian random variable with parameter $\nu \lesssim \delta H\|a\|_1/\sqrt{n}$; $\gamma_t(\delta)$ is a d-dimensional vector that satisfies

$$\|\widetilde{\gamma}_t(\delta)\|_\infty \lesssim L\delta^2 + \frac{\sigma s \log d}{n\delta} + s\delta H\sqrt{\frac{\log d}{n}}.$$

Note that $\widetilde{\zeta}_t(\delta)$ and $\widetilde{\beta}_t(\delta)$ might be correlated conditioned on x_t. Comparing Lemma 15 with Lemma 14, we observe that the bias term $\widetilde{\gamma}_t(\delta)$ is significantly smaller ($O(\delta^2)$ instead of $O(\delta)$); while the second term $\frac{\delta}{2}\mathbb{E}[(z^\top H_t z)z]$ is still a bias term with non-zero mean, it only depends on δ and can be easily removed. This motivates the following definition of a "twice de-biased" gradient estimator:

The twice de-biased estimator:

$$\widetilde{g}_t^{\text{tw}} := 2\widetilde{g}_t(\delta/2) - \widetilde{g}_t(\delta). \tag{3.25}$$

Corollary 2. *Suppose the conditions in Lemma 15 are satisfied. Then with probability $1 - O(d^{-2})$,*

$$\widetilde{g}_t^{\text{tw}} - g_t = \widetilde{\zeta}_t + \widetilde{\beta}_t + \widetilde{\gamma}_t,$$

where $\widetilde{\zeta}_t = 2\widetilde{\zeta}_t(\delta/2) - \widetilde{\zeta}_t(\delta)$, $\widetilde{\beta}_t = 2\widetilde{\beta}_t(\delta/2) - \widetilde{\beta}_t(\delta)$ and $\widetilde{\gamma}_t = \widetilde{\gamma}_t(\delta/2) - \widetilde{\gamma}_t(\delta)$.

The twice de-biased estimator is, in principle, similar to the "twicing" trick in nonparametric kernel smoothing (Newey et al., 2004) that reduces estimation bias. In particular, Corollary 2 shows that the $\frac{\delta}{2}\mathbb{E}[(z^\top H_t z)z]$ bias term is cancelled by the "twicing" trick, and the remaining bias term $\widetilde{\gamma}$ is an order of magnitude smaller than γ in the bias term before twicing (e.g., Lemma 14). We also remark that the twice de-biased estimator $\widetilde{g}_t^{\text{tw}}$ does *not* significantly increase the computational burden, because the method remains first-order and only (two copies of) the de-biased gradient estimate needs to be computed.

Plugging the "twice" de-biased gradient estimator $\widetilde{g}_t^{\text{tw}}$ into the stochastic mirror descent procedure and choosing tuning parameters n, λ, δ and η appropriately, we obtain the following improved convergence rate:

Theorem 8. *Suppose (A1) through (A4) and (A6) hold. Suppose also that $T = \Omega(s^3 \log^2 d + (1 + L)^2 s^2 + H^2 B^2(1 + L)s \log d)$ and $T \leqslant d$. Let $\eta := Bn^{2/3}\sqrt{\frac{\log d}{T}}$, $n := \lfloor(1 + L)s^{2/3}\sqrt{T}\rfloor$ and $\delta := (s \log d/n)^{1/3}$. Then with probability $1 - O(d^{-1})$*

$$\mathbb{E}f(\widehat{x}_n) - f(x^*) \lesssim \widetilde{\xi}_{\sigma,s}B\sqrt{\log d}\left(\frac{(1 + L)s^{2/3}}{T}\right)^{1/3} + \widetilde{O}(T^{-5/12}),$$

(a) $s = 5, d = 100$　　　(b) $s = 10, d = 100$　　　(c) $s = 20, d = 100$
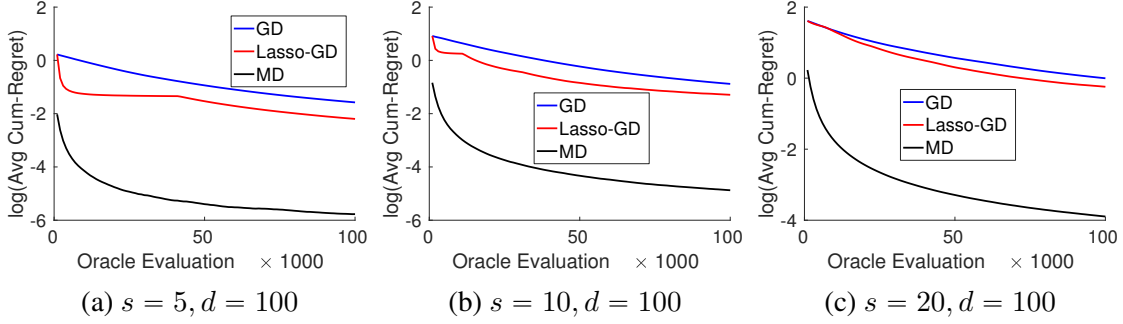
Figure 3.2: Sparse quadratic optimization with identity quadratic term.

(a) decay rate $= 1.5$　　　(b) decay rate $= 2$　　　(c) decay rate $= 3$
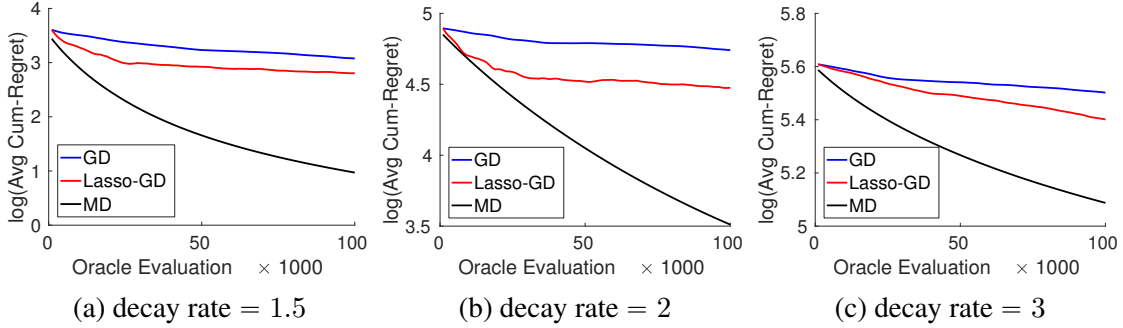
Figure 3.3: Sparse quadratic optimization with polynomial decay of eigenvalues.

where $\widetilde{\xi}_{\sigma,s} = (1 + \sigma + \sigma^2/s^{2/3})$ and $\widehat{x}_n$ is the average of all $\{x_t\}_{t=1}^{T'}$ with T' being the number of epochs in which $2m$ design points are constructed.

As a simple illustration consider the following example:

Example 3. Consider a quadratic function $f(x) = \frac{1}{2}(x - x^*)^\top Q(x - x^*)$ with (unknown) $Q \succeq 0$ being positive semi-definite and supported on $S \subseteq [d]$ with $|S| \leqslant s$, meaning that $Q_{ij} = 0$ if $i \notin S$ or $j \notin S$. It is easy to verify that f satisfies (A1) through (A5), and also (A6) with $L = 0$ because $\nabla^2 f(x) \equiv Q$, independent of x. Subsequently, applying results in Theorem 8 we obtain a convergence rate of $O(T^{-1/3})$.

More broadly, compared to Theorem 7, the stochastic mirror descent algorithm with the twice de-biased gradient estimator ($\widetilde{g}_t^{\mathrm{tw}}$) has the convergence rate of $O(T^{-1/3})$, which is a strict improvement over the $O(T^{-1/4})$ rate in Theorem 7. Such improvement is at the cost of the additional assumption of Hessian smoothness (A6); however, the optimization algorithm remains almost unchanged and no second-order information is required at runtime.

3.2.5 Numerical results

We compare our proposed algorithm with the baseline method on synthetic function examples, including low-dimensional zeroth-order optimization (proposed in (Flaxman et al., 2005)) as well as the intuitive method of first doing Lasso support selection and then low-dimensional zeroth-order optimization on the selected variables. We use GD to represent "zeroth order" gradient descent algorithm proposed in (Flaxman et al., 2005), Lasso-GD to represent the model-

(a) $s = 5, d = 100$ (b) $s = 10, d = 100$ (c) $s = 20, d = 100$
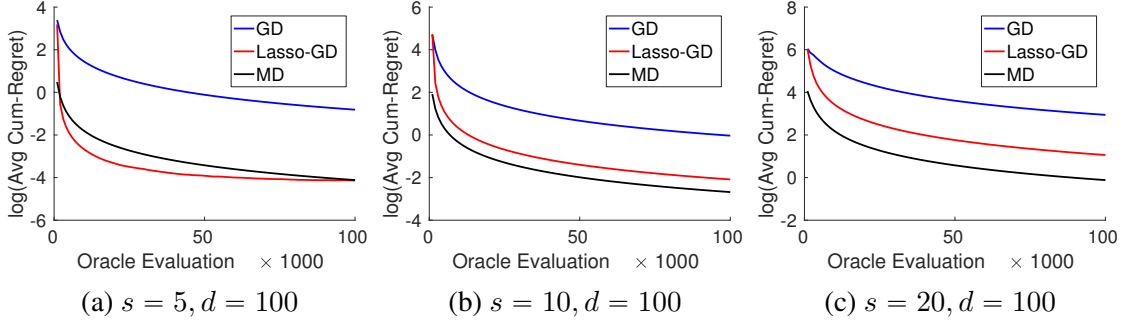
Figure 3.4: Sparse fourth-degree polynomial optimization with identity quadratic term.

selection-then-optimize approach, and MD to represent the zeroth-order mirror descent algorithm with sparse gradient estimates. For our synthetic function examples, we first construct a convex low-dimensional function $f_S : \mathbb{R}^{|S|} \to \mathbb{R}$ on a uniformly chosen subset $S \subseteq [d]$ with size s, and then "extend" f_S to f defined on the high-dimensional domain $\mathbb{R}^d$ by $f(x) \equiv f_S(x_S)$. Functions constructed as such naturally satisfy the sparsity assumptions (A3), (A4) and (A5). In all plots we start at the 1000th iterations (oracle evaluations) of all algorithms to avoid clutter caused by the volatile burn-in phases. Thus, the starting points in the plots are slightly different for different algorithms.

In Figure 3.2 we consider sparse quadratic optimization problem with $f_S(x_S) = x_S^\top Q x_S + b^\top x_S$ where we set $Q_{ii} = 1$ and $b_i = 1$ for $i \in S$ and other entries to 0. In Figure 3.3 we consider sparse quadratic optimization problem with $f_S(x_S) = x_S^\top Q x + b^\top x_S$ where we set $Q_{ii} = i^{-\gamma}$ where γ is the eigenvalue decay rate and $b_i = 1$ for $i \in S$ and other entries to 0. In Figure 3.4 we consider sparse degree-4 polynomial optimization problem with $f_S(x) = |(x_S - b)^\top Q(x_S - b)|^2 + (x_S - b)^\top Q(x_S - b)$ where we set $Q_{ii} = 1$ and $b_i = 1$ for $i \in S$ and other entries to 0. All hyper-parameters are tuned by grid search. The cumulative optimization error $\frac{1}{t} \sum_{t'=0}^{t-1} f(x_t) - f^*$ is reported for all algorithms and selected time epochs $t \leqslant n$.

We observe that in all our simulation settings, the vanilla gradient descent algorithm is dominated by our proposed algorithms. Our simulation results also suggest that the mirror descent algorithm is superior to the successive component selection algorithm. MD is also easier to use in practice as it has fewer parameters. Thus, we recommend mirror descent algorithm for practical use.

3.2.6 Extension to $\ell_{\mathfrak{p}}$ geometry: the unconstrained case

High-dimensional derivative-free stochastic optimization arises in many scientific and engineering applications. While most of the time additional structural assumptions on the objectives or the optimal solutions do exist, exact sparsity conditions could be too strong to hold in many cases. In this section we discuss how the zeroth-order mirror descent framework could be extended to cases where conditions weaker than the sparsity structural assumptions are imposed.

For $\mathfrak{p} \in (1, 2]$ and $x \in \mathbb{R}^d$, let $\|x\|_{\mathfrak{p}} := (\sum_{i=1}^d |x_i|^{\mathfrak{p}})^{1/\mathfrak{p}}$ denote the vector-$\ell_{\mathfrak{p}}$ norm of x, and $\mathfrak{q} = 1/(1 - 1/\mathfrak{p})$ be the dual norm of $\ell_{\mathfrak{p}}$ (if $\mathfrak{p} = 1$ then define $\mathfrak{q} = \infty$). The following conditions are imposed upon the objective function $f : \mathcal{X} \to \mathbb{R}$ as well as its minimizer x^*:

54

(B1) (*Unconstrained optimization*): $\mathcal{X} = \mathbb{R}^d$;

(B2) (*Bounded minima*): there exists $x^* \in \mathcal{X}$ such that $f(x^*) = \min_{x \in \mathcal{X}} f(x)$ and $\|x^*\|_{\mathfrak{p}} \leqslant B$;

(B3) (*Bounded gradients*): f is differentiable on $\mathbb{R}^d$ and furthermore $\sup_x \|\nabla f(x)\|_{\mathfrak{q}} \leqslant L$.

The constraint $\|x^*\|_{\mathfrak{p}} \leqslant B$ for $\mathfrak{p} \in [1, 2]$ is a weaker form of constraining the optimal solution x^* to be sparse. More specifically, as the norm $\mathfrak{p}$ moves from 2 to 1, the optimal solution x^* has to be sparser in order to satisfy the more stringent constraint $\|x^*\|_{\mathfrak{p}} \leqslant B$. Nevertheless, the condition is considerably weaker than the exact sparsity constraint (e.g., $\|x^*\|_0 \leqslant s$), as the minimizer x^* itself could still be dense and spread across all its components.

Our algorithmic framework is the follows:

1. First, we construct a *smoothed* version $\widetilde{f}(x) := \mathbb{E}_{u \sim \mu}[f(x + u)]$, where μ is a distribution supported on $\mathcal{X} = \mathbb{R}^d$. The smoothed function $\widetilde{f}$ is constructed so that $|\widetilde{f}(x) - f(x)|$ is small, and furthermore the gradient $\nabla \widetilde{f}(x)$ can be *unbiasedly* estimated;

2. At each iteration t, the algorithm observes $y_t = f(x_t + v_t)$ for some random variable v_t and construct an *unbiased* estimator $\widehat{g}_t \in \mathbb{R}^d$ of $g_t := \nabla \widetilde{f}(x_t)$ such that $\mathbb{E}\widehat{g}_t = g_t$;

3. The mirror descent update is preformed:

$$x_{t+1} \in \arg\min_{x \in \mathcal{X}} \{\eta_t \langle \widehat{g}_t, x \rangle + D_\psi(x, x_t)\}, \tag{3.26}$$

where $\{\eta_t\}$ are the step sizes, and $D_\psi(x, y) = \psi(x) - \psi(y) - \langle \nabla \psi(y), x - y \rangle$ is a Bregman divergence with respect to a potential function ψ, which is κ-strongly convex with respect to $\| \cdot \|_{\mathfrak{p}}$ (i.e., $D_\psi(x, y) \geqslant \kappa \|x - y\|_{\mathfrak{p}}^2 / 2$).

Let p be the density of μ. The gradient estimates $\widehat{g}_t$ are constructed via the celebrated *Stein-Hudson identity* ([Hudson, 1978](); [Stein, 1981]()) which asserts that, under minimal regularity conditions, for any differentiable $h : \mathbb{R}^d \to \mathbb{R}$,

$$\mathbb{E}\left[h(u)\nabla \log p(u)\right] = -\mathbb{E}\left[\nabla h(u)\right] \qquad \text{where } u \sim \mu. \tag{3.27}$$

Consider $h(u) \equiv f(x + u)$ and recall that $\widetilde{f}(x_t) = \mathbb{E}_\mu[f(x_t + u)]$. Eq. (3.27) then reduces to

$$\nabla \widetilde{f}(x_t) = \mathbb{E}_\mu\left[f(x_t + u)\nabla \log p(u)\right]. \tag{3.28}$$

Hence, a natural estimator of $g_t = \nabla \widetilde{f}(x_t)$ is

$$\widehat{g}_t := (y_t - y_t')\nabla \log p(u_t) \quad \text{where } y_t = f(x_t) + \xi_t, \ y_t' = f(x_t + u_t) + \xi_t', \ u_t \sim \mu. \tag{3.29}$$

Remark 13. The y_t term acts like a *control variate* to reduce variance and to avoid dependency on the magnitude of $|f(x_t)|$.

Let $\mathfrak{p} \in (1, 2]$ be the fixed norm parameter. Consider the generalized Gaussian distribution (see, e.g., [Song & Gupta (1997)](); [Toulias & Kitsos (2014)]())

$$p(x) = p(x_1, \cdots, x_d) = \prod_{i=1}^{d} p_i(x_i) \quad \text{where } p_i(x_i) = \frac{\mathfrak{p}^{1-1/\mathfrak{p}}}{2\delta\Gamma(1/\mathfrak{p})} \exp\left\{-\frac{|x_i|^{\mathfrak{p}}}{\mathfrak{p}\delta^{\mathfrak{p}}}\right\}. \tag{3.30}$$

When $\mathfrak{p} \in (1, 2]$, $\log p$ is differentiable everywhere, and the gradient of $\log p$ takes the form of $-\nabla \log p(x) = \widetilde{x}/\delta^{\mathfrak{p}}$, where $\widetilde{x} = [\mathrm{sgn}(x_1)|x_1|^{\mathfrak{p}-1}, \cdots, \mathrm{sgn}(x_d)|x_d|^{\mathfrak{p}-1}]$. The gradient estimate $\widetilde{g}_t$ can then be written as

$$\widehat{g}_t = (y_t' - y_t)\delta^{-\mathfrak{p}}\widetilde{u}_t \quad \text{where} \quad \widetilde{u}_t = [\mathrm{sgn}(u_{t1})|u_{t1}|^{\mathfrak{p}-1}, \cdots, \mathrm{sgn}(u_{td})|u_{td}|^{\mathfrak{p}-1}]. \quad (3.31)$$

The following lemmas establish several properties of the gradient estimate $\widetilde{g}_t$ (assuming (B1) through (B3) hold):

Lemma 16. *$\widehat{g}_t$ is an unbiased estimator of g_t, meaning that $\mathbb{E}\widehat{g}_t = g_t = \nabla\widetilde{f}(x_t)$.*

Lemma 17. *$|\widetilde{f}(x) - f(x)| \leqslant L\mathbb{E}_\mu[\|u\|_{\mathfrak{p}}] \lesssim L(1 + \delta)\delta d^{1/\mathfrak{p}} \log d$ for all $x \in \mathcal{X}$.*

Lemma 18. *$\mathbb{E}_\mu\|\widehat{g}_t\|_*^2 \leqslant \mathbb{E}_\mu[(2\sigma^2 + L^2\|u\|_{\mathfrak{p}}^2)\|\nabla \log p(u)\|_{\mathfrak{q}}^2] \lesssim \sigma^2(1 + \delta)\delta^{-2}d^{2/\mathfrak{q}} \log^2 d + L^2(1 + \delta)d^2 \log^4 d$.*

Combining Lemmas 16, 17 and 18, we arrive at the following theorem on the convergence rate of the zeroth-order mirror descent with Stein-Hudson gradient estimates, when step sizes ($\{\eta_t\}$) and probing radius (δ) are carefully selected:

Theorem 9. *Suppose (B1) through (B3) hold, and the parameters $\eta_t \equiv \eta$ and δ are selected as $\delta \asymp (\sigma^2 B/L^2 n)^{1/4} \cdot d^{1/2 - 1/\mathfrak{p}}$, $\eta \asymp \sqrt{B/(n \times (\sigma^2\delta^{-2}d^{2/\mathfrak{q}} \log^2 d + L^2 d^2 \log^4 d))}$. Then for sufficiently large n, the average $\widehat{x}_n = (\sum_{t=1}^n x_t)/n$ satisfies*

$$\mathbb{E}f(\widehat{x}_n) - f(x^*) = \widetilde{O}\left(\left[\frac{\sigma^2 d^2 BL}{n}\right]^{1/4} + \sqrt{\frac{d^2 BL^2}{n}}\right). \quad (3.32)$$

Remark 14. Theorem 9 holds when the origin is taken as the initial point (i.e., $x_0 = 0$), and n is sufficiently large such that $\delta \leqslant 1$. Also, in the $\widetilde{O}(\cdot)$ notation we omit poly-logarithmic dependency on d.

Remark 15. Regardless of the values of $\mathfrak{p}$ and $\mathfrak{q}$ appearing in the conjugate norms that define the boundedness of x^* and ∇f, Eq. (3.32) has the same dependency on domain dimension d.

Remark 16. In the noiseless case $\sigma = 0$ much better convergence rate is reflected in Eq. (3.32); i.e., $n^{-1/2}$ instead of $n^{-1/4}$. This is similar to the "two-point query" models studied in the literature (Agarwal et al., 2010; Duchi et al., 2015; Shamir, 2017) which were known to yield faster convergence rates for the zeroth-order optimization problem.

3.2.7 Extension to $\ell_{\mathfrak{p}}$ geometry: the constrained case

One disadvantage of the Stein-Hudson's gradient estimator is that the support of the probing distribution μ spans the entire $\mathbb{R}^d$ domain, making it applicable only in unconstrained optimization. While certain truncation arguments could be applied when constraints are present, such approaches are quite messy and difficult to analyze.

To overcome such difficulties, in this section we consider alternative gradient estimates for $\ell_{\mathfrak{p}}$ geometry whose probing distribution is supported on a compact set, and are therefore more appropriate for constrained optimization when the optimal solution x^* is not too close from the boundary.

More specifically, let $K \subset \mathbb{R}^d$ be a non-empty, compact symmetric [4] convex set in $\mathbb{R}^d$. Let ν_K and $\sigma_{\partial K}$ be the uniform measure on K and its boundary ∂K, respectively. For any convex, differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, a "smoothed version" $\widetilde{f} : \mathbb{R}^d \to \mathbb{R}$ is defined as

$$\widetilde{f}(x) := \mathbb{E}_{u \sim \nu_K} \left[f(x + \delta u) \right], \tag{3.33}$$

where $\delta > 0$ is a tuning parameter. Define also $\rho(K) := \mathrm{vol}_{d-1}(\partial K)/\mathrm{vol}_d(K)$ and $\ell(v) \in \mathbb{R}^d$ the outer normal vector at v on ∂K. The gradient estimate $\widehat{g}_t$ of $\nabla \widetilde{f}(x_t)$ is then defined as

$$\widehat{g}_t(x_t) = \delta^{-1} \rho(K) f_t(z_t) \ell(v_t) \quad \text{where } z_t = x_t + \delta v_t, v_t \sim \sigma_{\partial K}. \tag{3.34}$$

For fixed $\mathfrak{p} \in (1, 2]$ norm parameter, the convex body K is taken to be the unit $\ell_{\mathfrak{p}}$ ball $\mathbb{B}_{\mathfrak{p}}^d = \{x \in \mathbb{R}^d : \|x\|_{\mathfrak{p}} \leqslant 1\}$. For such K, the outer normal vector $\ell(v)$ is well-defined for all $v \in \partial K$, taking the form of

$$[\ell(v)]_i = \mathrm{sgn}(v_i) \cdot |v_i|^{\mathfrak{p}-1} / \|v\|_{2(\mathfrak{p}-1)}^{\mathfrak{p}-1} \quad \text{where } \|v\|_{2(\mathfrak{p}-1)}^{\mathfrak{p}-1} = \sqrt{\sum_{j=1}^d |v_j|^{2(\mathfrak{p}-1)}}.$$

The following lemmas establish several properties of the gradient estimate $\widehat{g}_t$, as well as the approximation error of $\widetilde{f}$ in terms of the probing convex body K.

Lemma 19. *For any $x_t \in \mathbb{R}^d$, $\nabla \widetilde{f}(x_t) = \mathbb{E}\widehat{g}_t(x_t)$.*

Lemma 20. *For any $x \in \mathbb{R}^d$, $\mathfrak{p} \in (1, 2]$ and $\mathfrak{q} = 1/(1 - 1/\mathfrak{p})$, $|\widetilde{f}(x) - f(x)| \leqslant \delta L \cdot \mathbb{E}_{u \sim \nu(K)}[\|u\|_{\mathfrak{p}}]$.*

Combining Lemmas 19, 20 and using the potential function $\psi(x) = \frac{1}{\mathfrak{p}-1} \|x\|_{\mathfrak{p}}^2$, by standard analysis of mirror descent algorithms ((Agarwal et al., 2012; Beck & Teboulle, 2003; Yudin & Nemirovskii, 1983), see also Eq. (3.138)), we have the following result:

Lemma 21. *Suppose (B2), (B3) hold and $|f(x)| \leqslant C$ for all $x \in \mathcal{X}$. Suppose also that $\|x^*\|_{\mathfrak{p}} \leqslant b$, and define $\widehat{x}_n = (\sum_{t=1}^n x_t)/n$. Then*

$$\mathbb{E}f(\widehat{x}_n) - f(x^*) \lesssim \frac{b}{\eta n} + \eta \cdot \max_{t \leqslant n} \sqrt{\mathbb{E}\|\widehat{g}_t(x_t)\|_{\mathfrak{q}}^2} + \delta L \cdot \mathbb{E}_{u \sim \nu(K)}[\|u\|_{\mathfrak{p}}]$$

$$\lesssim \frac{b}{\eta n} + \eta \cdot \frac{\rho(K)C}{\delta} \sqrt{\mathbb{E}_{\sigma_{\partial K}}[\|\ell(v)\|_{\mathfrak{q}}^2]} + \delta L.$$

From Lemma 21, it remains to upper bound $\rho(K)$ and $\mathbb{E}_{\sigma_{\partial K}}[\|\ell(v)\|_{\mathfrak{q}}^2]$ as well as the selection of appropriate values of η and δ. This is accomplished by the following two key lemmas:

Lemma 22. *For any $p \in (1, \infty)$, we have*

$$\lim_{d \to \infty} \frac{\rho(\mathbb{B}_{\mathfrak{p}}^d)}{d^{1/2+1/\mathfrak{p}}} = \sqrt{\frac{\kappa(\mathfrak{p}, 2(\mathfrak{p}-1))}{\kappa(\mathfrak{p}, \mathfrak{p})^{2(\mathfrak{p}-1)/\mathfrak{p}}}} \quad \text{where } \kappa(\mathfrak{p}, r) := \frac{\Gamma((1+r)/\mathfrak{p})}{\Gamma(1/\mathfrak{p})}. \tag{3.35}$$

---

[4] A convex set $K \subset \mathbb{R}^d$ is symmetric if for all $x \in \mathbb{R}^d$, $x \in K \iff -x \in K$.

Lemma 23. *For any pair of conjugate norms* $\mathfrak{p}, \mathfrak{q} \in (1, \infty)$, $1/\mathfrak{p} + 1/\mathfrak{q} = 1$, *there exists a finite positive constant* $C_\mathfrak{p}$ *such that for sufficiently large* d, *for* $K = \mathbb{B}_\mathfrak{p}^d$,

$$\mathbb{E}_{v \sim \sigma_{\partial K}}[\|\ell(v)\|_\mathfrak{q}^2] \leqslant C_\mathfrak{p} \cdot d^{1/\mathfrak{q} - 1/\mathfrak{p}}. \tag{3.36}$$

Combining Lemmas 21, 22 and 23, we arrive at the following theorem:

Theorem 10. *Under conditions (B2), (B3) and the assumptions that* $|f(x)| \leqslant C$ *for all* $x \in \mathcal{X}$ *and* $x^* + v \in \mathcal{X}$ *for all* $v \in \delta\mathbb{B}_\mathfrak{p}^d$, *if* $\eta \asymp \sqrt{B/nL^2}$ *and* $\delta \asymp \sqrt{BCd/L} \cdot n^{-1/4}$, *we have*

$$\mathbb{E}f(\widehat{x}_n) - f(x^*) \lesssim \sqrt{BLd} \cdot n^{-1/4}.$$

3.3 Non-stationary optimization with local variation criteria

In this section we consider a *non-stationary* setting of nonparametric optimization, in which the underlying function f is allowed to slightly change over time.

More specifically, at each time epoch t there is a different unknown function $f_t : \mathcal{X} \to \mathbb{R}$ to be optimized, associated with its minimizer $x_t^* \in \arg\min_{x \in \mathcal{X}} f_t(x)$. At each time epoch t a policy (algorithm) π queries a specific point $x_t \in \mathcal{X}$ and receives noisy feedback $f_t(x_t) + \xi_t$, where $\xi_t \sim \mathcal{N}(0, 1)$. The objective is to minimize the *dynamic regret* (or *strong regret*) of $\{x_t\}_{t=1}^T$:

$$R^\pi(f) := \mathbb{E}^\pi \sum_{t=1}^T f_t(x_t) - f_t(x_t^*). \tag{3.37}$$

Apart from classical convexity and smoothness assumptions, additional conditions are required to constrain the changes between neighboring objectives f_t and f_{t+1} to make the dynamic regret minimization problem feasible.

3.3.1 Backgrounds and assumptions

Apart from $\mathcal{X}$ being closed convex and $f_1, \cdots, f_T$ being convex and differentiable, we also make the following additional assumptions on the domain $\mathcal{X}$ and functions $f_1, \cdots, f_T$:

(A1) (*Bounded domain*): there exists constant $D > 0$ such that $\sup_{x,x' \in \mathcal{X}} \|x - x'\|_2 \leqslant D$;

(A2) (*Bounded function and gradient*): there exists constant $H > 0$ such that $\sup_{x \in \mathcal{X}} |f_t(x)| \leqslant H$ and $\sup_{x \in \mathcal{X}} \|\nabla f_t(x)\|_2 \leqslant H$;

(A3) (*Non-empty interior*): the interior of $\mathcal{X}$ is non-empty; that is, $\mathcal{X}^o \neq \varnothing$;

(A4) (*Smoothness*): there exists constant $L > 0$ such that $f_t(x') \leqslant f_t(x) + \nabla f_t(x)^\top (x' - x) + \frac{L}{2}\|x' - x\|_2^2$ for all $x, x' \in \mathcal{X}$.

(A5) (*Strong convexity*): there exists constant $M > 0$ such that $f_t(x') \geqslant f_t(x) + \nabla f_t(x)^\top (x' - x) + \frac{M}{2}\|x' - x\|_2^2$ for all $x, x' \in \mathcal{X}$.

The assumptions (A1), (A2) are standard assumptions that were imposed in previous works on both stationary and non-stationary stochastic optimization (Agarwal et al., 2013; Besbes et al., 2015; Flaxman et al., 2005; Shamir, 2017). The conditions (A4) and (A5) concern second-order properties of f_t and enable smaller regret rates for gradient descent algorithms. We note that the condition $MI_d \leq \nabla^2 f_t(x) \leq LI_d, \forall x \in \mathcal{X}$ in (Besbes et al., 2015) (see Eq. (10) in (Besbes et al., 2015)) is stronger and implies our (A4) and (A5) since we do not assume that f_t is twice differentiable. We also consider parameters D, H, L, M in (A1)–(A5) and domain dimensionality d as constants throughout the paper and omit their (polynomial) multiplicative dependency in regret bounds.

3.3.2 Local variation criteria

We generalize the results of (Besbes et al., 2015) so that local spatial and temporal changes of smooth and strongly convex function sequences are taken into consideration. For any measurable function $f : \mathcal{X} \to \mathbb{R}$, define

$$\|f\|_p := \begin{cases} \left(\frac{1}{\mathrm{vol}(\mathcal{X})} \int_{\mathcal{X}} |f(x)|^p \mathrm{d}x \right)^{1/p} & 1 \leq p < \infty; \\ \sup_{x \in \mathcal{X}} |f(x)| & p = \infty. \end{cases} \tag{3.38}$$

Here $\mathrm{vol}(\mathcal{X}) = \int_{\mathcal{X}} 1 \mathrm{d}x$ is the Lebesgue measure of the domain $\mathcal{X}$ and is finite because of the compactness of $\mathcal{X}$. We shall refer to $\|f\|_p$ as the L_p-norm of f in the rest of this paper. (Conventionally in functional analysis the L_p norm of a function is defined as the *unnormalized* integration $\left(\int_{\mathcal{X}} |f(x)|^p \mathrm{d}x \right)^{1/p}$. Nevertheless, we adopt the volume normalized definition in this paper for convenience. The Minkowski's inequality $\|f + g\|_p \leq \|f\|_p + \|g\|_p$, as well as other basic properties of L_p norm, remain valid.) Also, for a sequence of convex functions $f_1, \cdots, f_T : \mathcal{X} \to \mathbb{R}$, define the $L_{p,q}$-*variation functional* of $f = (f_1, \cdots, f_T)$ as

$$\mathrm{Var}_{p,q}(f) := \begin{cases} \left(\frac{1}{T} \sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_p^q \right)^{1/q} & 1 \leq p \leq \infty, 1 \leq q < \infty; \\ \sup_{1 \leq t \leq T-1} \|f_{t+1} - f_t\|_p & 1 \leq p \leq \infty, q = \infty. \end{cases} \tag{3.39}$$

Note that in both Eqs. (3.38) and (3.39) we restrain ourselves to convex norms $p \geq 1$ and $q \geq 1$. We can then define function classes

$$\mathcal{F}_{p,q}(V_T) := \{ f : \mathrm{Var}_{p,q}(f) \leq V_T \}, \tag{3.40}$$

which serves as the budget constraint for a function sequence f. The definition of $\mathcal{F}_{p,q}$ is more general than $\mathcal{F}_{\infty,1}$ introduced in (Besbes et al., 2015) since it better reflects the spatial and temporal locality of f in the subscripts p and q.

Example 4 (spatial locality). Let $\mathcal{X} = [0, 1]$. Consider univariate piecewise cubic spline functions $f_t(x) = \sum_{i=1}^m \mathbb{I}[x \in \mathcal{X}_i] \cdot (a_{ti}x^3 + b_{ti}x^2 + c_{ti}x + d_{ti})$, where $\mathbb{I}[\cdot]$ is the indicator function, $\{\mathcal{X}_i\}_{i=1}^m$ is a uniform partition of $\mathcal{X} = [0, 1]$ (i.e., $\mathcal{X}_i = \left[\frac{i-1}{m}, \frac{i}{m} \right]$) and $\{a_{ti}, b_{ti}, c_{ti}, d_{ti}\}_{i=1}^m$ are selected such that f_t is strongly convex and sufficiently smooth. Also suppose that f_t and f_{t+1} differ only on *two* neighboring pieces $\mathcal{X}_i \cup \mathcal{X}_{i+1}$, and the difference on $\mathcal{X}_i \cup \mathcal{X}_{i+1}$ between f_t and f_{t+1} is uniformly bounded. Formally, $f_t(x) = f_{t+1}(x)$ for all $x \in \mathcal{X} \backslash (\mathcal{X}_i \cup \mathcal{X}_{i+1})$ and

$\sup_{x \in \mathcal{X}_i \cup \mathcal{X}_{i+1}} |f_t(x) - f_{t+1}(x)| \leqslant \delta < \infty$. We then have that (noting that $\text{vol}(\mathcal{X}) = 1$ and $\text{vol}(\mathcal{X}_i \cup \mathcal{X}_{i+1}) = 2/m$)

$$\|f_t - f_{t+1}\|_p = \left(\int_{\mathcal{X}_i \cup \mathcal{X}_{i+1}} |f_t(x) - f_{t+1}(x)|^p \mathrm{d}x \right)^{1/p} \leqslant \delta \cdot (m/2)^{-1/p}.$$

Because $\delta \cdot (m/2)^{-1/p}$ is an increasing function of p, the parameter p controls the *spatial locality* of function changes between epochs. For example, for $p = 1$ we have that $\|f_t - f_{t+1}\|_1 = 2\delta/m$ and for $p = \infty$ we have that $\|f_t - f_{t+1}\|_\infty = \delta$. Therefore, when the number of regions m is large, meaning that the changes in functions are local, $\|f_t - f_{t+1}\|_1$ is much smaller than $\|f_t - f_{t+1}\|_\infty$ and captures the concept of local spatial change of functions. In other words, when V_T is fixed the function class $\mathcal{F}_{1,q}(V_T)$ is *richer* (i.e., contains more functions) than $\mathcal{F}_{\infty,q}(V_T)$, meaning that more functions with local spatial variations are contained in $\mathcal{F}_{1,q}(V_T)$ compared to $\mathcal{F}_{\infty,q}(V_T)$.

Example 5 (temporal locality). Define $\delta_t = \|f_{t+1} - f_t\|_p$ for $t = 1, \cdots, T - 1$ to be the amount of change at epoch t. Suppose a total amount of Δ change of functions is fixed (i.e., $\sum_{t=1}^{T-1} \delta_t = \Delta$), and the changes are distributed uniformly across $s < T - 1$ epochs. That is, $\delta_i = \Delta/s$ for s epochs in $\{1, \cdots, T - 1\}$ and $\delta_i = 0$ for the other $T - 1 - s$ epochs. We then have that

$$\text{Var}_{p,q}(f) = \left(\frac{1}{T} \cdot s \cdot \delta^q \right)^{1/q} = \delta \cdot (s/T)^{1/q}.$$

Because $(s/T)^{1/q}$ is an increasing function of q, the parameter q controls the *temporal locality* of function changes. For example, for $q = 1$ we have that $\text{Var}_{p,1}(f) = \delta \cdot s/T$ and for $q = \infty$ we have $\text{Var}_{p,\infty}(f) = \delta$. Therefore, when the number of changes s is small compared to T, $\text{Var}_{p,1}(f)$ is much smaller than $\text{Var}_{p,\infty}(f)$ and captures the concept of local temporal change of functions.

3.3.3 Minimax upper bounds

We establish the following upper bounds on the worst-case regret of our designed policy, with details to be introduced later.

Theorem 11 (Upper bound). *Fix arbitrary $1 \leqslant p < \infty$ and $1 \leqslant q < \infty$. Suppose (A1) through (A5) hold and $0 \leqslant V_T \leqslant 1$. Then there exists a computationally efficient policy π and $C_1 > 0$ as a polynomial function of $\log T$ and $\log V_T$ such that*

$$\sup_{f \in \mathcal{F}_{p,q}(V_T)} R^{\pi'}(f) \leqslant C_1 \cdot T \cdot V_T^{2p/(6p+d)}.$$

Remark 17 (On the constant C_1). The dependency of C_1 on domain dimension d and variation parameters p, q is of the form $D^{d/2p}$, which arises from our main affinity technical lemma (Lemma 44). Since d, D, and p are all treated as constants in our paper, the quantity $D^{d/2p}$ is also a constant. We remark that it does *not* depend on T or V_T, and thus will be much smaller than the main $T \cdot V_T^{2p/(6p+d)}$ term.

Remark 18 (On the term $\sqrt{T}$). The regret bound in Theorem 11 consists of two terms. The $\sqrt{T}$ term arise from regret bounds for *stationary* stochastic optimization problems (i.e., $V_T = 0$), which were proved in (Hazan & Kale, 2014; Jamieson et al., 2012). The other terms involving polynomial dependency on V_T are the main regret terms for typical dynamic function sequences whose perturbation V_T is not too small.

Remark 19 (The role of the parameter q). We remark that the q parameter does *not* affect the optimal rate of convergence in Theorem 12 (provided that $q \geqslant 1$ is assumed for convexity of the norms). While this appears counter-intuitive, this is a property of our worst-case analytical framework, as the function sequence that leads to the worst-case regret is the one that distributes function changes $\|f_{t+1} - f_t\|_p$ evenly across all $t \in \mathcal{T}$ (see for example our detailed construction of adversarial function sequences in the online supplement), in which case the $L_{p,q}$-variation measure is the same for all $q \in [1, \infty]$.

Remark 20 (On the condition $V_T \leqslant 1$). The condition $V_T \leqslant 1$ in Theorem 11 is necessary for obtaining a non-trivial sub-linear regret. In particular, our lower bound results will show that for $V_T = \Omega(1)$, no algorithm can achieve sub-linear regret in either feedback models (see Theorem 12 in the lower bound section). On the other hand, a trivial algorithm that outputs $x_1 = \cdots = x_T = x_0$ for an arbitrary $x_0 \in \mathcal{X}$ leads to a linear regret.

Remark 21 (Curse of dimensionality). A significant difference between $p = \infty$ and $p < \infty$ settings is the *curse of dimensionality*. In particular, when $p < \infty$ the (optimal) regret depends *exponentially* on dimension d, while for $p = \infty$ the dependency on V_T is independent of d on the exponent. The curse of dimensionality is a well-known phenomenon in non-parametric statistical estimation (Tsybakov, 2009).

Remark 22 (Comparing with Besbes et al. (2015)). Besbes et al. (2015) considered the special case of $p = \infty$ and $q = 1$, and established the following result:

$$\inf_{\pi \in \mathcal{P}_T^\pi} \sup_{f \in \mathcal{F}_{p,q}(V_T)} R_\phi^\pi(f) \asymp T \cdot V_T^{1/3} \qquad \text{for} \quad p = \infty, q = 1. \tag{3.41}$$

Note that in Eq. (3.41) we adopt a slightly different notation from Besbes et al. (2015). In particular, the parameter V_T in our paper is $1/T$ times the parameter V_T in (Besbes et al., 2015). Such normalization is for presentation clarity only (to single out the T term in the regret bounds).

It is clear that our results reduce to Eq. (3.41) as $p \to \infty$. In particular, for fixed domain dimension d we have that $\lim_{p \to \infty} 2p/(6p + d) = 1/3$, matching regrets in Eq. (3.41). Therefore, the result from Besbes et al. (2015) (for strongly convex function sequences) is a special case of our results.

3.3.4 Policy design

There are two main components in our policy design:

1. A general restarting "meta-policy" from Besbes et al. (2015), where the interval/batch length Δ_T is tuned as a function of p.

2. Within each interval/batch of the meta-policy, a proper sub-policy π_s is invoked depending on the type of the feedback.

We first describe the "meta-policy" based on a re-starting procedure (Besbes et al., 2015):

> META-POLICY (RESTARTING PROCEDURE): input parameters T and Δ_T; sub-policy π_s.
> 1. Divide epochs $\{1, \cdots, T\}$ into $J = \lceil T/\Delta_T \rceil$ batches $B_1, \cdots, B_J$ such that $B_1 = \{\underline{b}_1, \cdots, \overline{b}_1\}$, $B_2 = \{\underline{b}_2, \cdots, \overline{b}_2\}$, etc., with $\underline{b}_1 = 1$, $\overline{b}_J = T$ and $\underline{b}_{\ell+1} = \overline{b}_\ell + 1$ for $\ell = 1, \cdots, J-1$. The epochs are divided as evenly as possible, so that $|B_\ell| \in \{\Delta_T, \Delta_T+1\}$ for all $\ell = 1, \cdots, J$.
> 2. For each batch B_ℓ, $\ell = 1, \cdots, J$, do the following:
>   (a) Run sub-policy π_s with $\underline{b}_\ell$ and $\overline{b}_\ell$, corresponding to $f_{\underline{b}_\ell}, f_{\underline{b}_\ell+1}, \cdots, f_{\overline{b}_\ell}$.

The key idea behind the meta-policy is to "restart" certain sub-policy π_s after Δ_T epochs. This strategy ensures that the sub-policy π_s has sufficient number of epochs to exploit feedback information, while at the same time avoids usage of outdated feedback. Scalings of Δ_T in the meta-policy is set as $\Delta_T \asymp V_T^{-4p/(6p+d)}$, which is motivated by our proof to our regret upper bound in Theorem 11.

In the rest of this section we describe the sub-policy π_s mentioned in the meta-policy. the classical approach is to first obtain an estimator of the gradient $\nabla f_t(x_t)$ by perturbing x_t along a random coordinate $e_j = (0, \cdots, 1, \cdots, 0) \in \mathbb{R}^d$. This idea originates from the seminal work of Yudin & Nemirovskii (1983) and was applied to convex bandits problems (e.g., Besbes et al. (2015); Flaxman et al. (2005)). Such an approach, however, fails to deliver the optimal rate of regret when the optimal solution x_t^* lies particularly close to the boundary of the domain $\mathcal{X}$. Here we describe a regularized ellipsoidal (RE) algorithm from Hazan & Levy (2014), which attains the optimal rate of regret even when x_t^* is very close to $\partial \mathcal{X}$.

The RE algorithm in Hazan & Levy (2014) is based on the idea of *self-concordant barriers*:

Definition 3 (self-concordant barrier). *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$ is convex and $\mathcal{X}^o \neq \varnothing$. A convex function $\varphi : \mathcal{X}^o \to \mathbb{R}$ is a κ-self-concordant barrier of $\mathcal{X}$ if it is three times continuously differentiable on $\mathcal{X}^o$ and has the following properties:*

1. For any $\{x_n\}_{n=1}^\infty \subseteq \mathcal{X}^o$, if $\lim_{n\to\infty} x_n \in \partial \mathcal{X}$ then $\lim_{n\to\infty} \varphi(x_n) = +\infty$.

2. For any $z \in \mathbb{R}^d$ and $x \in \mathcal{X}^o$ it holds that $|\nabla^3 \varphi(x)[z,z,z]| \leqslant 2|z^\top \nabla^2 \varphi(x)z|^{3/2}$ and $|z^\top \nabla \varphi(x)| \leqslant \kappa^{1/2}|z^\top \nabla^2 \varphi(x)z|^{1/2}$, where $\nabla^3 \varphi(x)[z,z,z] = \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} \varphi(x + t_1 z + t_2 z + t_3 z)\big|_{t_1=t_2=t_3=0}$.

It is well-known that for any convex set $\mathcal{X} \subseteq \mathbb{R}^d$ with non-empty interior $\mathcal{X}^o$, there exists a κ-self-concordant barrier function φ with $\kappa = O(d)$, and furthermore for bounded $\mathcal{X}$ the barrier φ can be selected such that it is strictly convex; i.e., $\nabla^2 \varphi(x) > 0$ for all $x \in \mathcal{X}^o$ (Boyd & Vandenberghe, 2004; Nesterov & Nemirovskii, 1994). For example, for linear constraints $\mathcal{X} = \{x : Ax \leqslant b\}$ with $A \in \mathbb{R}^{m \times d}$, a logarithmic barrier function $\varphi(x) = \sum_{i=1}^m -\log(b_i - a_i x)$ can be used to satisfy all the above properties (note that a_i denotes the i-th row of A).

We are now ready to describe the RE sub-policy that handles noisy function value feedback. The policy is similar to the algorithm proposed in Hazan & Levy (2014), except that *noisy* function value feedback is allowed in our policy, while Hazan & Levy (2014) considered only exact function evaluations. The analysis of our policy is also more involved for dealing with noise.

SUB-POLICY π_s (RE): input parameters $\underline{b}_\ell, \overline{b}_\ell$; constant step size η; self-concordant barrier φ;

1. Select $z_0 = \operatorname{argmin}_{y \in \mathcal{X}} \varphi(z)$;

2. For $t = 0$ to $\overline{b}_\ell - \underline{b}_\ell$ do the following:

   (a) Compute $A_t = (\nabla^2 \varphi(z_t) + \eta M(t+1)I_d)^{-1/2}$, where I_d is the identity matrix in $\mathbb{R}^{d \times d}$.

   (b) Sample u_t from the uniform distribution on the unit d-dimensional sphere $\mathbb{S}_d$.

   (c) Select $x_{\underline{b}_\ell + t} = z_t + A_t u_t$; suffer loss $f_{\underline{b}_\ell + t}(x_{\underline{b}_\ell + t})$ and obtain feedback $y_{\underline{b}_\ell + t} = f_{\underline{b}_\ell + t}(x_{\underline{b}_\ell + t}) + \xi_{\underline{b}_\ell + t}$.

   (d) Compute gradient estimate $\widehat{g}_t = d \cdot y_{\underline{b}_\ell + t} \cdot A_t^{-1} u_t$.

   (e) FTRL update: $z_{t+1} = \operatorname{argmin}_{z \in \mathcal{X}} \sum_{\tau=0}^{t} \left\{ \widehat{g}_\tau^\top z + \frac{M}{2} \|z - z_\tau\|_2^2 \right\} + \eta^{-1} \varphi(z)$.

In step 2(d), the gradient estimate $\widehat{g}_t = d \cdot y_{\underline{b}_\ell + t} \cdot A_t^{-1} u_t$ satisfies $\mathbb{E}[\widehat{g}_t] \approx \nabla f_{\underline{b}_\ell + t}(y_t)$ by the change-of-variable formula and the smoothness of $f_{\underline{b}_\ell + t}$. In step 2(e), instead of the projected gradient step, a Follow-The-Regularized-Leader (FTRL) step is executed to prevent y_{t+1} from being too close to the boundary of $\mathcal{X}$. The FTRL step is essentially a mirror descent, which uses a regularization term ($\varphi(\cdot)$ in our policy) and its associated Bregman divergence to improve the convergence rates of optimization algorithms measured in non-standard metric. It was shown in McMahan (2017) (Sec. 6) that the FTRL step is equivalent to mirror descent under minimal regularity conditions. Finally, step 2(c) is a random perturbation step originally considered in (Hazan & Levy, 2014). An important aspect of step 2(c) is the clever choice of the matrix A_t, which ensures the optimal regret bound even if the optimal solution x_t^* is very close to the boundary of $\mathcal{X}$. More specifically, the following proposition shows that $x_{\underline{b}_\ell + t} = z_t + A_t u_t$ always belongs to the domain $\mathcal{X}$, justifying the correctness of policy π_s.

Proposition 10. *Suppose φ is strictly convex on $\mathcal{X}^o$. Then for any $x \in \mathcal{X}^o$, $\delta \geqslant 0$ and $u \in \mathbb{S}_d$, $x + (\nabla^2 \varphi(x) + \delta I_d)^{-1/2} u \in \mathcal{X}$.*

3.3.5 Minimax lower bounds

We prove the following result, establishing a lower bound of worst-case regret possible for *any* policy π.

Theorem 12 (Lower bound). *Suppose the same conditions hold as in Theorem 11. Then there exists a constant $C_3 > 0$ independent of T and V_T such that*

$$\inf_{\pi} \sup_{f \in \mathcal{F}_{p,q}(V_T)} R^\pi(f) \geqslant C_3 \cdot T \cdot V_T^{2p/(6p+d)}.$$

Remark 23. The constant C_3 depends polynomially on d and p, q in our construction of adversarial problem instances. We again emphasize that this constant does *not* depend on either T or V_T.

3.4 Summary and related works

Optimizing a nonparametric function with access to (noisy) function evaluations only is in general a very important questions. With different regularity conditions imposed on the objective functions (which would certainly depend on the target application domains), the techniques, analysis and results also differ significantly. In this section, we present our results for low-dimensional smooth functions, high-dimensional convex functions and also dynamically changing convex functions.

Due to the very large volume of existing related works, it is certainly impossible to cover all of them. Below we give representative works on directions that are directly related to our results.

Derivative-free optimization Derivative-free optimization (also known as blackbox/zeroth-order optimization) is an extensively studied topic in mathematical optimization. Conn et al. (2009) proved global convergence of first- and second-order trust-region methods for derivative-free optimization. See also Audet & Dennis Jr (2006); Cartis et al. (2012); Conn et al. (2009); Dodangeh & Vicente (2016); Kolda et al. (2003); Torczon (1997); Vicente & Custódio (2012) for the study of other methods. Meanwhile, Bandeira et al. (2012, 2014); Billups et al. (2013); Chen et al. (2018a); Conn et al. (2008a,b); Powell (2003, 2004); Scheinberg & Toint (2010) considered trust-region and local probabilistic modeling methods for *high-dimensional* zeroth-order optimization problems, in which the problem dimension far exceeds the number of oracle queries but certain sparsity structures on gradients/hessians are expected. Ghadimi & Lan (2013b) considered non-convex objective functions under the zeroth-order optimization setting.

Bandit convex optimization In the literature of bandit convex optimization, the objective function sequence is subject to constant dynamic changes and only one or two evaluations for each function in the sequence is possible. The idea of using noisy function values subject to random perturbations to estimate function gradients first appeared in the seminal work of (Yudin & Nemirovskii, 1983) and was applied to bandit convex optimization problems in Flaxman et al. (2005). Agarwal et al. (2010) obtained improved convergence rates with additional strongly smooth/convex assumptions. Duchi et al. (2015); Shamir (2013, 2017) considered convex optimization/bandit problems with *noiseless* zeroth-order oracles and derived nearly matching convergence rates. Nesterov & Spokoiny (2017) studied the convergence rate under *noisy* zeroth-order oracles. Also, Hazan & Levy (2014) used an elliptical probing distribution to study the constrained zeroth-order optimization problem with optimal solution x^* very close to the boundary of feasible sets. Bubeck et al. (2017) proposed kernel based methods that attain optimal regret bounds for bandit convex optimization without strong convexity or smoothness assumptions. While most works consider *stationary* regret only, results on *dynamic* regret for bandit convex optimization also exist (Besbes et al., 2015).

Global or simulation optimization Traditionally, global optimization aims at finding the *global* optima of a multi-modal function, typically at the cost of an exponential number of queries/samples in domain dimensions. The question has a long history in the optimization research community

(Kan & Timmer, 1987a,b) and has also received a significant amount of recent interest in statistics and machine learning (Bubeck et al., 2011; Bull, 2011; Hazan et al., 2017; Malherbe et al., 2016; Malherbe & Vayatis, 2017; Rasmussen & Williams, 2006). Many previous works (Bubeck et al., 2011; Kleinberg, 2005) have derived rates for non-convex smooth payoffs in "continuum-armed" bandit problems. Grill et al. (2015); Minsker (2013) studies the problem of estimating the set of all optima of a smooth function in Hausdorff's distance. For Hölder smooth functions with polynomial growth, Minsker (2013) derives an $n^{-1/(2\alpha+d-\alpha\beta)}$ minimax rate for $\alpha < 1$ (later improved to $\alpha \geqslant 1$ in his thesis Minsker (2012)), Grill et al. (2015); Minsker (2013) also discussed adaptivity to unknown smoothness parameters. (Malherbe et al., 2016; Malherbe & Vayatis, 2017) impose additional assumptions on the level sets of the underlying function to obtain an improved convergence rate.

Methodology-wise, the success elimination algorithm in Sec. 3.1 is conceptually similar to the abstract *Pure Adaptive Search (PAS)* framework proposed and analyzed in (Zabinsky & Smith, 1992). The iterative procedure also resembles disagreement-based active learning methods (Balcan et al., 2009; Dasgupta et al., 2008; Hanneke, 2007) and the "successive rejection" algorithm in bandit problems (Even-Dar et al., 2006). The intermediate steps of candidate point elimination can also be viewed as sequences of level set estimation problems (Polonik, 1995; Rigollet & Vert, 2009; Singh et al., 2009) or cluster tree estimation (Balakrishnan et al., 2013; Chaudhuri et al., 2014) with active queries.

Bandit convex optimization In the literature of bandit convex optimization, the objective function sequence is subject to constant dynamic changes and only one or two evaluations for each function in the sequence is possible. The idea of using noisy function values subject to random perturbations to estimate function gradients first appeared in the seminal work of (Yudin & Nemirovskii, 1983) and was applied to bandit convex optimization problems in Flaxman et al. (2005). Agarwal et al. (2010) obtained improved convergence rates with additional strongly smooth/convex assumptions. Duchi et al. (2015); Shamir (2013, 2017) considered convex optimization/bandit problems with *noiseless* zeroth-order oracles and derived nearly matching convergence rates. Nesterov & Spokoiny (2017) studied the convergence rate under *noisy* zeroth-order oracles. Also, Hazan & Levy (2014) used an elliptical probing distribution to study the constrained zeroth-order optimization problem with optimal solution x^* very close to the boundary of feasible sets. Bubeck et al. (2017) proposed kernel based methods that attain optimal regret bounds for bandit convex optimization without strong convexity or smoothness assumptions. While most works consider *stationary* regret only, results on *dynamic* regret for bandit convex optimization also exist (Besbes et al., 2015).

Online convex optimization In online convex optimization, an *arbitrary* convex function sequence $f_1, \cdots, f_T$ is allowed, and the regret of a policy π is compared against the optimal *stationary* benchmark $\inf_{x \in \mathcal{X}} \{\sum_{t=1}^{T} f_t(x)\}$ in hindsight. Unlike the bandit convex optimization setting, in online convex optimization the full information of f_t is revealed to the optimizing algorithm after epoch t, which allows for exact gradient methods. It is known that for unconstrained online convex optimization, the simplest gradient descent method attains $O(\sqrt{T})$ regret for convex functions, and $O(\log T)$ regret for strongly convex and smooth functions, both of which are optimal

in the worst-case sense (Hazan, 2016). For constrained optimization problems, projection-free methods exist following mirror descent or follow-the-regularized-leader (FTRL) methods (Hazan & Levy, 2014). Hall & Willett (2015); Zinkevich (2003) considered the question of online convex optimization by competing against the optimal dynamic *solution sequence* $x_1^*, \cdots, x_T^*$ subject to certain smoothness constraints like $\sum_t \|x_{t+1}^* - x_t^*\| \leqslant C$. Jadbabaie et al. (2015); Mokhtari et al. (2016) further imposed the constraint on both solution sequences and function sequences in terms of $L_{\infty,1}$-variation and showed that adaptivity to the unknown smoothness parameter V_T is possible with noiseless gradient and the information of $\|f_t - f_{t-1}\|_\infty$. Daniely et al. (2015); Zhang et al. (2018) also designed algorithms that *adapt* to the unknown smoothness parameter, under the model that the entire function f_t is revealed after time t. However, the adaptation still remains an open problem in the "bandit" feedback setting considered in our paper, in which only noisy evaluations of $f_t(x_t)$ or $\nabla f_t(x_t)$ are revealed. Under the bandit feedback setting, the function perturbations (e.g., $\|f_{t+1} - f_t\|_\infty$) cannot be easily estimated, making it unclear whether adaptation to V_T is possible.

3.5 Proofs of results in Sec. 3.1

3.5.1 Proof of Lemma 12

We will need the following standard concentration inequality for Gaussian random vectors:

Lemma 24 ((Hsu et al., 2012)). *Suppose $x \sim \mathcal{N}_d(0, I_{d \times d})$ and let A be a $d \times d$ positive semi-definite matrix. Then for all $t > 0$,*

$$\Pr\left[x^\top A x > \operatorname{tr}(A) + 2\sqrt{\operatorname{tr}(A^2)t} + 2\|A\|_{\operatorname{op}} t\right] \leqslant e^{-t}.$$

Our proof closely follows the analysis of asymptotic convergence rates for series estimators in the seminal work of Newey (1997). We further work out all constants in the error bounds to arrive at a completely finite-sample result, which is then used to construct finite-sample confidence intervals.

We start with as polynomial interpolation results for all Hölder smooth functions in $B_{h_t}^\infty(x; \mathcal{X})$.

Lemma 25. *Suppose f satisfies Eq. (3.6) on $B_h^\infty(x; \mathcal{X})$. Then there exists $\widetilde{f}_x \in \mathcal{P}_k$ such that*

$$\sup_{z \in B_h^\infty(x;\mathcal{X})} \left|f(z) - \widetilde{f}_x(z)\right| \leqslant M d^k h^\alpha. \tag{3.42}$$

Proof. Consider

$$\widetilde{f}_x(z) := f(x) + \sum_{j=1}^k \sum_{\alpha_1+\ldots+\alpha_d=j} \frac{\partial^j f(x)}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}} \prod_{\ell=1}^d (z_\ell - x_\ell)^{\alpha_\ell}. \tag{3.43}$$

By Taylor expansion with Lagrangian remainders, there exists $\xi \in (0, 1)$ such that

$$\left|\widetilde{f}_x(z) - f(z)\right| \leqslant \sum_{\alpha_1+\ldots+\alpha_d=k} \left|f^{(\alpha)}(x + \xi(z - x)) - f^{(\alpha)}(x)\right| \cdot \prod_{\ell=1}^d |z_\ell - x_\ell|^{\alpha_\ell}. \tag{3.44}$$

Because f satisfies Eq. (3.6) on $B_h^\infty(x; \mathcal{X})$, we have that $|f^{(\alpha)}(x + \xi(z - x)) - f^{(\alpha)}(x)| \leqslant M \cdot \|z - x\|_\infty^{\alpha-k}$. Also note that $|z_\ell - x_\ell| \leqslant \|z - x\|_\infty \leqslant h$ for all $z \in B_h^\infty(x; \mathcal{X})$. The lemma is thus proved. $\square$

Using Eq. (3.9), the local polynomial estimate $\widehat{f}_h$ can be written as $\widehat{f}_h(z) \equiv \psi_{x,h}(z)^\top \widehat{\theta}_h$, where

$$\widehat{\theta}_h = (\Psi_{t,h}^\top \Psi_{t,h})^{-1} \Psi_{t,h}^\top Y_{t,h}. \tag{3.45}$$

In addition, because $\widetilde{f}_x \in \mathcal{P}_k$, there exists $\widetilde{\theta} \in \mathbb{R}^D$ such that $\widetilde{f}_x(z) \equiv \psi_{x,h}(z)^\top \widetilde{\theta}$. Denote also that $F_{t,h} := (f(x_{t'}))_{1 \leqslant t' \leqslant t, x_{t'} \in B_h^\infty(x)}$, $\Delta_{t,h} := (f(x_{t'}) - \widetilde{f}_x(x_{t'}))_{1 \leqslant t' \leqslant t, x_{t'} \in B_h^\infty(x)}$ and $W_{t,h} := (w_{t'})_{1 \leqslant t' \leqslant t, x_{t'} \in B_h^\infty(x)}$. Eq. (3.45) can then be re-formulated as

$$\widehat{\theta}_h = (\Psi_{t,h}^\top \Psi_{t,h})^{-1} \Psi_{t,h}^\top \left[\Psi_{t,h}\widetilde{\theta} + \Delta_{t,h} + W_{t,h} \right] \tag{3.46}$$

$$= \widetilde{\theta} + \left[\frac{1}{m} \Psi_{t,h}^\top \Psi_{t,h} \right]^{-1} \left[\frac{1}{m} \Psi_{t,h}^\top (\Delta_{t,h} + W_{t,h}) \right]. \tag{3.47}$$

Because $\frac{1}{m}\Psi_{t,h}^\top \Psi_{t,h} \succeq \sigma I_{D \times D}$ and $\sup_{z \in B_h^\infty(x)} \|\psi_{x,h}(z)\|_2 \leqslant b$, we have that

$$\|\widehat{\theta}_h - \widetilde{\theta}\|_2 \leqslant \frac{b}{\sigma}\|\Delta_{t,h}\|_\infty + \left\| \left[\frac{1}{m} \Psi_{t,h}^\top \Psi_{t,h} \right]^{-1} \frac{1}{m} \Psi_{t,h}^\top W_t \right\|_2. \tag{3.48}$$

Invoking Lemma 25 we have $\|\Delta_{t,h}\|_\infty \leqslant M d^k h^\alpha$. In addition, because $W_t \sim \mathcal{N}_m(0, I_{m \times n})$, we have that

$$\left[\frac{1}{m} \Psi_{t,h}^\top \Psi_{t,h} \right]^{-1} \frac{1}{m} \Psi_{t,h}^\top W_t \sim \mathcal{N}_D \left(0, \frac{1}{m} \left[\frac{1}{m} \Psi_{t,h}^\top \Psi_{t,h} \right]^{-1} \right). \tag{3.49}$$

Applying concentration inequalities for quadratic forms of Gaussian random vectors (Lemma 90), with probability $1 - \delta$ it holds that

$$\left\| \left[\frac{1}{m} \Psi_{t,h}^\top \Psi_{t,h} \right]^{-1} \frac{1}{m} \Psi_{t,h}^\top W_t \right\|_2 \leqslant \sqrt{\frac{5D \log(1/\delta)}{\sigma m}}. \tag{3.50}$$

We then have that with probability $1 - \delta$ that

$$\|\widehat{\theta}_h - \widetilde{\theta}\|_2 \leqslant \frac{b}{\sigma_h} M d^k h_t^\alpha + \sqrt{\frac{5D \log(1/\delta)}{\sigma m}}. \tag{3.51}$$

Finally, noting that

$$|\widehat{f}_h(x) - f(x)| = |\widehat{f}_h(x) - \widetilde{f}_x(x)| = |\psi(x)^\top (\widehat{\theta}_h - \widetilde{\theta})| \leqslant b\|\widehat{\theta}_h - \widetilde{\theta}\|_2 \tag{3.52}$$

we complete the proof of Lemma 12.

3.5.2 Proof of Proposition 6

To prove Proposition 6, we need the following lemma showing that the grid G_n is "dense" in approximating the target function f under assumptions (B1) and (B2). Define $x_n^* := \arg\min_{x \in G_n} f(x)$, $f_n^* = f(x_n^*)$ and $f^* = \inf_{x \in \mathcal{X}} f(x)$.

Lemma 26. *Suppose (B1) and (B2) hold. Then with probability $1 - O(n^{-1})$ the following holds:*

1. *$\sup_{x \in \mathcal{X}} \min_{x' \in G_n} \|x - x'\|_\infty = \widetilde{O}(n^{-3/\min(\alpha,1)})$;*
2. *$f_n^* - f^* = \widetilde{O}(n^{-3})$.*

Proof. Let $H_N \subseteq \mathcal{X}$ be the finite subset of $\mathcal{X}$ such that $|H_N| = N$ and $\sup_{x \in \mathcal{X}} \min_{x' \in H_n} \|x - x'\|_\infty$ is maximized. By standard results of metric entropy number of the d-dimensional unit box (see for example, (van de Geer, 2000, Lemma 2.2)), we have that $\sup_{x \in \mathcal{X}} \min_{x' \in H_n} \|x - x'\|_\infty \lesssim N^{-1/d}$.

For any $x \in H_n$, consider an ℓ_∞ ball $B_{r_n}^\infty(x)$ or radius r_n centered at x, with r_n to be specified later. Because the density of P_X is uniformly bounded away from below on $\mathcal{X}$, we have that $P_X(x \in B_{r_n}^\infty(x)) \gtrsim r_n^d$. Therefore, applying union bound over all $x \in H_n$ we have that

$$P_X\left[\exists x \in H_N, G_n \cap B_{r_n}^\infty(x) = \varnothing\right] \leqslant N(1 - r_n^d)^{|G_n|} \lesssim \exp\left\{-r_n^d|G_n| + \log N\right\}. \quad (3.53)$$

Set $N = |G_n|$ and $r_n \asymp n^{-3/\min(\alpha,1)} \log n$. The right-hand side of the above inequality is then upper bounded by $O(1/n^2)$, thanks to the assumption (A1) and that $|G_n| \gtrsim n^{3d/\min(\alpha,1)}$. The first property is then proved by noting that

$$\sup_{x \in \mathcal{X}} \min_{x' \in G_n} \|x - x'\|_\infty \leqslant \sup_{x \in \mathcal{X}} \min_{x' \in H_n} \|x - x'\|_\infty + \max_{x \in H_n} \min_{x' \in G_n} \|x - x'\|_\infty. \quad (3.54)$$

To prove the second property, note that for any $x, x' \in \mathcal{X}$, $|f(x) - f(x')| \leqslant M \cdot \|x - x'\|_\infty^{\min(\alpha,1)}$. The first property then implies that $f_n^* - f^* = \widetilde{O}(n^{-3})$. $\qquad\square$

We are now ready to prove Proposition 6. By Chernoff bound and union bound, with probability $1 - O(n^{-1})$ uniformly over all $x \in G_n$, there are $\Omega(\sqrt{n_0} \log^2 n)$ uniform samples in $B_{h_0}^\infty(x; \mathcal{X})$. Subsequently, by standard Gaussian concentration inequality, with probability $1 - O(n^{-1})$ we have

$$\inf_{z \in B_{h_0}^\infty(x;\mathcal{X})} f(z) - O(n_0^{-1/4}) \leqslant \check{f}(x) \leqslant \sup_{z \in B_{h_0}^\infty(x;\mathcal{X})} f(z) + O(n_0^{-1/4}) \qquad \forall x \in G_n. \quad (3.55)$$

Fix arbitrary $\widetilde{x}^* \in \arg\min_{x \in G_n} f(x)$. Because $f \in \Sigma_\kappa^\alpha(M)$ for constant κ and $h_0 \to 0$, f is smooth on $B_{h_0}^\infty(\widetilde{x}^*; \mathcal{X})$ and therefore $\sup_{z \in B_{h_0}^\infty(\widetilde{x}^*;\mathcal{X})} f(z) \leqslant f(\widetilde{x}^*) + O(h_0^{\min\{\alpha,1\}}) \leqslant f(\widetilde{x}^*) + O(1/\log^2 n) \leqslant f^* + O(1/\log^2 n)$, where the last inequality holds due to Lemma 26. On the other hand, for all $x \in G_n$, $\check{f}(x) \geqslant f^* - O(n_0^{-1/4})$. Therefore, for sufficiently large n we must have $\check{f}(\widetilde{x}^*) \leqslant \min_{z \in G_n} \check{f}(z) + 1/\log n$ and subsequently $\widetilde{x}^* \in S_0'$.

We next prove the statement that $S_0' \subseteq \bigcup_{x \in L_f(\kappa/2)} B_{h_0}^\infty(x; \mathcal{X})$. Consider arbitrary $z \in G_n$ and $z \notin \bigcup_{x \in L_f(\kappa/2)} B_{h_0}^\infty(x; \mathcal{X})$. By definition, $f(z') \geqslant f^* + \kappa/2$ for all $z' \in B_{h_0}^\infty(z; \mathcal{X})$. Subsequently, $\check{f}(z) \geqslant f^* + \kappa/2 - O(n_0^{-1/4}) > f^* + 1/\log n$ for constant $\kappa > 0$ and sufficiently large n, which implies $z \notin S_0'$.

3.5.3 Proof of Theorem 4

We prove the theorem by considering every reference function $f_0 \in \Sigma_\kappa^\alpha(M) \cap \Theta_C$ separately. For simplicity, we assume $\kappa = \infty$ throughout the proof. The $0 < \kappa < \infty$ can be handled by replacing $\mathcal{X}$ with S_0 which is the grid after the pre-screening step described in Section 3.1.3. We also suppress dependency on $d, \alpha, M, C, \underline{p}_0, \overline{p}_0$ in $O(\cdot)$, $\Omega(\cdot)$, $\Theta(\cdot)$, $\gtrsim$, $\lesssim$ and $\asymp$ notations. We further suppress logarithmic terms of n in $\widetilde{O}(\cdot)$ and $\widetilde{\Omega}(\cdot)$ notations.

The following lemma is our main lemma, which shows that the active set S_τ in our proposed algorithm shrinks geometrically before it reaches a certain level. To simplify notations, denote $\widetilde{c}_0 := 10 c_0$ and (A2) then hold for all $\epsilon, \delta \in [0, \widetilde{c}_0]$ for all $f_0 \in \Theta_C$.

Lemma 27. *For $\tau = 1, \ldots, T$ define $\varepsilon_\tau := \max\{\widetilde{c}_0 \cdot 2^{-\tau}, C_3[\varepsilon_n^{\mathsf{U}}(f_0) + n^{-1/2}] \log^2 n\}$, where $C_3 > 0$ is a constant depending only on $d, \alpha, M, \underline{p}_0, \overline{p}_0$ and C. Then for sufficiently large n, with probability $1 - O(n^{-1})$ the following holds uniformly for all outer iterations $\tau = 1, \ldots, T$:*

$$S_\tau \subseteq L_f(\varepsilon_\tau). \tag{3.56}$$

Lemma 27 shows that the level ε_τ in $L_f(\varepsilon_\tau)$ that contains $S_{\tau-1}$ shrinks *geometriclly*, until the condition $\varepsilon_\tau \geqslant C_3[\varepsilon_n^{\mathsf{U}}(f_0) + n^{-1/2}] \log^2 n$ is violated. If the condition is never violated, then at the end of the last epoch τ^* we have $\varepsilon_{\tau*} = O(n^{-1})$ because $\tau^* = \log n$, in which case Theorem 4 clearly holds. On the other hand, because $S_\tau \subseteq S_{\tau-1}$ always holds, we have $\varepsilon_{\tau*} \lesssim [\varepsilon_n^{\mathsf{U}}(f_0) + n^{-1/2}] \log^2 n$ which justifies the convergence rate in Theorem 4.

In the rest of this section we prove Lemma 27. We need several technical lemmas and propositions. Except for Proposition 11 that is straightforward, the proofs of the other technical lemmas are deferred to the end of this section.

The next proposition shows that with high probability, the confidence intervals constructed in the algorithm are truthful and the successive rejection procedure will never exclude the true optimizer of f on G_n.

Proposition 11. *Suppose $\delta = 1/n^4 |G_n|$. Then with probability $1 - O(n^{-1})$ the following holds:*
  1. *$f(x) \in [\ell_t(x), u_t(x)]$ for all $1 \leqslant t \leqslant n$ and $x \in G_n$;*
  2. *$x_n^* \in S_\tau$ for all $0 \leqslant \tau \leqslant n$.*

Proof. The first property is true by applying the union bound over all $t = 1, \ldots, n$ and $x \in G_n$. The second property then follows, because $\ell_t(x_n^*) \leqslant f_n^*$ and $\min_{x \in S_{\tau-1}} u_t(x) \geqslant f_n^*$ for all τ. $\square$

The following lemma shows that every small box centered around a certain sample point $x \in G_n$ contains a sufficient number of sample points whose least eigenvalue can be bounded with high probability under the polynomial mapping $\psi_{x,h}$.

Lemma 28. *For any $x \in G_n$, $1 \leqslant m \leqslant n$ and $h > 0$, let $K_{h,m}^1(x), \ldots, K_{h,m}^n(x)$ be n independent point sets, where each point set consists of m points sampled i.i.d. uniformly at random from $B_h^\infty(x; G_n) = G_n \cap B_h^\infty(x)$. With probability $1 - O(n^{-1})$ the following holds true uniformly for all $x \in G_n$, $h \in \{j/n^2 : j \in \mathbb{N}, j \leqslant n^2\}$ and $K_{h,m}^\ell(x)$, $\ell \in [n]$ as $n \to \infty$:*
  1. *$\sup_{h>0} \sup_{z \in B_h^\infty(x)} \|\psi_{x,h}(z)\|_2 \asymp \Theta(1)$;*
  2. *$|B_h^\infty(x; G_n)| \asymp h^d |G_n|$;*

3. $\sigma_{\min}(K_{h,m}^{\ell}(x)) \asymp \Theta(1)$ *for all* $m \geqslant \Omega(\log^2 n)$ *and* $m \leqslant |G_n|$, *where* $\sigma_{\min}(K_{h,m}^{\ell}(x))$ *is the least eigenvalue of* $\frac{1}{m}\sum_{z \in K_{h,m}^{\ell}(x)} \psi_{x,h}(z)\psi_{x,h}(z)^{\top}$.

Remark 24. It is possible to improve the concentration result in Eq. (3.73) using the strategies adopted in (Chaudhuri et al., 2014) based on sharper Bernstein type concentration inequalities. Such improvements are, however, not important in establishing the main results of this paper.

The next lemma shows that, the bandwidth h_t selected at the end of each outer iteration τ is near-optimal, being sandwiched between two quantities determined by the size of the active sample grid $\widetilde{S}_{\tau-1} := S_{\tau-1}^{\circ}(\varrho_{\tau-1})$.

Lemma 29. *There exist constants $C_1, C_2 > 0$ depending only on $d, \alpha, M, \underline{p}_0, \overline{p}_0$ and C such that with probability $1 - O(n^{-1})$, the following holds for every outer iteration $\tau \in \{1, \dots, T\}$ and all $x \in S_{\tau-1}$:*

$$C_1[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)} - \tau/n \leqslant \varrho_{\tau}(x) \leqslant h_t(x) \leqslant C_2[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)}\log n + \tau/n, \qquad (3.57)$$

where $\widetilde{\nu}_{\tau-1} := |G_n|/|\widetilde{S}_{\tau-1}|$.

We are now ready to state the proof of Lemma 27, which is based on an inductive argument over the epochts $\tau = 1, \dots, T$.

Proof. We use induction to prove this lemma. For the base case $\tau = 1$, because $\|f - f_0\|_{\infty} \leqslant \varepsilon_n^{\mathsf{U}}(f_0)$ and $\varepsilon_n^{\mathsf{U}}(f_0) \to 0$ as $n \to \infty$, it suffices to prove that $S_1 \subseteq L_{f_0}(\widetilde{c}_0/4)$ for sufficiently large n. Because $\widetilde{S}_0 = S_0 = G_n$, invoking Lemmas 29 and 12 we have that $|u_t(x) - \ell_t(x)| = \widetilde{O}(n^{-\alpha/(2\alpha+d)})$ for all $x \in G_n$ with high probability at the end of the first outer iteration $\tau = 1$. Therefore, for sufficiently large n we conclude that $\sup_{x \in G_n}|u_t(x) - \ell_t(x)| \leqslant c_0/8$ and hence $S_1 \subseteq L_{f_0}(\widetilde{c}_0/4)$.

We now prove the lemma for $\tau \geqslant 2$, assuming it holds for $\tau - 1$. We also assume that n (and hence n_0) is sufficiently large, such that the maximum CI length $\max_{x \in G}|u_t(x) - \ell_t(x)|$ after the first outer iteration $\tau = 1$ is smaller than c_0, where c_0 is a constant such that

Because $\|f - f_0\|_{\infty} \leqslant \varepsilon_n^{\mathsf{U}}(f_0)$ and $\varepsilon_{\tau-1} \geqslant C_3\varepsilon_n^{\mathsf{U}}(f_0)\log^2 n$, for appropriately chosen constant C_3 that is not too small, we have that $\|f - f_0\|_{\infty} \leqslant \varepsilon_{\tau-1}$. By the inductive hypothesis we have

$$S_{\tau-1} \subseteq L_f(\varepsilon_{\tau-1}) \subseteq L_{f_0}(\varepsilon_{\tau-1} + \|f - f_0\|_{\infty}) \subseteq L_{f_0}(2\varepsilon_{\tau-1}). \qquad (3.58)$$

Subsequently, denoting $\rho_{\tau-1}^* := \max_{x \in S_{\tau-1}}\varrho_{\tau-1}(x)$ we have

$$\widetilde{S}_{\tau-1} = S_{\tau-1}^{\circ} \subseteq L_{f_0}^{\circ}(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*). \qquad (3.59)$$

Let $\bigcup_{x \in H_n} B_{\rho_{\tau-1}^*}^2(x)$ be the smallest covering set of $L_{f_0}(2\varepsilon_{\tau-1})$, meaning that $L_{f_0}(2\varepsilon_{\tau-1}) \subseteq \bigcup_{x \in H_n} B_{\rho_{\tau-1}^*}^2(x)$, where $B_{\rho_{\tau-1}^*}^2(x) = \{z \in \mathcal{X} : \|z - x\|_2 \leqslant \rho_{\tau-1}^*\}$ is the ℓ_2 ball of radius $\rho_{\tau-1}^*$ centered at x. By (A2), we know that $|H_n| \lesssim 1 + [\rho_{\tau-1}^*]^{-d}\mu_{f_0}(2\varepsilon_{\tau-1})$. In addition, the enlarged level set satisfies $L_{f_0}^{\circ}(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*) \subseteq \bigcup_{x \in H_n} B_{2\rho_{\tau-1}^*}^{\infty}(x)$. Subsequently,

$$\mu_{f_0}^{\circ}(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*) \lesssim |\mathcal{H}_n| \cdot [\rho_{\tau-1}^*]^d \lesssim \mu_{f_0}(2\varepsilon_{\tau-1}) + [\rho_{\tau-1}^*]^d. \qquad (3.60)$$

70

By Lemma 29, the monotonicity of $|\widetilde{S}_{\tau-1}|$ and the fact that $\underline{p}_0 \leqslant p_X(z) \leqslant \overline{p}_0$ for all $z \in \mathcal{X}$, we have

$$\rho_{\tau-1}^* \lesssim \left[\mu_f^\circ(\varepsilon_{\tau-1}, \rho_{\tau-1}^*) \right]^{1/(2\alpha+d)} n_0^{-1/(2\alpha+d)} \log n \tag{3.61}$$

$$\leqslant \left[\mu_{f_0}^\circ(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*) \right]^{1/(2\alpha+d)} n_0^{-1/(2\alpha+d)} \log n \tag{3.62}$$

$$\lesssim \left(\mu_{f_0}(2\varepsilon_{\tau-1}) + [\rho_{\tau-1}^*]^d \right)^{1/(2\alpha+d)} n_0^{-1/(2\alpha+d)} \log n. \tag{3.63}$$

Re-arranging terms on both sides of Eq. (3.63) we have

$$\rho_{\tau-1}^* \lesssim \max\left\{ \left[\mu_{f_0}(2\varepsilon_{\tau-1}) \right]^{\frac{1}{2\alpha+d}} n_0^{-\frac{1}{2\alpha+d}} \log n, \ n_0^{-\frac{1}{2\alpha}} \log n \right\}. \tag{3.64}$$

On the other hand, according to the selection procedure of the bandwidth $h_t(x)$, we have that $\eta_{h_t(x),\delta}(x) \lesssim \mathfrak{b}_{h_t(x),\delta}(x)$. Invoking Lemma 29 we have for all $x \in S_{\tau-1}$ that

$$\eta_{h_t(x),\delta}(x) \lesssim \mathfrak{b}_{h_t(x),\delta}(x) \lesssim [h_t(x)]^\alpha \tag{3.65}$$

$$\lesssim [\widetilde{\nu}_{\tau-1} n_0]^{-\alpha/(2\alpha+d)} \log n \tag{3.66}$$

$$\lesssim [\widetilde{\nu}_{\tau-2} n_0]^{-\alpha/(2\alpha+d)} \log n \tag{3.67}$$

$$\lesssim [\rho_{\tau-1}^*]^\alpha \log n. \tag{3.68}$$

Here Eq. (3.66) holds by invoking the upper bound on $h_t(x)$ in Lemma 29, Eq. (3.67) holds because $\widetilde{\nu}_{\tau-1} \geqslant \widetilde{\nu}_{\tau-2}$, and Eq. (3.68) holds by again invoking the lower bound on $\varrho_{\tau-1}(x)$ in Lemma 29. Combining Eqs. (3.64,3.68) we have

$$\max_{x \in S_{\tau-1}} \eta_{h_t(x),\delta}(x) \lesssim \max\left\{ \left[\mu_{f_0}(2\varepsilon_{\tau-1}) \right]^{\frac{\alpha}{2\alpha+d}} n_0^{-\frac{\alpha}{2\alpha+d}} \log^2 n, \ n_0^{-\frac{1}{2}} \log n \right\}. \tag{3.69}$$

Recall that $n_0 = n/\log n$ and $\varepsilon_n^{\mathsf{U}}(f_0) \leqslant \varepsilon_{\tau-1}$, provided that C_3 is not too small. By definition, every $\varepsilon \geqslant \varepsilon_n^{\mathsf{U}}(f_0)$ satisfies $\varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \leqslant n/\log^\omega n$ for some large constant $\omega > 5 + d/\alpha$. Subsequently,

$$\left[\mu_{f_0}(2\varepsilon_{\tau-1}) \right]^{\frac{\alpha}{2\alpha+d}} n_0^{-\frac{\alpha}{2\alpha+d}} \log^2 \lesssim 2\varepsilon_{\tau-1} n^{\frac{\alpha}{2\alpha+d}} \log^{-\frac{\omega\alpha}{2\alpha+d}} n \cdot n_0^{-\frac{\alpha}{2\alpha+d}} \log^2 n \tag{3.70}$$

$$\lesssim \varepsilon_{\tau-1} / [\log n]^{\frac{(\omega-5-d/\alpha)\alpha}{2\alpha+d}}. \tag{3.71}$$

Because $\omega > 5 + d/\alpha$, the right-hand side of Eq. (3.71) is asymptotically dominated [5] by $\varepsilon_{\tau-1}$. In addition, $n_0^{-1/2} \log n$ is also asymptotically dominated by $\varepsilon_{\tau-1}$ because $\varepsilon_{\tau-1} \geqslant C_3 n^{-1/2} \log^\omega n$. Therefore, for sufficiently large n we have

$$\max_{x \in S_{\tau-1}} \eta_{h_t(x),\delta}(x) \leqslant \varepsilon_{\tau-1}/4. \tag{3.72}$$

Lemma 27 is thus proved. $\qquad\square$

---

[5] We say $\{a_n\}$ is asymptotically dominated by $\{b_n\}$ if $\lim_{n\to\infty} |a_n|/|b_n| = 0$.

Proof of Lemma 28 We first show that the first property holds almost surely. Recall the definition of $\psi_{x,h}$, we have that $1 \leqslant \|\psi_{x,h}(z)\|_2 \leqslant D \cdot [\max_{1 \leqslant j \leqslant d} h^{-1}|z_j - x_j|]^k$. Because $\|z - x\|_\infty \leqslant h$ for all $z \in B_h^\infty(x)$, $\sup_{z \in B_h^\infty(x)} \|\psi_{x,h}(z)\|_2 \lesssim O(1)$ for all $h > 0$. Thus, $\sup_{h>0} \sup_{z \in B_h^\infty(x)} \|\psi_{x,h}(z)\|_2 \asymp \Theta(1)$ for all $x \in G_n$.

For the second property, by Hoeffding's inequality (Lemma 89) and the union bound, with probability $1 - O(n^{-1})$ we have that

$$\max_{x,h} \left| \frac{|B_h^\infty(x; G_n)|}{|G_n|} - P_X(z \in B_h^\infty(x)) \right| \lesssim \sqrt{\frac{\log n}{|G_n|}}. \tag{3.73}$$

In addition, note that $P_X(z \in B_h^\infty(x; \mathcal{X})) \geqslant \underline{p}_0 \lambda(B_h^\infty(x; \mathcal{X})) \gtrsim h^d$ and $P_X(z \in B_h^\infty(x; \mathcal{X})) \leqslant \overline{p}_0 \lambda(B_h^\infty(x; \mathcal{X})) \lesssim h^d$, where $\lambda(\cdot)$ denotes the Lebesgue measure on $\mathcal{X}$. Subsequently, $|B_h^\infty(x; G_n)|$ is lower bounded by $\Omega(h^d|G_n| - \sqrt{|G_n| \log n})$ and upper bounded by $O(h^d|G_n| + \sqrt{|G_n| \log n})$. The second property is then proved by noting that $h_d \gtrsim n^{-d}$ and $|G_n| \gtrsim n^{3d/\min(\alpha,1)}$.

We next prove the third property. Because $\underline{p}_0 \leqslant p_X(z) \in \overline{p}_0$ for all $z \in \mathcal{X}$, we have that

$$\underline{p}_0 \int_{B_h^\infty(x; \mathcal{X})} \psi_{x,h}(z)\psi_{x,h}(z)^\top dU_{x,h}(z) \leq \mathbb{E}\left[\frac{1}{m} \sum_{z \in K_{h,m}^\ell} \psi_{x,h}(z)\psi_{x,h}(z)^\top \right] \tag{3.74}$$

$$\leq \overline{p}_0 \int_{B_h^\infty(x; \mathcal{X})} \psi_{x,h}(z)\psi_{x,h}(z)^\top dU_{x,h}(z), \tag{3.75}$$

where $U_{x,h}$ is the uniform distribution on $B_h^\infty(x; \mathcal{X})$. Note also that

$$\int_{\mathcal{X}} \psi_{0,1}(z)\psi_{0,1}(z)^\top dU(z) \leq \int_{B_h^\infty(x; \mathcal{X})} \psi_{x,h}(z)\psi_{x,h}(z)^\top dU_{x,h}(z) \tag{3.76}$$

$$\leq 2^d \int_{\mathcal{X}} \psi_{0,1}(z)\psi_{0,1}(z)^\top dU(z) \tag{3.77}$$

where U is the uniform distribution on $\mathcal{X} = [0,1]^d$. The following proposition upper and lower bounds the eigenvalues of $\int_{\mathcal{X}} \psi_{0,1}(z)\psi_{0,1}(z)^\top dU(z)$, which is proved in the appendix.

Proposition 12. *There exist constants $0 < \psi_0 \leqslant \Psi_0 < \infty$ depending only on d, D such that*

$$\psi_0 I_{D \times D} \leq \int_{\mathcal{X}} \psi_{0,1}(z)\psi_{0,1}(z)^\top dU(z) \leq \Psi_0 I_{D \times D}. \tag{3.78}$$

Using Proposition 12 and Eqs. (3.76,3.77), we conclude that

$$\Omega(1) \cdot I_{D \times D} \leq \mathbb{E}\left[\frac{1}{m} \sum_{z \in K_{h,m}^\ell} \psi_{x,h}(z)\psi_{x,h}(z)^\top \right] \leq O(1) \cdot I_{D \times D}. \tag{3.79}$$

Applying matrix Chernoff bound (Lemma 98) and the union bound, we have that with probability $1 - O(n^{-1})$,

$$\max_{x,h,m,\ell} \left\| \frac{1}{m} \sum_{z \in K_{h,m}^\ell(x)} \psi_{x,h}(z)\psi_{x,h}(z)^\top - \mathbb{E}\left[\psi_{x,h}(z)\psi_{x,h}(z)^\top | z \in B_h(x) \right] \right\|_{\text{op}} \lesssim \sqrt{\frac{\log n}{m}}. \tag{3.80}$$

72

Combining Eqs. (3.79,3.80) and applying Weyl's inequality (Lemma 99) we have

$$\Omega(1) - O(\sqrt{\log n/m}) \lesssim \sigma_{\min}(K_{h,m}^{\ell}(x)) \lesssim O(1) - O(\sqrt{\log n/m}). \tag{3.81}$$

The third property is therefore proved.

Proof of Lemma 29   We use induction to prove this lemma. For the base case of $\tau = 1$, we have $\widetilde{S}_0 = S_0 = G_n$ and therefore $\widetilde{\nu}_{\tau-1} = 1$. Furthermore, applying Lemma 28 we have that for all $h = j/n^2$,

$$\mathfrak{b}_{h,\delta}(x) \asymp h^{\alpha} \quad \text{and} \quad \mathfrak{s}_{h,\delta}(x) \asymp \sqrt{\frac{\log n}{h^d n_0}}. \tag{3.82}$$

Thus, for h selected according to Eq. (3.11) as the largest bandwidth of the form j/n^2, $j \in \mathbb{N}$ such that $\mathfrak{b}_{h,\delta}(x) \leqslant \mathfrak{s}_{h,\delta}(x)$, both $\mathfrak{b}_{h,\delta}(x), \mathfrak{s}_{h,\delta}(x)$ are on the order of $n_0^{-1/(2\alpha+d)}$ up to logarithmic terms of n, and therefore one can pick appropriate constants $C_1, C_2 > 0$ such that $C_1 n_0^{-1/(2\alpha+d)} \leqslant \varrho_1(x) \leqslant C_2 n_0^{-1/(2\alpha+d)} \log n$ holds for all $x \in G_n$.

We next prove the lemma for $\tau > 1$, assuming it holds for $\tau - 1$. We first establish the lower bound part. Define $\rho_{\tau-1}^* := \min_{z \in S_{\tau-1}} \varrho_{\tau-1}(z)$. By inductive hypothesis, $\rho_{\tau-1}^* \geqslant C_1[\widetilde{\nu}_{\tau-2}n_0]^{-1/(2\alpha+d)} - (\tau-1)/n$. Note also that $\widetilde{\nu}_{\tau-1} \geqslant \widetilde{\nu}_{\tau-2}$ because $\widetilde{S}_{\tau-1} \subseteq \widetilde{S}_{\tau-2}$, which holds because $S_{\tau-1} \subseteq S_{\tau-2}$ and $\varrho_{\tau-1}(z) \leqslant \varrho_{\tau-2}(z)$ for all z. Let h_t^* be the smallest number of the form j_t^*/n^2, $j_t^* \in [n^2]$ such that $h_t^* \geqslant C_1[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)} - \tau/n$. We then have $h_t^* \leqslant \rho_{\tau-1}^*$ and therefore query points in epoch τ are uniformly distributed in $B_{h_t^*}^{\infty}(x; G_n)$. Subsequently, applying Lemma 28 we have with probability $1 - O(n^{-1})$ that

$$\mathfrak{b}_{h_t^*,\delta}(x) \leqslant C'[h_t^*]^{\alpha} \quad \text{and} \quad \mathfrak{s}_{h_t^*,\delta}(x) \geqslant C''\sqrt{\frac{\log n}{[h_t^*]^d \widetilde{\nu}_{\tau-1} n}}, \tag{3.83}$$

where $C', C'' > 0$ are constants that depend on $d, \alpha, M, \underline{p}_0, \overline{p}_0$ and C, but *not* C_1, C_2, τ or h_t^*. By choosing C_1 appropriately (depending on C' and C'') we can make $\mathfrak{b}_{h_t^*,\delta}(x) \leqslant \mathfrak{s}_{h_t^*,\delta}(x)$ holds for all $x \in S_{\tau-1}$, thus establishing $\varrho_\tau(x) \geqslant \min\{\varrho_{\tau-1}(x), h_t^*\} \geqslant C_1[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)} - \tau/n$.

We next prove the upper bound part. For any $h_t = j_t/n^2$ where $j_t \in [n^2]$, invoking Lemma 28 we have that

$$\mathfrak{b}_{h,\delta}(x) \geqslant \widetilde{C}'h^{\alpha} \quad \text{and} \quad \mathfrak{s}_{h,\delta}(x) \leqslant \widetilde{C}''\sqrt{\frac{\log n}{\min\{h, \rho_{\tau-1}^*\}^d \cdot \widetilde{\nu}_{\tau-1} n_0}}, \tag{3.84}$$

where $\widetilde{C}'$ and $\widetilde{C}''$ are again constants depending on $d, \alpha, M, \underline{p}_0, \overline{p}_0$ and C, but *not* C_1, C_2. Note also that $\rho_{\tau-1}^* \geqslant C_1[\widetilde{\nu}_{\tau-2}n_0]^{-1/(2\alpha+d)} - (\tau-1)/n \geqslant C_1[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)} - \tau/n$, because $\widetilde{\nu}_{\tau-1} \geqslant \widetilde{\nu}_{\tau-2}$. By selecting constant $C_2 > 0$ carefully (depending on $\widetilde{C}', \widetilde{C}''$ and C_1), we can ensure $\mathfrak{b}_{h,\delta}(x) > \mathfrak{s}_{h,\delta}(x)$ for all $h \geqslant C_2[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)} + \tau/n$. Therefore, $\varrho_\tau(x) \leqslant h_t(x) \leqslant C_2[\widetilde{\nu}_{\tau-1}n_0]^{-1/(2\alpha+d)} + \tau/n$.

73

3.5.4 Proof of Theorem 5

In this section we prove the main negative result in Theorem 5. To simplify presentation, we suppress dependency on α, d, c_0 and C_0 in $\lesssim, \gtrsim, \asymp, O(\cdot)$ and $\Omega(\cdot)$ notations. However, we do *not* suppress dependency on $\underline{C}_R$ or M in any of the above notations.

Let $\varphi_0 : [-2, 2]^d \to \mathbb{R}^*$ be a non-negative function defined on $\mathcal{X}$ such that $\varphi_0 \in \Sigma_\kappa^{\lceil\alpha\rceil}(1)$ with $\kappa = \infty$, $\sup_{x\in\mathcal{X}} \varphi_0(x) = \Omega(1)$ and $\varphi_0(z) = 0$ for all $\|z\|_2 \geqslant 1$. Here $\lceil\alpha\rceil$ denotes the smallest integer that upper bounds α. Such functions exist and are the cornerstones of the construction of information-theoretic lower bounds in nonparametric estimation problems (Castro & Nowak, 2008). One typical example is the "smoothstep" function (see for example (Ebert, 2003))

$$S_N(x) := \frac{1}{Z} x^{N+1} \sum_{n=0}^{N} \binom{N+n}{n} \binom{2N+1}{N-n} (-x)^n, \qquad N = 0, 1, 2, \dots$$

where $Z > 0$ is a scaling parameter. The smoothstep function S_N is defined on $[0, 1]$ and satisfies the Hölder condition in Eq. (3.6) of order $\alpha = N$ on $[0, 1]$. It can be easily extended to $\widetilde{S}_{N,d} : [-2, 2]^d \to \mathbb{R}$ by considering $\widetilde{S}_{N,d}(x) := 1/Z - S_N(a\|x\|_1)$ where $\|x\|_1 = |x_1| + \dots + |x_d|$ and $a = 1/(2d)$. It is easy to verify that, with Z chosen appropriately, $\widetilde{S}_{N,d} \in \Sigma_\infty^N(1)$, $\sup_{x\in\mathcal{X}} \widetilde{S}_{N,d}(x) = 1/Z = \Omega(1)$ and $\widetilde{S}_{N,d}(z) = 0$ for all $\|z\|_2 \geqslant 1$, where $M > 0$ is a constant.

For any $x \in \mathcal{X}$ and $h > 0$, define $\varphi_{x,h} : \mathcal{X} \to \mathbb{R}^*$ as

$$\varphi_{x,h}(z) := \mathbb{I}[z \in B_h^\infty(x)] \cdot \frac{Mh^\alpha}{2} \varphi_0 \left(\frac{z - x}{h} \right). \tag{3.85}$$

It is easy to verify that $\varphi_{x,h} \in \Sigma_\infty^\alpha(M/2)$, and furthermore $\sup_{z\in\mathcal{X}} \varphi_{x,h}(z) \asymp Mh^\alpha$ and $\varphi_{x,h}(z) = 0$ for all $z \notin B_h^\infty(x)$.

Let $L_{f_0}(\varepsilon_n^{\mathsf{L}}(f_0))$ be the level set of f_0 at $\varepsilon_n^{\mathsf{L}}(f_0)$. Let $H_n \subseteq L_{f_0}(\varepsilon_n^{\mathsf{L}}(f_0))$ be the largest *packing* set such that $B_h^\infty(x)$ are disjoint for all $x \in H_n$, and $\bigcup_{x\in H_n} B_h^\infty(x) \subseteq L_{f_0}(\varepsilon_n^{\mathsf{L}}(f_0))$. By (A2') and the definition of $\varepsilon_n^{\mathsf{L}}(f_0)$, we have that

$$|H_n| \geqslant M(L_{f_0}(\varepsilon_n^{\mathsf{L}}(f_0)), 2\sqrt{d}h) \gtrsim \mu_{f_0}(\varepsilon_n^{\mathsf{L}}(f_0)) \cdot h^{-d} \geqslant [\varepsilon_n^{\mathsf{L}}(f_0)]^{2+d/\alpha} \cdot nh^{-d}. \tag{3.86}$$

For any $x \in H_n$, construct $f_x : \mathcal{X} \to \mathbb{R}$ as

$$f_x(z) := f_0(z) - \varphi_{x,h}(z). \tag{3.87}$$

Let $\mathcal{F}_n := \{f_x : x \in H_n\}$ be the class of functions indexed by $x \in H_n$. Let also $h \asymp (\varepsilon_n^{\mathsf{L}}(f_0)/M)^{1/\alpha}$ such that $\|\varphi_{x,h}\|_\infty = 2\varepsilon_n^{\mathsf{L}}(f_0)$. We then have that $\|f_x - f_0\|_\infty \leqslant 2\varepsilon_n^{\mathsf{L}}(f_0)$ and $f_x \in \Sigma_\infty^\alpha(M)$, because $f_0, \varphi_{x,h} \in \Sigma_\infty^\alpha(M/2)$.

The next lemma shows that, with n adaptive queries to the noisy zeroth-order oracle $y_t = f(x_t) + w_t$, it is information theoretically not possible to identify a certain f_x in $\mathcal{F}_n$ with high probability.

Lemma 30. *Suppose $|\mathcal{F}_n| \geqslant 2$. Let $\mathcal{A}_n = (\chi_1, \dots, \chi_n, \phi_n)$ be an active optimization algorithm operating with a sample budget n, which consists of* samplers $\chi_\ell : \{(x_i, y_i)\}_{i=1}^{\ell-1} \mapsto x_\ell$ *and an estimator* $\phi_n : \{(x_i, y_i)\}_{i=1}^{n} \mapsto \widehat{f}_x \in \mathcal{F}_n$, *both can be deterministic or randomized functions. Then*

$$\inf_{\mathcal{A}_n} \sup_{f_x\in\mathcal{F}_n} \Pr_{f_x} \left[\widehat{f}_x \neq f_x \right] \geqslant \frac{1}{2} - \sqrt{\frac{n \cdot \sup_{f_x\in\mathcal{F}_n} \|f_x - f_0\|_\infty^2}{2|\mathcal{F}_n|}}. \tag{3.88}$$

Lemma 31. *There exists constant $M > 0$ depending on α, d, c_0, C_0 such that the right-hand side of Eq. (3.88) is lower bounded by $1/3$.*

Lemmas 30 and 31 are proved at the end of this section. Combining both lemmas and noting that for any distinct $f_x, f_{x'} \in \mathcal{F}_n$ and $z \in \mathcal{X}$, $\max\{\mathfrak{L}(z; f_x), \mathfrak{L}(z; f_{x'})\} \geqslant \varepsilon_n^{\mathsf{L}}(f_0)$, we proved the minimax lower bound formulated in Theorem 5.

Proof of Lemma 30 Our proof is inspired by the negative result of multi-arm bandit pure exploration problems established in (Bubeck et al., 2009). For any $x \in H_n$, define

$$n_x := \mathbb{E}_{f_0}\left[\sum_{i=1}^{n} \mathbb{I}[x \in B_h^{\infty}(x)]\right]. \tag{3.89}$$

Because $B_h^{\infty}(x)$ are disjoint for $x \in H_n$, we have $\sum_{x \in H_n} n_x \leqslant n$. Also define, for every $x \in H_n$,

$$\wp_x := \Pr_{f_0}\left[\widehat{f}_x = f_x\right]. \tag{3.90}$$

Because $\sum_{x \in H_n} \wp_x = 1$, by pigeonhole principle there is at most one $x \in H_n$ such that $\wp_x > 1/2$. Let $x_1, x_2 \in H_n$ be the points that have the largest and second largest n_x. Then there exists $x \in \{x_1, x_2\}$ such that $\wp_x \leqslant 1/2$ and $n_x \leqslant 2n/|\mathcal{F}_n|$. By Le Cam's and Pinsker's inequality (see, for example, (Tsybakov, 2009)) we have that

$$\Pr_{f_x}\left[\widehat{f}_x = f_x\right] \leqslant \Pr_{f_0}\left[\widehat{f}_x = f_x\right] + d_{\mathrm{TV}}(P_{f_0}^{\mathcal{A}_n} \| P_{f_x}^{\mathcal{A}_n}) \tag{3.91}$$

$$\leqslant \Pr_{f_0}\left[\widehat{f}_x = f_x\right] + \sqrt{\frac{1}{2}\mathrm{KL}(P_{f_0}^{\mathcal{A}_n} \| P_{f_x}^{\mathcal{A}_n})} \tag{3.92}$$

$$= \wp_x + \sqrt{\frac{1}{2}\mathrm{KL}(P_{f_0}^{\mathcal{A}_n} \| P_{f_x}^{\mathcal{A}_n})} \tag{3.93}$$

$$\leqslant \frac{1}{2} + \sqrt{\frac{1}{2}\mathrm{KL}(P_{f_0}^{\mathcal{A}_n} \| P_{f_x}^{\mathcal{A}_n})}. \tag{3.94}$$

It remains to upper bound KL divergence of the active queries made by $\mathcal{A}_n$. Using the standard lower bound analysis for active learning algorithms (Castro, 2014; Castro & Nowak, 2008) and the fact that $f_x \equiv f_0$ on $\mathcal{X} \backslash B_h^{\infty}(x)$, we have

$$\mathrm{KL}(P_{f_0}^{\mathcal{A}_n} \| P_{f_x}^{\mathcal{A}_n}) = \mathbb{E}_{f_0, \mathcal{A}_n}\left[\log \frac{P_{f_0, \mathcal{A}_n}(x_{1:n}, y_{1:n})}{P_{f_x, \mathcal{A}_n}(x_{1:n}, y_{1:n})}\right] \tag{3.95}$$

$$= \mathbb{E}_{f_0, \mathcal{A}_n}\left[\log \frac{\prod_{i=1}^{n} P_{f_0}(y_i|x_i) P_{\mathcal{A}_n}(x_i|x_{1:(i-1)}, y_{1:(i-1)})}{\prod_{i=1}^{n} P_{f_x}(y_i|x_i) P_{\mathcal{A}_n}(x_i|x_{1:(i-1)}, y_{1:(i-1)})}\right] \tag{3.96}$$

$$= \mathbb{E}_{f_0, \mathcal{A}_n}\left[\log \frac{\prod_{i=1}^{n} P_{f_0}(y_i|x_i)}{\prod_{i=1}^{n} P_{f_x}(y_i|x_i)}\right] \tag{3.97}$$

$$= \mathbb{E}_{f_0, \mathcal{A}_n}\left[\sum_{x_i \in B_h(x)} \log \frac{P_{f_0}(y_i|x_i)}{P_{f_x}(y_i|x_i)}\right] \tag{3.98}$$

75

$$\leqslant n_x \cdot \sup_{z \in B_h^\infty(x; \mathcal{X})} \mathrm{KL}(P_{f_0}(\cdot|z) \| P_{f_x}(\cdot|z)) \tag{3.99}$$

$$\leqslant n_x \cdot \| f_0 - f_x \|_\infty^2. \tag{3.100}$$

Therefore,

$$\Pr_{f_x} \left[\widehat{f}_x = f_x \right] \leqslant \frac{1}{2} + \sqrt{\frac{1}{4} n_x \varepsilon_n^2} \leqslant \frac{1}{2} + \sqrt{\frac{n \| f_x - f_0 \|_\infty^2}{2|\mathcal{F}_n|}}. \tag{3.101}$$

Proof of Lemma 31 By construction, $n \sup_{f_x \in \mathcal{F}_x} \| f_x - f_0 \|_\infty^2 \lesssim M^2 n h^{2\alpha}$ and $|\mathcal{F}_n| = |H_n| \gtrsim [\underline{C}_\varepsilon \varepsilon_n^\mathsf{L}(f_0)]^{2+d/\alpha} n h^{-d}$. Note also that $h \asymp (\varepsilon/M)^{1/\alpha} \asymp (\underline{C}_\varepsilon \varepsilon_n^\mathsf{L}(f_0)/M)^{1/\alpha}$ because $\| f_x - f_0 \|_\infty = \varepsilon = \underline{C}_\varepsilon \varepsilon_n^\mathsf{L}(f_0)$. Subsequently,

$$\frac{n \sup_{f_x \in \mathcal{F}_x} \| f_x - f_0 \|_\infty^2}{2|\mathcal{F}_n|} \lesssim \frac{n[\underline{C}_\varepsilon \varepsilon_n^\mathsf{L}(f_0)]^2}{n[\underline{C}_\varepsilon \varepsilon_n^\mathsf{L}(f_0)]^2 \cdot M^{d/\alpha}} = M^{-d/\alpha}. \tag{3.102}$$

By choosing the constant $M > 0$ to be sufficiently large, the right-hand side of the above inequality is upper bounded by $1/36$. The lemma is thus proved.

3.5.5 Proof of Theorem 6

The proof of Theorem 6 is similar to the proof of Theorem 5, but is much more standard by invoking the *Fano's inequality* (Tsybakov, 2009). In particular, adapting the Fano's inequality on any finite function class $\mathcal{F}_n$ constructed we have the following lemma:

Lemma 32 (Fano's inequality). *Suppose $|\mathcal{F}_n| \geqslant 2$, and $\{(x_i, y_i)\}_{i=1}^n$ are i.i.d. random variables. Then*

$$\inf_{\widehat{f}_x} \sup_{f_x \in \mathcal{F}_n} \Pr_{f_x} \left[\widehat{f}_x \neq f_x \right] \geqslant 1 - \frac{\log 2 + n \cdot \sup_{f_x, f_{x'} \in \mathcal{F}_n} \mathrm{KL}(P_{f_x} \| P_{f_{x'}})}{\log |\mathcal{F}_n|}, \tag{3.103}$$

where P_{f_x} denotes the distribution of (x, y) under the law of f_x.

Let $\mathcal{F}_n$ be the function class constructed in the previous proof of Theorem 6, corresponding to the largest packing set H_n of $L_{f_0}(\widetilde{\varepsilon}_n^\mathsf{L})$ such that $B_h^\infty(x)$ for all $x \in H_n$ are disjoint, where $h \asymp (\widetilde{\varepsilon}_n^\mathsf{L}/M)^{1/\alpha}$ such that $\| \varphi_{x,h} \|_\infty = 2\widetilde{\varepsilon}_n^\mathsf{L}$ for all $x \in H_n$. Because f_0 satisfies (A2'), we have that $|\mathcal{F}_n| = |H_n| \gtrsim \mu_{f_0}(\widetilde{\varepsilon}_n^\mathsf{L}) h^{-d}$. Under the condition that $\varepsilon_n^\mathsf{U}(f_0) \leqslant \widetilde{\varepsilon}_n^\mathsf{L}$, it holds that $\mu_{f_0}(\widetilde{\varepsilon}_n^\mathsf{L}) \gtrsim [\widetilde{\varepsilon}_n^\mathsf{L}]^{2+d/\alpha} n$. Therefore,

$$|\mathcal{F}_n| \gtrsim [\widetilde{\varepsilon}_n^\mathsf{L}]^{2+d/\alpha} \cdot n h^{-d} \gtrsim [\widetilde{\varepsilon}_n^\mathsf{L}]^2 \cdot n M^{d/\alpha}. \tag{3.104}$$

Because $\log(n/\widetilde{\varepsilon}_n^\mathsf{L}) \gtrsim \log n$ and $M > 0$ is a constant, we have that $\log |\mathcal{F}_n| \geqslant c \log n$ for all $n \geqslant N$, where $c > 0$ is a constant depending only on α, d and $N \in \mathbb{N}$ is a constand depending on M.

Let U be the uniform distribution on $\mathcal{X}$. Because $x \sim U$ and $f_x \equiv f_{x'}$ on $\mathcal{X} \backslash B_h^\infty(x)$, we have that

$$\mathrm{KL}(P_{f_x} \| P_{f_{x'}}) = \frac{1}{2} \int_{\mathcal{X}} |f_x(z) - f_{x'}(z)|^2 \mathrm{d}U(z) \tag{3.105}$$

$$\leqslant \frac{1}{2}\Pr_{U}\left[z \in B_h^\infty(x)\right] \cdot \|f_x - f_{x'}\|_\infty^2 \tag{3.106}$$

$$\leqslant \frac{1}{2}\lambda(B_h^\infty(x)) \cdot [\varepsilon_n^{\mathsf{L}}]^2 \tag{3.107}$$

$$\lesssim h^d[\widehat{\varepsilon}_n^{\mathsf{L}}]^2 \lesssim [\widehat{\varepsilon}_n^{\mathsf{L}}]^{2+d/\alpha}/M^{d/\alpha}. \tag{3.108}$$

By choosing M to be sufficiently large, the right-hand side of Eq. (3.103) can be lower bounded by an absolute constant. The theorem is then proved following the same argument as in the proof of Theorem 5.

3.6 Proofs of results in Sec. 3.2

3.6.1 Proof of Lemma 13

We first prove a technical lemma that bounds the ℓ_∞ norm of error vectors.

Lemma 33. *For any $x \in \mathbb{R}^d$ and $z_i \in \{\pm 1\}^d$, with probability $1 - O(d^{-3})$ (conditioned on x_t and z_i)*

$$\left\|\sum_{i=1}^n \varepsilon_i z_i\right\|_\infty \lesssim \frac{\sigma}{\delta}\sqrt{\frac{\log d}{m}} + H\delta.$$

Proof. Let $\overline{\xi}_i = \xi_i/\delta \sim \mathcal{N}(0, \sigma^2/\delta^2)$. Consider the following decomposition:

$$\left\|\sum_{i=1}^n \varepsilon_i z_i\right\|_\infty \leqslant \frac{1}{m\delta}\left\|\sum_{i=1}^m \overline{\xi}_i z_i\right\|_\infty + \delta \cdot \sup_{1 \leqslant i \leqslant m}\left|z_i^\top H_t(\kappa_i, z_i)z_i\right| \cdot \|z_i\|_\infty.$$

The second term on the right-hand side of the above inequality is upper bounded by $O(H\delta)$ almost surely, because $\|z_i\|_\infty \leqslant 1$ and $|z_i^\top H_t(\kappa_i, z_i)z_i| \leqslant \|H_t(\kappa_i, z_i)\|_1\|z_i\|_\infty^2 \leqslant H$. For the first term, because $\overline{\xi}_i$ are centered sub-Gaussian random variables independent of z_i and $\|z_i\|_\infty \leqslant 1$, we have that $1/m \cdot \|\sum_{i=1}^m \overline{\xi}_i z_i\|_\infty \lesssim \sqrt{\sigma^2 \log d/m}$ with probability $1 - O(d^{-3})$, by invoking standard sub-Gaussian concentration inequalities. $\square$

Now define $\widehat{\theta} = (\widehat{g}_t, \widehat{\mu}_t)$, $\theta_0 = (g_t, \delta^{-1}f(x_t))$ and $\overline{Z} = (\overline{z}_1, \ldots, \overline{z}_m)$ where $\overline{z}_i = (z_i, 1) \in \mathbb{R}^{d+1}$. Define also that $Y = (\widetilde{y}_1, \ldots, \widetilde{y}_m)$. The estimator can then be written as $\widehat{\theta} = \arg\min_{\theta \in \mathbb{R}^{d+1}} \frac{1}{m}\|\widetilde{Y} - \overline{Z}\theta\|_2^2 + \lambda\|\theta\|_1$ where $\widetilde{Y} = \overline{Z}\theta_0 + \varepsilon$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_m)$. We first establish a "basic inequality" type results that are essential in performance analysis of Lasso type estimators. By optimality of $\widehat{\theta}$, we have that

$$\frac{1}{m}\|Y - \overline{Z}\widehat{\theta}\|_2^2 + \lambda\|\widehat{\theta}\|_1 \leqslant \frac{1}{m}\|Y - \overline{Z}\theta_0\|_2^2 + \lambda\|\theta_0\|_1 = \frac{1}{m}\|\varepsilon\|_2^2 + \lambda\|\theta_0\|_1.$$

Re-organizing terms we obtain

$$\lambda\|\widehat{\theta}\|_1 \leqslant \lambda\|\theta_0\|_1 + \frac{2}{m}(\widehat{\theta} - \theta_0)^\top\overline{Z}^\top\varepsilon.$$

77

On the other hand, by Hölder's inequality and Lemma 33 we have, with probability $1 - O(d^{-2})$,

$$\frac{2}{m}(\widehat{\theta} - \theta_0)^\top \overline{Z}^\top \varepsilon \leqslant 2\|\widehat{\theta} - \theta_0\|_1 \cdot \left\| \frac{1}{m}\overline{Z}^\top \varepsilon \right\|_\infty \lesssim \|\widehat{\theta} - \theta_0\|_1 \cdot \left(\frac{\sigma}{\delta}\sqrt{\frac{\log d}{m}} + H\delta \right).$$

Subsequently, if $\lambda \leqslant c_0(\sigma\delta^{-1}\sqrt{\log d/m} + H\delta)$ for some sufficiently small $c_0 > 0$, we have that $\|\widehat{\theta}\|_1 \leqslant \|\theta_0\|_1 + 1/2\|\widehat{\theta} - \theta_0\|_1$. Multiplying by 2 and adding $\|\widehat{\theta} - \theta_0\|_1$ on both sides of the inequality we obtain $\|\widehat{\theta} - \theta_0\|_1 \leqslant 2(\|\widehat{\theta} - \theta_0\|_1 + \|\widehat{\theta}_0\|_1 - \|\widehat{\theta}\|_1)$. Recall that θ_0 is sparse and let $\overline{S} = S \cup \{d+1\}$ be the support of θ_0. We then have $\|(\widehat{\theta} - \theta_0)_{\overline{S}^c} + \|(\theta_0)_{\overline{S}^c}\|_1 - \|\widehat{\theta}_{\overline{S}^c}\|_1 = 0$ and hence $\|(\widehat{\theta} - \theta_0)_{\overline{S}^c}\|_1 - \|(\widehat{\theta} - \theta_0)_{\overline{S}}\|_1 \leqslant \|\widehat{\theta} - \theta_0\|_1 \leqslant 2\|(\widehat{\theta} - \theta_0)_{\overline{S}}\|_1$. Thus,

$$\|(\widehat{\theta} - \theta_0)_{\overline{S}^c}\|_1 \leqslant 3\|(\widehat{\theta} - \theta_0)_{\overline{S}}\|_1. \tag{3.109}$$

Now consider $\widehat{\theta}$ that minimizes $\frac{1}{m}\|Y - \overline{Z}\theta\|_2^2 + \lambda\|\theta\|_1$. By KKT condition we have that

$$\left\| \frac{1}{m}\overline{Z}^\top(Y - \overline{Z}\widehat{\theta}) \right\|_\infty \leqslant \frac{\lambda}{2}.$$

Define $\widehat{\Sigma} = \frac{1}{m}\overline{Z}^\top \overline{Z}$ and recall that $Y = \overline{Z}\theta_0 + \varepsilon$. Invoking Lemma 33 and the scaling of λ we have that, with probability $1 - O(d^{-2})$

$$\|\widehat{\Sigma}(\widehat{\theta} - \theta_0)\|_\infty \leqslant \frac{\lambda}{2} + \left\| \frac{1}{m}\overline{Z}^\top \varepsilon \right\| \lesssim \frac{\sigma}{\delta}\sqrt{\frac{\log d}{m}} + \delta H. \tag{3.110}$$

By definition of $\{\overline{z}_i\}_{i=1}^m$, we know that $\widehat{\Sigma}_{jj} = 1$ for all $j = 1, \ldots, d+1$ and $\mathbb{E}[\widehat{\Sigma}_{jk}] = 0$ for $j \neq k$. By Hoeffding's inequality (Hoeffding, 1963) and union bound we have that with probability $1 - O(d^{-2})$, $\|\widehat{\Sigma} - I_{(d+1)\times(d+1)}\|_\infty \lesssim \sqrt{\log d/m}$, where $\|\cdot\|_\infty$ denotes the maximum absolute value of matrix entries. Also note that $\widehat{\theta} - \theta_0$ satisfies $\|(\widehat{\theta} - \theta_0)_{\overline{S}^c}\|_1 \leqslant 3\|(\widehat{\theta} - \theta_0)_{\overline{S}}\|_1$ thanks to Eq. (3.109). Subsequently,

$$\begin{aligned}
\|\widehat{\theta} - \theta_0\|_\infty &\leqslant \|\widehat{\Sigma}(\widehat{\theta} - \theta_0)\|_\infty + \|(\widehat{\Sigma} - I)(\widehat{\theta} - \theta_0)\|_\infty \\
&\leqslant \|\widehat{\Sigma}(\widehat{\theta} - \theta_0)\|_\infty + \|\widehat{\Sigma} - I\|_\infty\|\widehat{\theta} - \theta_0\|_1 \\
&\leqslant \|\widehat{\Sigma}(\widehat{\theta} - \theta_0)\|_\infty + \|\widehat{\Sigma} - I\|_\infty \cdot 4\|(\widehat{\theta} - \theta_0)_{\overline{S}}\|_1 \\
&\leqslant \|\widehat{\Sigma}(\widehat{\theta} - \theta_0)\|_\infty + \|\widehat{\Sigma} - I\|_\infty \cdot 4(s+1)\|\widehat{\theta} - \theta_0\|_\infty \\
&\lesssim \frac{\sigma}{\delta}\sqrt{\frac{\log d}{m}} + \delta H + \sqrt{\frac{s^2\log d}{m}} \cdot \|\widehat{\theta} - \theta_0\|_\infty.
\end{aligned} \tag{3.111}$$

Combining Eq. (3.111) together with the scaling $m = \Omega(s^2 \log d)$ we complete the proof of Lemma 13. Note that the statement on the ℓ_1 error $\|\widehat{\theta} - \theta_0\|_1$ is a simple consequence of the basic inequality Eq. (3.109).

3.6.2 Proof of Lemma 14

Before proving Lemma 14 we state a simple observation on sub-exponentiality of products of sub-Gaussian random variables.

Lemma 34. *Suppose X and Y are centered sub-Gaussian random variables with parameters ν_1^2 and ν_2^2, respectively. Then XY is a centered sub-exponential random variable with parameter $\nu = \sqrt{2}v$ and $\alpha = 2v$, where $v = 2e^{2/e+1}\nu_1\nu_2$.*

Proof. XY is clearly centered because $\mathbb{E}XY = \mathbb{E}X \cdot \mathbb{E}Y = 0$, thanks to independence. We next bound $\mathbb{E}[|XY|^k]$ for $k \geqslant 3$ (i.e., verification of the Bernstein's condition). Because X and Y are independent, we have that $\mathbb{E}[|XY|^k] = \mathbb{E}|X|^k \cdot \mathbb{E}|Y|^k$. In addition, because X is a centered sub-Gaussian random variable with parameter ν_1^2, it holds that $(\mathbb{E}|X|^k)^{1/k} \leqslant \nu_1 e^{1/e}\sqrt{k}$. Similarly, $(\mathbb{E}|X|^k)^{1/k} \leqslant \nu_2 e^{1/e}\sqrt{k}$. Subsequently,

$$\mathbb{E}|XY|^k \leqslant \left(e^{2/e}\nu_1\nu_2\right)^k \cdot k^k \leqslant \left(e^{2/e}\nu_1\nu_2\right)^k \cdot e^k k! \leqslant \frac{1}{2}k! \cdot \left(2e^{2/e+1}\nu_1\nu_2\right)^k.$$

where in the second inequality we use the Stirling's approximation inequality that $\sqrt{2\pi k}k^k e^{-k} \leqslant k!$. The sub-exponential parameter of XY can then be determined. $\qquad\square$

We use the "full-length" parameterization $\widetilde{\theta}_t = \widehat{\theta}_t + \frac{1}{m}\overline{Z}_t^\top(\widetilde{Y}_t - \overline{Z}_t\widehat{\theta}_t)$, where $\widehat{\theta}_t, \overline{Z}_t$ and $\widetilde{Y}_t$ are notations defined in the proof of Lemma 13 (with subscripts t added to emphasize that both Z_t and $\widetilde{Y}_t$ are specific to the tth epoch in Algorithm 5). Because $\widetilde{Y}_t = \overline{Z}_t\theta_{0t} + \varepsilon_t$ (where $\theta_{0t} = \nabla f(x_t)$ and $\varepsilon = (\varepsilon_{t1}, \ldots, \varepsilon_{tm})$, with ε_{ti} defined in Eq. (3.22)). we have

$$\widetilde{\theta}_t = \widehat{\theta}_t + \frac{1}{m}\overline{Z}_t^\top(\overline{Z}_t\theta_{0t} + \varepsilon_t - \overline{Z}_t\widehat{\theta}_t) = \theta_{0t} + \frac{1}{m}\overline{Z}_t^\top\varepsilon_t + (\widehat{\Sigma} - I_{(d+1)\times(d+1)})(\widehat{\theta}_t - \theta_{0t}),$$

where $\widehat{\Sigma} = \frac{1}{n}\overline{Z}_t^\top\overline{Z}_t$. Recall that $\varepsilon_{ti} = \xi_i/\delta + \delta z_i^\top H_t(\kappa_i, z_i)z_i$. Define $b_i = z_i^\top H_t(\kappa_i, z_i)z_i$ and $b = (b_1, \ldots, b_m)$. Also note that the first d components of $\widetilde{\theta}_t$ are identical to $\widetilde{g}_t$ defined in Eq. (3.24). Subsequently,

$$\widehat{g}_t = g_t + \underbrace{\frac{1}{m\delta}Z_t^\top\xi}_{:=\zeta_t} + \underbrace{\frac{\delta}{m}Z_t^\top b + \left[(\widehat{\Sigma} - I_{(d+1)\times(d+1)})(\widehat{\theta}_t - \theta_{0t})\right]_{1:d}}_{:=\gamma_t}. \tag{3.112}$$

In Eq. (3.112) we divide $\widehat{g}_t - g_t$ into two terms. We first consider the term $\zeta_t := \frac{1}{m\delta}Z_t^\top\xi$. It is clear that $\mathbb{E}[\zeta_t|x_t] = 0$ because $\mathbb{E}[\xi|x_t, Z_t] = 0$. Now consider any d-dimensional vector $a \in \mathbb{R}^d$, and to simplify notations all derivations below are conditioned on x_t. For any $i \in [m]$, $z_{ti}^\top a$ are i.i.d. sub-Gaussian random variables with common parameter $\nu^2 = \|a\|_2^2$. Also, $\overline{\xi}_i$ is a sub-Gaussian random variable with parameter σ^2 and is independent of $z_{ti}^\top a$. Thus, invoking Lemma 34 we have that $\xi_i z_{ti}^\top a$ is a sub-exponential random variable with parameters $\nu = \alpha/\sqrt{2} \lesssim \sigma\|a\|_2$. Consequently, $\langle\zeta_t, a\rangle = \frac{1}{m\delta}\sum_{i=1}^m \xi_i z_{ti}^\top a$ is a centered sub-exponential random variable with parameters $\nu = \sqrt{m/2} \cdot \alpha \lesssim \sigma\|a\|_2/\delta\sqrt{m}$.

We next consider the term $\gamma_t = \frac{\delta}{m}Z_t^\top b + (\widehat{\Sigma} - I)(\widehat{\theta}_t - \theta_{0t})$. By Assumption (A3) we know that $\|b\|_\infty \leqslant \delta H$. Subsequently, by Hölder's inequality we have that

$$\|\gamma_t\|_\infty \leqslant \frac{\delta}{m}\|Z_t\|_{1,\infty}\|b\|_\infty + \|\widehat{\Sigma} - I\|_\infty\|\widehat{\theta}_t - \theta_{t0}\|_1$$

$$\lesssim H\delta + \sqrt{\frac{\log d}{m}}\left(\frac{\sigma s}{\delta}\sqrt{\frac{\log d}{m}} + s\delta H\right).$$

where the second inequality holds with probability $1 - O(d^{-2})$ thanks to Lemma 13.

3.6.3 Proof of Theorem 7

Because of the convexity of f, to prove Theorem 7 it suffices to upper bound $\frac{1}{T'}\sum_{t=0}^{T'-1} f(x_t) - f^*$. We next cite the result in (Lan, 2012) that gives explicit cumulative regret bounds for mirror descent with approximate gradients:

Lemma 35 (Lan (2012), Lemma 3)**.** *Let $\|\cdot\|_\psi$ and $\|\cdot\|_{\psi*}$ be a pair of conjugate norms, and let $\Delta_\psi(\cdot,\cdot)$ be a Bregman divergence that is κ-strongly convex with respect to $\|\cdot\|_\psi$. Suppose f is $\widetilde{H}$-smooth with respect to $\|\cdot\|_\psi$, meaning that $f(y) \leqslant f(x) + \nabla f(x)^\top(y-x) + \frac{\widetilde{H}}{2}\|x-y\|_\psi^2$ for all $x, y \in \mathcal{X}$, and $\eta < \kappa/\widetilde{H}$. Define $g_t = \nabla f(x_t)$, and let $x_0, \ldots, x_{T'-1}$ be iterations in Algorithm 5. Then for every $0 \leqslant t \leqslant T' - 1$ and any $x^* \in \widetilde{\mathcal{X}}$,*

$$\eta\left[f(x_{t+1}) - f(x^*)\right] + \Delta_\psi(x_{t+1}, x^*) \leqslant \Delta_\psi(x_t, x^*) + \eta\langle\widetilde{g}_t - g_t, x^* - x_t\rangle + \frac{\eta^2\|\widetilde{g}_t - g_t\|_{\psi*}^2}{2(\kappa - \widetilde{H}\eta)}. \quad (3.113)$$

Adding both sides of Eq. (3.113) from $t = 0$ to $t = T' - 1$, telescoping and noting that $\Delta_\psi(x_{T'}, x^*) \geqslant 0$, we obtain

$$\frac{1}{T'}\sum_{t=0}^{T'-1} f(x_t) - f(x^*) \leqslant \frac{\Delta_\psi(x_0, x^*)}{\eta T'} + \frac{1}{T'}\sum_{t=0}^{T'-1}\langle\widetilde{g}_t - g_t, x_t - x^*\rangle + \frac{\eta}{2(\kappa - H\eta)}\cdot\sup_{0\leqslant t<T'}\|\widetilde{g}_t - g_t\|_{\psi*}^2. \quad (3.114)$$

Set $\|\cdot\|_\psi = \|\cdot\|_a$ for $a = \frac{2\log d}{2\log d - 1}$. It is easy to verify that under Assumption (A3), the function f satisfies

$$f(y) \geqslant f(x) + \nabla f(x)^\top(y-x) + H\|y-x\|_\infty^2$$
$$\geqslant f(x) + \nabla f(x)^\top(y-x) + \widetilde{H}\|y-x\|_\psi^2$$

for all $x, y \in \mathcal{X}$ with $\widetilde{H} \leqslant eH$, because $\|x-y\|_1^2 \leqslant d^{2(1-1/a)}\|x-y\|_a^2 \leqslant d^{1/\log d}\|x-y\|_1^2 = e\|x-y\|_1^2$ by Hölder's inequality. In addition, by definition of Bregman divergence we have that

$$\Delta_\psi(x_0, x^*) \leqslant \frac{1}{2(a-1)}\|x^*\|_a^2 \leqslant \frac{1}{2(a-1)}\|x^*\|_1^2 \leqslant \|x^*\|_1^2\log d \leqslant B^2\log d, \quad (3.115)$$

where the first inequality holds because $\psi_a(x_0) = \psi_a(0) = 0$ and $\nabla\psi_a(x_0) = \nabla\psi_a(0) = 0$ for $a > 1$.

We next upper bound the $\frac{1}{T'}\sum_{t=0}^{T'-1}\langle\widetilde{g}_t - g_t, x^* - x_t\rangle$ and $\|\widetilde{g}_t - g_t\|_{\psi*}^2$ terms. By Lemma 14 and sub-exponential concentration inequalities (e.g., Lemma 91), we have that with probability $1 - O(d^{-1})$

$$\|\widetilde{g}_t - g_t\|_\infty \leqslant \|\zeta_t\|_\infty + \|\gamma_t\|_\infty \lesssim \frac{\sigma}{\delta}\left(\sqrt{\frac{\log d}{m}} + \frac{\log d}{m}\right) + H\delta + \frac{\sigma s\log d}{\delta m} \lesssim \frac{\sigma}{\delta}\sqrt{\frac{\log d}{m}} + H\delta$$

uniformly over all $t' \in \{0, \ldots, T' - 1\}$, where the last inequality holds because $m = \Omega(s^2 \log d)$. Subsequently, by Hölder's inequality we have that

$$\sup_{0 \leqslant t < T'} \|\widetilde{g}_t - g_t\|_{\psi*}^2 \leqslant d^{2(a-1)/a} \cdot \sup_{0 \leqslant t < T'} \|\widetilde{g}_t - g_t\|_\infty^2 \lesssim \frac{\sigma^2 \log d}{\delta^2 m} + H^2 \delta^2. \tag{3.116}$$

We now consider the first term $\frac{1}{T'} \sum_{t=0}^{T'-1} \langle \widetilde{g}_t - g_t, x^* - x_t \rangle \leqslant \frac{1}{T'} \sum_{t=0}^{T'-1} X_t + \sup_{0 \leqslant t \leqslant T'-1} \|\gamma_t\|_\infty \|x^* - x_t\|_1$, where $X_t := \langle \zeta_t, x^* - x_t \rangle$. By Lemma 14, we know that $X_t | X_1, \ldots, X_{t-1}$ is a centered sub-exponential random variable with parameters $\nu = \sqrt{m/2} \cdot \alpha \lesssim \sigma \|x^* - x_t\|_2 / \delta\sqrt{n} \lesssim \sigma \|x^*\|_1 / \delta\sqrt{m}$. Invoking concentration inequalities for sub-exponential martingales ((Victor, 1999), also phrased as Lemma 92 for a simplified version in the appendix) and the definition that $T' = n/m$, we have with probability $1 - O(d^{-1})$

$$\left| \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \zeta_t, x^* - x_t \rangle \right| \lesssim \frac{\sigma \|x^*\|_1}{\delta} \left(\sqrt{\frac{\log d}{n}} + \frac{\log d}{n} \right) \lesssim \frac{\sigma \|x^*\|_1}{\delta} \sqrt{\frac{\log d}{n}},$$

where the last inequality holds because $n \geqslant m = \Omega(s^2 \log d)$. Thus,

$$\left| \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \widetilde{g}_t - g_t, x^* - x_t \rangle \right| \lesssim \frac{\sigma \|x^*\|_1}{\delta} \sqrt{\frac{\log d}{n}} + \|x^*\|_1 \left(H\delta + \frac{\sigma s \log d}{\delta m} \right). \tag{3.117}$$

Combining Eqs. (3.115,3.116,3.117) with Eq. (3.114) and taking x^* to be a minimizer of f on $\mathcal{X}$ that satisfies $\|x^*\|_1 \leqslant B$, we obtain

$$\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*)$$
$$\lesssim \frac{\|x^*\|_1^2 \log d}{\eta} \frac{m}{n} + \frac{\sigma \|x^*\|_1}{\delta} \sqrt{\frac{\log d}{n}} + \|x^*\|_1 \left(H\delta + \frac{\sigma s \log d}{\delta m} \right) + \eta \left(\frac{\sigma^2 \log d}{\delta^2 m} + H^2 \delta^2 \right)$$
$$\leqslant \frac{B^2 \log d}{\eta} \frac{m}{n} + \frac{\sigma B}{\delta} \sqrt{\frac{\log d}{n}} + B \left(H\delta + \frac{\sigma s \log d}{\delta m} \right) + \eta \left(\frac{\sigma^2 \log d}{\delta^2 m} + H^2 \delta^2 \right) \tag{3.118}$$

with probability $1 - O(d^{-1})$, provided that $\eta < \kappa/2H = 1/2H$.

We are now ready to prove Theorem 7. By the conditions we impose on n and the choices of η and m, it is easy to verify that $\eta < 1/2H$, $m = \Omega(s^2 \log d)$ and $m = O(n)$. Subsequently,

$$\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*)$$
$$\lesssim B \sqrt{\frac{m \log d}{n}} + \sigma B \sqrt{\frac{m}{sn}} + B(\sigma + H) \sqrt{\frac{s \log d}{m}} + B \sqrt{\frac{m \log d}{n}} \left(\frac{\sigma^2}{s} + \widetilde{O}(m^{-1}) \right)$$
$$\lesssim B \left(\frac{(1+H)^2 s \log^2 d}{n} \right)^{1/4} + \frac{\sigma B \sqrt{(1+H)}}{s^{1/4} n^{1/4}} + \frac{B(\sigma + H)}{\sqrt{1+H}} \left(\frac{s \log^2 d}{n} \right)^{1/4}$$

81

$$+ B \left(\frac{(1+H)^2 s \log d}{n} \right)^{1/4} \left(\frac{\sigma^2}{s} + \widetilde{O}(n^{-1/2}) \right)$$

$$\lesssim \left(B\sqrt{\log d} + \frac{\sigma B}{\sqrt{s}} + \frac{\sigma^2 B}{s} \right) \left[\frac{(1+H)^2 s}{n} \right]^{1/4} + B(\sigma + \sqrt{H})\sqrt{\log d} \left[\frac{s}{n} \right]^{1/4} + \widetilde{O}(n^{-1/2})$$

$$\lesssim (1 + \sigma + \sigma^2/s) B\sqrt{\log d} \left[\frac{(1+H)^2 s}{n} \right]^{1/4} + \widetilde{O}(n^{-1/2}).$$

3.6.4 Proof of Lemma 15

Using the model Eq. (3.22) we can decompose $\widetilde{g}_t(\delta) - g_t$ as

$$\widetilde{g}_t(\delta) - g_t = \frac{\delta}{2}\mathbb{E}\left[(z^\top H_t z)z\right] + \underbrace{\frac{1}{n\delta}Z_t^\top \xi}_{:=\widetilde{\zeta}_t(\delta)} + \underbrace{\frac{\delta}{2n}\sum_{i=1}^n (z_i^\top H_t z_i)z_i - \mathbb{E}[(z^\top H_t z)z]}_{:=\widetilde{\beta}_t(\delta)}$$

$$+ \underbrace{\frac{\delta}{2n}\sum_{i=1}^n (z_i^\top (H_t(\delta z_i) - H_t)z_i)z_i + \left[(\widehat{\Sigma} - I)(\widehat{\theta}_t - \theta_{0t})\right]_{1:d}}_{:=\widetilde{\gamma}_t(\delta)},$$

where $\widehat{\Sigma}$, $\widehat{\theta}_t$ and θ_{0t} are similarly defined as in the proof of Lemma 14. The sub-exponentiality of $\langle \widetilde{\zeta}_t(\delta), a \rangle$ for any $a \in \mathbb{R}^d$ is established in Lemma 14. We next consider $\widetilde{\beta}_t(\delta)$. For any $a \in \mathbb{R}^d$ consider $\langle \widetilde{\beta}_t(\delta), a \rangle = \frac{\delta}{2n}\sum_{i=1}^n X_i(a)$ where $X_i(a) = (z_i^\top H_t z_i)(z_i^\top a) - \mathbb{E}[(z_i^\top H_t z_i)(z_i^\top a)]$ are centered i.i.d. random variables conditioned on H_t and x_t. In addition, $|X_i(a)| \leq 2\|H_t\|_1\|z_i\|_\infty^2 \cdot \|a\|_1\|z_i\|_\infty \lesssim H\|a\|_1$ almost surely. Therefore, $X_i(a)$ is a sub-Gaussian random variable with parameter $\nu = H\|a\|_1$, and hence $\langle \widetilde{\beta}_t(\delta), a \rangle$ is a sub-Gaussian random variable with parameter $\nu = \delta H\|a\|_1/\sqrt{n}$. Finally, for the deterministic term $\widetilde{\gamma}_t(\delta)$, we have that

$$\|\widetilde{\gamma}_t(\delta)\|_\infty \leq \frac{\delta}{2} \sup_{z \in \{\pm 1\}^d} \|H_t(\delta z) - H_t\|_1\|z\|_\infty^2 + \|(\widehat{\Sigma} - I)(\widehat{\theta}_t - \theta_{0t})\|_\infty$$

$$\leq \frac{\delta}{2} \sup_{z \in \{\pm 1\}^d} L \cdot \|\delta z\|_\infty\|z\|_\infty^2 + \|\widehat{\Sigma} - I\|_{\max}\|\widehat{\theta}_t - \theta_{0t}\|_\infty$$

$$\lesssim L\delta^2 + \sqrt{\frac{\log d}{n}} \left(\frac{\sigma s}{\delta}\sqrt{\frac{\log d}{n}} + s\delta H \right)$$

$$\lesssim L\delta^2 + \frac{\sigma s \log d}{n\delta} + s\delta H\sqrt{\frac{\log d}{n}}.$$

3.6.5 Proof of Theorem 8

Because f is convex, it suffices to upper bound $\frac{1}{T'}\sum_{t=0}^{T'-1} f(x_t) - f(x^*)$, where $x^* \in \mathcal{X}$, $\|x^*\|_1 \leq B$ is a minimizer of f over $\mathcal{X}$. Using the strategy in the proof of Theorem 7, this amounts to upper bound (with high probability) $\|\widetilde{g}_t^{\mathrm{tw}} - g_t\|_{\psi*}^2$ and $\frac{1}{T'}\sum_{t=0}^{T'-1}\langle \widetilde{g}_t^{\mathrm{tw}} - g_t, x^* - x_t \rangle$.

82

For the first term, using sub-exponentiality of $\widetilde{\zeta}_t$ and sub-gaussianity of $\widetilde{\beta}_t$, we have with probability $1 - O(d^{-1})$ uniformly over all $t \in \{0, \dots, T'-1\}$,

$$
\begin{aligned}
\|\widetilde{g}_t^{\mathrm{tw}} - g_t\|_\infty &\leqslant \|\widetilde{\zeta}_t\|_\infty + \|\widetilde{\beta}_t\|_\infty + \|\widetilde{\gamma}_t\|_\infty \\
&\lesssim \frac{\sigma}{\delta}\left(\sqrt{\frac{\log d}{m}} + \frac{\log d}{m}\right) + \delta H\sqrt{\frac{\log d}{m}} + L\delta^2 + H\delta\sqrt{\frac{s^2 \log d}{m}} + \frac{\sigma s \log d}{\delta m} \\
&\lesssim \left(\frac{\sigma}{\delta} + s\delta H\right)\sqrt{\frac{\log d}{m}} + L\delta^2,
\end{aligned}
$$

where the last inequality holds because $m = \Omega(s^2 \log d)$. Subsequently, with probability $1 - O(d^{-1})$

$$
\sup_{0 \leqslant t \leqslant T'-1} \|\widetilde{g}_t^{\mathrm{tw}} - g_t\|_{\psi*}^2 \lesssim \left(\frac{\sigma^2}{\delta^2} + s^2\delta^2 H^2\right)\frac{\log d}{m} + L^2\delta^4. \tag{3.119}
$$

For the other term $\frac{1}{T'}\sum_{t=0}^{T'-1}\langle \widetilde{g}_t^{\mathrm{tw}} - g_t, x^* - x_t\rangle$, again using concentration inequalities of sub-exponential/sub-Gaussian martingales and noting that $\|x^* - x_t\|_2 \leqslant \|x^* - x_t\|_1 \leqslant 2B$, we have

$$
\begin{aligned}
\frac{1}{T'}\sum_{t=0}^{T'-1}\langle \widetilde{g}_t^{\mathrm{tw}} - g_t, x^* - x_t\rangle &= \frac{1}{T'}\sum_{t=0}^{T'-1}\langle \widetilde{\zeta}_t + \widetilde{\beta}_t + \widetilde{\gamma}_t, x^* - x_t\rangle \\
&\lesssim \left(\frac{\sigma}{\delta} + s\delta H\right)B\sqrt{\frac{\log d}{n}} + B\left(L\delta^2 + \frac{\sigma s \log d}{\delta m} + s\delta H\sqrt{\frac{\log d}{m}}\right). \tag{3.120}
\end{aligned}
$$

Subsequently, combining Eqs. (3.119,3.120) with Eq. (3.114) we have

$$
\begin{aligned}
&\frac{1}{T'}\sum_{t=0}^{T'-1} f(x_t) - f(x^*) \\
&\lesssim \frac{B^2 \log d}{\eta}\frac{m}{n} + \left(\frac{\sigma}{\delta} + s\delta H\right)B\sqrt{\frac{\log d}{n}} + (B+\eta)\left(L\delta^2 + \frac{\sigma s \log d}{\delta m} + s\delta H\sqrt{\frac{\log d}{m}}\right) \\
&\quad + \eta\left(\frac{\sigma^2}{\delta^2} + s^2\delta^2 H^2\right)\frac{\log d}{m} + \eta L^2\delta^4. \tag{3.121}
\end{aligned}
$$

We are now ready to prove Theorem 8. It is easy to verify that with the condition imposed on n and the selection of η and m, it holds that $\eta < 1/2H$, $m = \Omega(s^2 \log d)$ and $m \leqslant n/10$. Subsequently,

$$
\begin{aligned}
&\frac{1}{T'}\sum_{t=0}^{T'-1} f(x_t) - f(x^*) \\
&\lesssim Bm^{1/3}\sqrt{\frac{\log d}{n}} + \left[\sigma\left(\frac{m}{s \log d}\right)^{1/3} + \widetilde{O}(m^{-1/3})\right]B\sqrt{\frac{\log d}{n}}
\end{aligned}
$$

83

$$+ \left(B + \tilde{O}\left(\frac{m^{2/3}}{\sqrt{n}} \right) \right) \left[(L + \sigma) \left(\frac{s \log d}{m} \right)^{2/3} + \tilde{O}(m^{-5/6}) \right]$$

$$+ B m^{2/3} \sqrt{\frac{\log d}{n}} \left(\sigma^2 \left(\frac{m}{s \log d} \right)^{2/3} + \tilde{O}(m^{-2/3}) \right) \frac{\log d}{m}$$

$$+ B m^{2/3} \sqrt{\frac{\log d}{n}} L^2 \left(\frac{s \log d}{m} \right)^{4/3}$$

$$\lesssim B m^{1/3} \sqrt{\frac{\log d}{n}} + \sigma B \left(\frac{m}{s \log d} \right)^{1/3} \sqrt{\frac{\log d}{n}} + B(L + \sigma) \left(\frac{s \log d}{m} \right)^{2/3}$$

$$+ \sigma^2 B \left(\frac{m}{s^2 \log^2 d} \right)^{1/3} \sqrt{\frac{\log d}{n}} + \tilde{O}(n^{-5/12})$$

$$\lesssim \left(B \sqrt{\log d} + \frac{\sigma B \sqrt{\log d}}{s^{1/3}} + \frac{\sigma^2 B \sqrt{\log d}}{s^{2/3}} \right) \left[\frac{(1 + L) s^{2/3}}{n} \right]^{1/3}$$

$$+ \frac{B(L + \sigma)}{(1 + L)^{2/3}} \left(\frac{s^{2/3} \log d}{n} \right)^{1/3} + \tilde{O}(n^{-5/12})$$

$$\lesssim \left(B \sqrt{\log d} + \frac{\sigma B \sqrt{\log d}}{s^{1/3}} + \frac{\sigma^2 B \sqrt{\log d}}{s^{2/3}} \right) \left[\frac{(1 + L) s^{2/3}}{n} \right]^{1/3}$$

$$+ B \sigma \sqrt{\log d} \left(\frac{(1 + L) s^{2/3}}{n} \right)^{1/3} + \tilde{O}(n^{-5/12})$$

$$\lesssim (1 + \sigma + \sigma^2 / s^{2/3}) B \sqrt{\log d} \left(\frac{(1 + L) s^{2/3}}{n} \right)^{1/3} + \tilde{O}(n^{-5/12}).$$

3.6.6 Proof of Lemma 16

Note that under minimal regularity conditions (under regularity conditions allowing the swapping of differentiation and integration operators, $\mathbb{E}[\nabla \log p(u)] = \int p(u) \frac{\nabla p(u)}{p(u)} du = \int \nabla p(u) du = \nabla [\int p(u) du] = 0$, where $\int p(u) du = 1$), $\mathbb{E}_\mu[\nabla \log p(u)] = 0$ and therefore $\mathbb{E}[y_t \nabla \log p(u_t)] = f(x_t) \mathbb{E}[\nabla \log p(u_t)] = 0$. The proposition is then immediate by Eq. (3.27).

3.6.7 Proof of Lemma 17

By the mean-value theorem, for any $u \in \mathbb{R}^d$ there exists $\delta(u) \in (0, 1)$ such that

$$\left| f(x + u) - f(x) \right| = \langle \nabla f(x + \delta(u) u), u \rangle. \tag{3.122}$$

By definition of dual norms (also known as the Cauchy-Schwarz inequality), for any vectors a, b and pairs of dual norms $\| \cdot \|_{\mathfrak{p}}, \| \cdot \|_{\mathfrak{q}}$ it holds that $|\langle a, b \rangle| \leq \|a\|_{\mathfrak{p}} \|b\|_{\mathfrak{q}}$. Subsequently

$$\left| f(x + u) - f(x) \right| \leq \|\nabla f(x + \delta(u) u)\|_{\mathfrak{q}} \|u\|_{\mathfrak{p}}. \tag{3.123}$$

Using the fact that $\|\nabla f(z)\|_{\mathfrak{q}} \leqslant L$ for all z we have $|f(x+u) - f(x)| \leqslant L\|u\|_{\mathfrak{p}}$ for all u, and hence

$$\left|\widetilde{f}(x) - f(x)\right| = \left|\mathbb{E}_\mu[f(x+u) - f(x)]\right| \leqslant \mathbb{E}_\mu|f(x+u) - f(x)| \leqslant L\mathbb{E}_\mu\|u\|_{\mathfrak{p}}. \tag{3.124}$$

The second inequality is implied by the following lemma, with $\beta = \mathfrak{p}$ and $r = 1$.

Lemma 36. *Suppose* $u = (u_1, \cdots, u_d)$ *with each component independently sampled from the density* $p(u) = \frac{\mathfrak{p}^{1-1/\mathfrak{p}}}{2\delta\Gamma(1/\mathfrak{p})} \exp\{-|u|^{\mathfrak{p}}/\mathfrak{p}\delta^{\mathfrak{p}}\}$ *for some* $\mathfrak{p} \in (1, 2]$. *Then for any* $\beta, r \geqslant 1$, *we have*

$$\mathbb{E}\|u\|_\beta^r \lesssim \delta^r d^{r/\beta} \log^r d \left[1 + \frac{\delta\mathfrak{p}^{1-1/\mathfrak{p}}r!}{\Gamma(1/\mathfrak{p})}\right].$$

Proof. For every $t > 0$,

$$\Pr[|u_1| \geqslant t] = \frac{\mathfrak{p}^{1-1/\mathfrak{p}}}{\delta\Gamma(1/\mathfrak{p})} \int_t^\infty e^{-x^{\mathfrak{p}}/\mathfrak{p}\delta^{\mathfrak{p}}} \mathrm{d}x = \frac{\mathfrak{p}^{1-1/\mathfrak{p}}}{\Gamma(1/\mathfrak{p})} \int_{t/\delta}^\infty e^{-z^{\mathfrak{p}}/\mathfrak{p}} \mathrm{d}z \tag{3.125}$$

$$\leqslant \frac{\mathfrak{p}^{1-1/\mathfrak{p}}}{\Gamma(1/\mathfrak{p})} \int_{t/\delta}^\infty e^{-z/2} \mathrm{d}z \quad \forall t/\delta \geqslant 1, \tag{3.126}$$

where the last inequality holds because $e^{-z^{\mathfrak{p}}/\mathfrak{p}} \leqslant e^{-z/2}$ for all $z \geqslant 1$ and $\mathfrak{p} \in (1, 2]$. Subsequently, using a union bound over all $j \in \{1, \cdots, d\}$, we have

$$\Pr_\mu[\|u\|_\infty \geqslant t] \leqslant \frac{2\mathfrak{p}^{1-1/\mathfrak{p}}}{\Gamma(1/\mathfrak{p})} \cdot de^{-t/2\delta}. \quad \forall t \geqslant \delta.$$

For any $\beta, r \geqslant 1$, we have

$$\mathbb{E}_\mu\|u\|_\beta^r = \mathbb{E}_\mu\left(\sum_{i=1}^d |u_i|^\beta\right)^{r/\beta} \leqslant \mathbb{E}_\mu d^{r/\beta}\|u\|_\infty^r \tag{3.127}$$

$$\leqslant d^{r/\beta}\left[(2\delta\log d)^r + \frac{2d\mathfrak{p}^{1-1/\mathfrak{p}}}{\Gamma(1/\mathfrak{p})} \int_{2\delta\log d}^\infty t^r e^{-t/2\delta} \mathrm{d}t\right] \tag{3.128}$$

$$= d^{r/\beta}\left[(2\delta\log d)^r + \delta^{r+1} \cdot \frac{2d\mathfrak{p}^{1-1/\mathfrak{p}}}{\Gamma(1/\mathfrak{p})} \int_{2\log d}^\infty z^r e^{-z/2} \mathrm{d}z\right]. \tag{3.129}$$

Applying integration by parts for r times, we have for all $r \in \mathbb{N}$ that

$$\int_a^\infty z^r e^{-z/2} \mathrm{d}z = 2a^r e^{-a} + 2r \int_a^\infty z^{r-1} e^{-z/2} \mathrm{d}z = \ldots$$

$$= \sum_{\ell=0}^{r-1} \frac{r!}{(r-\ell-1)!} 2^{\ell+1} a^{r-\ell} e^{-a} + r! \int_a^\infty e^{-z/2} \mathrm{d}z$$

$$\lesssim 2^r \cdot \sum_{\ell=0}^r a^\ell e^{-a} \leqslant r \cdot (2a)^r \cdot e^{-a}. \tag{3.130}$$

Subsequently,

$$\mathbb{E}_\mu \|u\|_\beta^r \lesssim d^{r/\beta}\left[(\delta \log d)^r + \delta^{r+1} \cdot \frac{d\mathfrak{p}^{1-1/\mathfrak{p}}}{\Gamma(1/\mathfrak{p})} \cdot r!(\log d)^r \cdot \frac{1}{d}\right]$$

$$\lesssim \delta^r d^{r/\beta} \log^r d + \delta^{r+1} d^{r/\beta} \log^r d \cdot \frac{\mathfrak{p}^{1-1/\mathfrak{p}}r!}{\Gamma(1/\mathfrak{p})}$$

$$= \delta^r d^{r/\beta} \log^r d \left[1 + \frac{\delta \mathfrak{p}^{1-1/\mathfrak{p}}r!}{\Gamma(1/\mathfrak{p})}\right]. \tag{3.131}$$

$\square$

3.6.8 Proof of Lemma 18

By definition of $\widehat{g}_t$ and the conditional independence between y_t, y_t' and u_t, we have that

$$\mathbb{E}_\mu \|\widehat{g}_t\|_\mathfrak{q}^2 = \mathbb{E}_\mu (y_t - y_t')^2 \|\nabla \log p(u_t)\|_\mathfrak{q}^2 \tag{3.132}$$

$$= \mathbb{E}_\mu (f(x_t) - f(x_t + u_t) + \xi_t - \xi_t')^2 \|\nabla \log p(u_t)\|_\mathfrak{q}^2 \tag{3.133}$$

$$= \mathbb{E}_\mu (|f(x_t) - f(x_t + u_t)|^2 + 2\sigma^2)\|\nabla \log p(u_t)\|_\mathfrak{q}^2. \tag{3.134}$$

Here the last identity holds because ξ_t, ξ_t' are independent of $f(x_t), f(x_t + u_t)$, and therefore $\mathbb{E}|f(x_t) - f(x_t + u_t) + \xi_t - \xi_t'|^2 = \mathbb{E}|f(x_t) - f(x_t + u_t)|^2 + 2\mathbb{E}(\xi_t - \xi_t')(f(x_t) - f(x_t + u_t)) + \mathbb{E}|\xi_t - \xi_t'|^2 = \mathbb{E}|f(x_t) - f(x_t + u_t)|^2 + \mathbb{E}|\xi_t - \xi_t'|^2 = \mathbb{E}|f(x_t) - f(x_t + u_t)|^2 + 2\sigma^2$, because $\xi_t, \xi_{t'}$ are independent $\mathcal{N}(0, \sigma^2)$ random variables.

By the mean-value theorem, for any u_t there exists $\delta(u_t) \in (0, 1)$ such that

$$f(x_t + u_t) - f(x_t) = \langle \nabla f(x_t + \delta(u_t)u_t), u_t \rangle.$$

Again using the Cauchy-Schwarz inequality with respect to norm pair $\|\cdot\|_\mathfrak{p}, \|\cdot\|_\mathfrak{q}$ and the uniform boundedness of $\|\nabla f(z)\|_\mathfrak{q}$, we have

$$\left|f(x_t + u_t) - f(x_t)\right| \leqslant \|\nabla f(x_t + \delta(u_t)u_t)\|_\mathfrak{q}^2 \|u_t\|_\mathfrak{p}^2 \leqslant L^2 \|u_t\|_\mathfrak{p}^2. \tag{3.135}$$

We next prove the second inequality. Note that $\mathbb{E}_\mu[\|u\|_\mathfrak{p}^2 \|\nabla \log p(u)\|_\mathfrak{q}^2] = \mathbb{E}_\mu[\|u\|_\mathfrak{p}^2 \|\widetilde{u}\|_\mathfrak{q}^2]/\delta^{2/\mathfrak{p}}$, where $|\widetilde{u}_i| = |u_i|^{\mathfrak{p}-1}$ Using Cauchy-Schwarz inequality, we have that

$$\mathbb{E}_\mu[\|u\|_\mathfrak{p}^2 \|\nabla \log p(u)\|_\mathfrak{q}^2] = \mathbb{E}_\mu[\|u\|_\mathfrak{p}^2 \|\widetilde{u}\|_\mathfrak{q}^2]/\delta^{2/\mathfrak{p}} \leqslant \sqrt{\mathbb{E}\|u\|_\mathfrak{p}^4} \cdot \sqrt{\mathbb{E}\|u\|_{\mathfrak{q}(\mathfrak{p}-1)}^{4(\mathfrak{p}-1)}}/\delta^{2\mathfrak{p}}.$$

Invoking Lemma 36 we complete the proof of Lemma 18.

3.6.9 Proof of Theorem 9

For any $x, y \in \mathbb{R}^d$ define $P_x(y) := \arg\min_{z \in \mathcal{X}}\{\langle y, z - x \rangle + D_\psi(z, x)\}$ as the *prox-mapping*. The following inequality is a classical result in mirror descent analysis.

Lemma 37 ([Nemirovski et al. (2009)](#), Lemma 2.1). *For any $u, x, y \in \mathbb{R}^d$, it holds that*

$$D_\psi(u, P_x(y)) \leqslant D_\psi(u, x) + \langle y, u - x \rangle + \frac{\|y\|_{\mathfrak{q}}^2}{2}. \tag{3.136}$$

The mirror descent update is given by $x_{t+1} = \arg\min_{x \in \mathcal{X}}\{\eta\langle x, \widehat{g}_t(x_t)\rangle + D_\psi(x, x_t)\}$. Using the language of prox-mapping, it can be re-written as $x_{t+1} = P_{x_t}(\eta\widehat{g}_t(x_t))$. Applying Lemma 37 with $u = x^*$, $x = x_t$ and $y = \eta\widehat{g}_t(x_t)$, we have

$$D_\psi(x^*, x_{t+1}) \leqslant D_\psi(x^*, x_t) + \eta\langle \widehat{g}_t(x_t), x^* - x_t \rangle + \frac{\eta^2\|\widehat{g}_t(x_t)\|_{\mathfrak{q}}^2}{2}. \tag{3.137}$$

For simplicity denote $\widetilde{f}_t(\cdot) \equiv \mathbb{E}_\mu[f(\cdot + u)]$. Because $\widetilde{f}_t$ is convex, we have that $\widetilde{f}_t(x^*) \geqslant \widetilde{f}_t(x_t) + \langle\nabla\widetilde{f}_t(x_t), x^* - x_t\rangle$ for any $x^* \in \mathcal{X}$. Plugging this into the above inequality, we have

$$D_\psi(x^*, x_{t+1}) \leqslant D_\psi(x^*, x_t) - \eta(\widetilde{f}_t(x_t) - \widetilde{f}_t(x^*)) + \eta\langle\widehat{g}_t(x_t) - \nabla\widetilde{f}_t(x_t), x^* - x_t\rangle + \frac{\eta^2\|\widehat{g}_t(x_t)\|_{\mathfrak{q}}^2}{2}.$$

Dividing both sides of the above inequality by η, telescoping and using the fact that $\mathbb{E}[\widehat{g}_t(x_t) - \nabla\widetilde{f}_t(x_t)] = 0$, we have

$$\mathbb{E}\left[\sum_{t=1}^n \widetilde{f}_t(x_t) - \widetilde{f}_t(x^*)\right] \leqslant \frac{D_\psi(x^*, x_0)}{\eta} + \eta \cdot \sum_{t=1}^n \mathbb{E}[\|\widehat{g}_t(x_t)\|_{\mathfrak{q}}^2]. \tag{3.138}$$

On the other hand, by Lemmas [17](#) and [18](#), $|\widetilde{f}_t(x_t) - f(x_t)| \leqslant L\mathbb{E}_\mu\|u\|_{\mathfrak{p}}$ and $\mathbb{E}\|\widehat{g}_t(x_t)\|_{\mathfrak{q}}^2 \leqslant \mathbb{E}_\mu[(2\sigma^2 + L^2\|u\|^2)\|\nabla\log p(u)\|_{\mathfrak{q}}^2]$. Subsequently, taking $x^* \in \arg\min_{x \in \mathcal{X}} f(x)$, we have

$$\mathbb{E}\left[\sum_{t=1}^n f(x_t) - f^*\right] \lesssim \frac{D_\psi(x^*, x_0)}{\eta} + L\sum_{t=1}^n \mathbb{E}_\mu[\|u\|_{\mathfrak{p}}] + \eta\sum_{t=1}^n \mathbb{E}_\mu[(\sigma^2 + L^2\|u\|_{\mathfrak{p}}^2)\|\nabla\log p(u)\|_{\mathfrak{q}}^2].$$

Note that $D_\psi(x^*, x_0) \leqslant \|x^*\|_{\mathfrak{p}}^2/2 \leqslant B$ when $x_0 = 0$. Applying again Lemmas [17](#), [18](#) and the scalings of η, δ we complete the proof of Theorem [9](#).

3.6.10 Proof of Lemma [19](#)

By the divergence theorem (a special case of the general Stokes theorem), for any differentiable vector field $F : \mathbb{R}^d \to \mathbb{R}^d$ we have

$$\int_K \sum_{i=1}^d \frac{\partial F_i}{\partial u_i} \mathrm{d}V = \int_{\partial K} \langle F, \ell \rangle \mathrm{d}S, \tag{3.139}$$

where $\mathrm{d}V$ and $\mathrm{d}S$ are the volume and surface integrals on K and ∂K, and ℓ denotes the outer normal vector on ∂K. Defining $F(u) = f(x + \delta u)e_i$ for all $u \in K$, where $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)$ is the ith standard basis vector, we have $\frac{\partial F_i}{\partial u_i} = [\nabla f(x + \delta u)]_i$, $\frac{\partial F_j}{\partial u_j} = 0$ for $j \neq i$ and therefore

$$\int_K \delta\nabla f(x + \delta u)\mathrm{d}V = \int_{\partial K} f(x + \delta v)\ell(v)\mathrm{d}S.$$

Subsequently,

$$\frac{\delta}{\mu_d(K)}\mathbb{E}_{u\sim\nu_K}\left[\nabla f(x+\delta u)\right] = \frac{1}{\mu_{d-1}(\partial K)}\mathbb{E}_{v\sim\sigma_{\partial K}}\left[f(x+\delta v)\ell(v)\right].$$

The lemma is then proved by using the definition that $\widetilde{f}(x) = \mathbb{E}_{u\sim\nu_K}[f(x+\delta u)]$ and interchanging the expectation and integration.

3.6.11 Proof of Lemma 20

The differentiability and convexity of $\widetilde{f}$ can be easily verified by definition. Because $\widetilde{f}$ is a local average of f, $\sup_{x\in\mathcal{X}}|\widetilde{f}(x)| \leqslant \sup_{x\in\mathcal{X}}|f(x)| \leqslant C$. Using the convexity of $\|\cdot\|_{\mathsf{q}}$ we have $\|\nabla\widetilde{f}(x)\|_{\mathsf{q}} = \|\mathbb{E}\nabla f(x+\delta u)\|_{\mathsf{q}} \leqslant \mathbb{E}\|\nabla f(x+\delta u)\|_{\mathsf{q}} \leqslant L$. In addition, by the mean value theorem, for any $x\in\mathbb{R}^d$ and $u\in K$, there exists $\delta'\in(0,\delta)$ such that

$$f(x+\delta u) = f(x) + \delta\langle\nabla f(x+\delta'u), u\rangle.$$

Because f satisfies $|f(x)| \leqslant C$ and $\|\nabla f(x)\|_{\mathsf{q}} \leqslant L$ for all $x\in\mathcal{X}$, we know that $\|\nabla f(x+\delta'u)\|_q \leqslant L$. Applying Hölder's inequality we have

$$\left|f(x+\delta u) - f(x)\right| \leqslant \delta L\|u\|_{\mathfrak{p}}.$$

Consequently,

$$\left|\widetilde{f}(x) - f(x)\right| \leqslant \mathbb{E}_{u\sim\nu_K}\left|f(x+\delta u) - f(x)\right| \leqslant \delta L\cdot\mathbb{E}_{u\sim\nu_K}\left[\|u\|_{\mathfrak{p}}\right].$$

3.6.12 Proof of Lemma 21

For the sake of readability, in this section we shall use p, q instead of $\mathfrak{p}, \mathfrak{q}$ for the norm parameters when no confusion can be caused

Recall that for all $v\in\partial\mathbb{B}_p^d$ with $p\in(1,\infty)$, the outer normal product $\ell(v)$ is unique and has the form

$$\ell(v) = \frac{(\mathrm{sgn}(v_1)|v_1|^{p-1}, \mathrm{sgn}(v_2)|v_2|^{p-1}, \cdots, \mathrm{sgn}(v_d)|v_d|^{p-1})}{\sqrt{|v_1|^{2(p-1)} + |v_2|^{2(p-1)} + \cdots + |v_d|^{2(p-1)}}}.$$

Let $V(\cdot)$ and $S(\cdot)$ denote the volume and surface elements on $\mathbb{B}_p^d$ and $\partial\mathbb{B}_p^d$. Using the divergence theorem in Eq. (3.139) with $F(u) = u$, we have

$$d\times\mu_d(\mathbb{B}_p^d) = \int_{\mathbb{B}_p^d}d\,\mathrm{d}V(u) = \int_{\mathbb{B}_p^d}\sum_{i=1}^d\frac{\partial F_i(u)}{\partial u_i}\mathrm{d}V(u) = \int_{\partial\mathbb{B}_p^d}\langle v, \ell(v)\rangle\mathrm{d}S(v) \tag{3.140}$$

$$= \int_{\partial\mathbb{B}_p^d}\Big\langle v, \big(\mathrm{sgn}(v_1)|v_1|^{p-1}, \cdots, \mathrm{sgn}(v_d)|v_d|^{p-1}\big)\Big\rangle\frac{\mathrm{d}S(v)}{\sqrt{\sum_{i=1}^d|v_i|^{2p-2}}} \tag{3.141}$$

$$= \int_{\partial\mathbb{B}_p^d}\frac{\mathrm{d}S(v)}{\sqrt{\sum_{i=1}^d|v_i|^{2p-2}}}. \tag{3.142}$$

Here the last identity holds because $\sum_{i=1}^{d} |v_i|^p = 1$ for all $v \in \partial\mathbb{B}_p^d$, by definition. Expressing the surface integral in Eq. (3.142) using the uniform measure σ_p^{d-1} on $\partial\mathbb{B}_p^d$, we obtain

$$
d \times \mu_d(\mathbb{B}_p^d) = \mu_{d-1}(\partial\mathbb{B}_p^d) \times \int_{\partial\mathbb{B}_p^d} \frac{\mathrm{d}\sigma_p^{d-1}(v)}{\sqrt{\sum_{i=1}^{d} |v_i|^{2p-2}}}. \tag{3.143}
$$

It remains to estimate the integration term in Eq. (3.143). However, the uniform surface measure σ_p^{d-1} is complicated. To simplify the problem, we consider a closely related measure γ_p^{d-1} on $\partial\mathbb{B}_p^d$, conventionally referred in the literature as the *cone measure* (Naor, 2007; Naor & Romik, 2003). More specifically, for any measurable $A \subseteq \partial\mathbb{B}_p^d$, the cone measure γ_p^{d-1} satisfies

$$
\gamma_p^{d-1}(A) = \mu_d(\{ta : a \in A, 0 \leqslant t \leqslant 1\})/\mu_d(\mathbb{B}_p^d). \tag{3.144}
$$

It is remarked that $\gamma_p^{d-1} \equiv \sigma_p^{d-1}$ for $p \in \{1, 2, \infty\}$, but the two measures are in general different for other values of p. Compared to the uniform surface measure σ_p^{d-1}, the cone measure γ_p^{d-1} has a relatively simple probabilistic interpretation, which was first proved in (Rachev & Ruschendorf, 1991; Schechtman & Zinn, 1990) and also summarized in (Naor, 2007; Naor & Romik, 2003).

Lemma 38 (Naor (2007); Naor & Romik (2003))**.** *Let $X_1, \cdots, X_d$ be i.i.d. random variables with PDF $f_p(x) = 1/(2\Gamma(1 + 1/p))e^{-|x|^p}$. Define $Z := (X_1/\|X\|_p, \cdots, X_d/\|X\|_p)$ as a normalized d-dimensional random vector. Then for any measurable $A \subseteq \partial\mathbb{B}_p^d$, $\Pr[Z \in A] = \gamma_p^d(A)$.*

Remark 25. Clearly $Z \in \partial\mathbb{B}_p^d$ with probability 1. The distribution of Z is commonly referred to as the L_p *norm distribution* (Song & Gupta, 1997), and has found wide applications in statistics and machine learning research (Gupta & Song, 1997; Sinz & Bethge, 2010).

The cone measure γ_p^{d-1} has several analytical advantages of the uniform surface measure σ_p^{d-1}. One of the most important properties of the cone measure is its equivalence to a normalized version of $X = (X_1, \cdots, X_d)$ with *independent* component distributions. Such independence structure gives rises to important concentration properties.

Proposition 13. *For any $p, r \in [1, \infty)$ let X be the random variable distributed according to f_p in Lemma 38. Then $\mathbb{E}|X|^r = \kappa(p, r)$.*

Proof. We have that

$$
\mathbb{E}|X|^r = \frac{1}{\Gamma(1 + 1/p)} \int_0^\infty t^r e^{-t^p} \mathrm{d}t = \frac{1}{\Gamma(1 + 1/p)} \frac{1}{p} \int_0^\infty t^{\frac{r+1}{p} - 1} e^{-t} \mathrm{d}t
$$
$$
= \frac{1}{\Gamma(1 + 1/p)} \frac{1}{p} \Gamma\left(\frac{r+1}{p}\right) = \frac{\Gamma((1+r)/p)}{\Gamma(1/p)}.
$$

Here in the second identity we apply the change-of-variable $t \mapsto t^{1/p}$ and the last equality holds because $\Gamma(1 + 1/p) = \Gamma(1/p)/p$. The proposition is therefore proved. $\qquad\square$

Lemma 39. *For any $p, r \in [1, \infty)$ let $X_1, \cdots, X_d$ be i.i.d. random variables distributed according to f_p in Lemma 38. Then for any $\omega \in [0, 1]$ it holds that*

$$\Pr\left[(1 - d^{-\frac{1-\omega}{2}})\kappa(p, r) \times d \leqslant \sum_{i=1}^{d} |X_i|^r \leqslant (1 + d^{-\frac{1-\omega}{2}})\kappa(p, r) \times d\right] \geqslant 1 - \frac{\kappa(p, 2r)}{\kappa(p, r)^2} d^{-\omega}. \tag{3.145}$$

Proof. Denote $S := \sum_{i=1}^{d} |X_i|^r$. By Proposition 13 we know that $\mathbb{E}S = \kappa(p, r) \times d$. In addition, because $|X_i|^r$ are independent, we have

$$\mathrm{Var}(S) = d \times \mathrm{Var}(|X_i|^r) \leqslant d \times \mathbb{E}|X_i|^{2r} = \kappa(p, 2r) \times d. \tag{3.146}$$

Applying the Chebyshev's inequality, we have that for all $\epsilon > 0$,

$$\Pr\left[|S - \mathbb{E}S| > \epsilon\right] \leqslant \frac{\kappa(p, 2r) \times d}{\epsilon^2}. \tag{3.147}$$

Setting $\epsilon = \kappa(p, r) \times d^{(1+\omega)/2}$ we complete the proof of Lemma 39. $\qquad\square$

By dividing $\partial \mathbb{B}_p^d$ into a regular part (on which the event in Eq. (3.145) holds) and an irregular part (on which the event in Eq. (3.145) potentially fails), we can estimate the integration of $1/\sqrt{\sum_i |v_i|^{2(p-1)}}$ with respect to the cone measure γ_p^{d-1} on $\partial \mathbb{B}_p^d$.

Lemma 40. *For any $p \in (1, \infty)$, we have*

$$\int_{\partial \mathbb{B}_p^d} \frac{\mathrm{d}\gamma_p^{d-1}(v)}{\sqrt{\sum_{i=1}^{d} |v_i|^{2(p-1)}}} = (1 + o(1))\sqrt{\frac{\kappa(p, p)^{2(p-1)/p}}{\kappa(p, 2(p-1))}} d^{1/2 - 1/p} \quad as \ d \to \infty. \tag{3.148}$$

Proof. Using Lemma 38, we can equivalently write $u_i = x_i / \|x\|_p$ where $x_1, \cdots, x_d$ are i.i.d. distributed according to $f_p(x) = 1/(2\Gamma(1 + 1/p))e^{-|x|^p}$. For any $r \in [1, \infty)$, let $\mathcal{A}_r^\omega$ denote the event that Eq. (3.145) occurs with parameter ω. We shall set in the rest of the proof that $\omega = 2/3$. Conditioned on the event $\mathcal{A}_p^{2/3} \cap \mathcal{A}_{2(p-1)}^{2/3}$, we have

$$\sum_{i=1}^{d} |x_i|^p = (1 \pm d^{-1/6})\kappa(p, p) \times d = (1 + o(1))\kappa(p, p) \times d; \tag{3.149}$$

$$\sum_{i=1}^{d} |x_i|^{2(p-1)} = (1 \pm d^{-1/6})\kappa(p, 2(p-1)) \times d = (1 + o(1))\kappa(p, 2(p-1)) \times d. \tag{3.150}$$

Subsequently, under $\mathcal{A}_p^{2/3} \cap \mathcal{A}_{2(p-1)}^{2/3}$ we have that

$$\sqrt{\sum_{i=1}^{d} |v_i|^{2(p-1)}} = \frac{\sqrt{\sum_{i=1}^{d} |x_i|^{2(p-1)}}}{\sqrt{(\sum_{i=1}^{d} |x_i|^p)^{2(p-1)/p}}} = \frac{\sqrt{(1 + o(1))\kappa(p, 2(p-1)) \times d}}{\sqrt{[(1 + o(1))\kappa(p, p) \times d]^{2(p-1)/p}}} \tag{3.151}$$

$$= (1 + o(1))\sqrt{\frac{\kappa(p, 2(p-1))}{\kappa(p,p)^{2(p-1)/p}}} \times d^{1/p-1/2}. \tag{3.152}$$

On the other hand, for all $v \in \partial\mathbb{B}_p^d$ satisfying $\|v\|_p = 1$, applying Hölder's inequality the term $\sqrt{\sum_i |v_i|^{2(p-1)}} = \|v\|_{2(p-1)}^{p-1}$ can be lower bounded as

$$\|v\|_{2(p-1)}^{p-1} \geqslant \left(d^{\frac{1}{2(p-1)} - \frac{1}{p}} \cdot \|v\|_p\right)^{p-1} = d^{1/p-1/2}, \qquad p \in [2, \infty); \tag{3.153}$$

$$\|v\|_{2(p-1)}^{p-1} \geqslant (\|v\|_p)^{p-1} = 1, \qquad p \in [1, 2]. \tag{3.154}$$

Combining Eqs. (3.152,3.153,3.154) and using the total expectation formula we have

$$\int_{\partial\mathbb{B}_p^d} \frac{\mathrm{d}\gamma_p^{d-1}(v)}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-1)}}} = (1 + o(1))\sqrt{\frac{\kappa(p,p)^{2(p-1)/p}}{\kappa(p, 2(p-1))}} \times d^{1/2-1/p} \times \gamma_d^{p-1}(\mathcal{A}_p^{2/3} \cap \mathcal{A}_{2(p-1)}^{2/3})$$

$$\pm \max\{d^{1/2-1/p}, 1\} \times \gamma_d^{p-1}(\neg(\mathcal{A}_p^{2/3} \cap \mathcal{A}_{2(p-1)}^{2/3})). \tag{3.155}$$

Invoking Lemma 39 with $\omega = 2/3$ and the union bound we know that $\gamma_p^{d-1}(\mathcal{A}_p^{2/3} \cap \mathcal{A}_{2(p-1)}^{2/3})) = 1 - O(d^{-2/3})$. Also note that both $\kappa(p,p)$ and $\kappa(p, 2(p-1))$ are positive constants depending only on p. Subsequently,

$$\int_{\partial\mathbb{B}_p^d} \frac{\mathrm{d}\gamma_p^{d-1}(u)}{\sqrt{\sum_{i=1}^d |u_i|^{2(p-1)}}} = (1 + o(1))\sqrt{\frac{\kappa(p,p)^{2(p-1)/p}}{\kappa(p, 2(p-1))}} \times d^{1/2-1/p} + O(d^{-2/3}) \tag{3.156}$$

and the lemma is proved, because $1/2 - 1/p \geqslant -2/3$ for all $p \in [1, \infty)$. $\qquad\square$

Lemma 40 constructs an estimate of the integration term in Eq. (3.143) by replacing the uniform surface measure σ_p^{d-1} with the cone measure γ_p^{d-1} that is easy to deal with. It remains to show that such substitution is valid asymptotically as $d \to \infty$. To this end, we cite the following result due to Naor & Romik (2003), which shows that σ_d^{p-1} and γ_d^{p-1} are close to each other in total variation under high-dimensional settings.

Lemma 41 (Naor & Romik (2003), Theorem 2). *For any $p \in [1, \infty)$, there exists a positive constant $c_p > 0$ depending only on p such that* $\mathrm{TV}(\sigma_p^{d-1}, \gamma_p^{d-1}) \leqslant c_p/\sqrt{d}$.

Combining Lemmas 40 and 41, we have that for all $p \in (1, \infty)$,

$$\int_{\partial\mathbb{B}_p^d} \frac{\mathrm{d}|\gamma_p^{d-1}(v) - \sigma_p^{d-1}(v)|}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-1)}}} \leqslant \mathrm{TV}(\gamma_p^{d-1}, \sigma_p^{d-1}) \times \sup_{v \in \partial\mathbb{B}_p^d} \frac{1}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-1)}}} \tag{3.157}$$

$$\leqslant \frac{c_p}{\sqrt{d}} \times \max\{d^{1/2-1/p}, 1\} = o(1) \times d^{1/2-1/p}. \tag{3.158}$$

Combining Eqs. (3.143,3.145,3.158) we obtain

$$\rho(\mathbb{B}_p^d) = \frac{\mu_{d-1}(\partial\mathbb{B}_p^d)}{\mu_d(\mathbb{B}_p^d)} = d \times (1 + o(1))\sqrt{\frac{\kappa(p, 2(p-1))}{\kappa(p,p)^{2(p-1)/p}}} \times d^{1/p-1/2}$$

$$= (1 + o(1)) \sqrt{\frac{\kappa(p, 2(p-1))}{\kappa(p,p)^{2(p-1)/p}}} \times d^{1/2+1/p}.$$

Lemma 22 is then proved.

3.6.13   Proof of Lemma 23

In this section we use the notation of p, q instead of $\mathfrak{p}, \mathfrak{q}$ for a cleaner presentation, when no confusion will be caused. We first separate the case of $p \in [2, \infty)$ that is relatively easier to prove. For $p = \infty$, with probability one $\ell(v) = e_i$ for some $i \in [d]$, implying $\mathbb{E}\|\ell(v)\|_q^2 = \mathbb{E}\|\ell(v)\|_1^2 = 1$. For $p \in [2, \infty)$, because $q \leqslant 2$, we can apply the Hölder's inequality to obtain

$$\mathbb{E}[\|\ell(v)\|_q^2] \leqslant \mathbb{E}\left[(d^{1/q-1/2}\|\ell(v)\|_2)^2\right] = d^{2/q-1} = d^{1/q-1/p}, \tag{3.159}$$

where the second to last equality holds because $\ell(v)$ is an outer normal vector and therefore always has unit ℓ_2 norm.

We next consider the relatively more difficult case of $p \in (1, 2)$. We shall again use the cone measure γ_p^{d-1} defined in Eq. (3.144) to approximate σ_p^{d-1}, the uniform surface measure on $\partial\mathbb{B}_p^d$. Recall that for all $v \in \partial\mathbb{B}_p^d$, $p \in (1, 2)$, $\ell(v)$ is unique and $\|\ell(v)\|_q^q$ can be written as

$$\|\ell(v)\|_q^2 = \frac{(\sum_{i=1}^d |v_i|^{q(p-1)})^{2/q}}{\sum_{i=1}^d |v_i|^{2(p-1)}} = \|v\|_{q(p-1)}^{2(p-1)}/\|v\|_{2(p-1)}^{2(p-1)} = 1/\|v\|_{2(p-1)}^{2(p-1)}. \tag{3.160}$$

Here the last identity holds because $q(p-1) = p$ and $\|v\|_p = 1$ for all $v \in \partial\mathbb{B}_p^d$.

Under the cone measure γ_p^{d-1}, the random variable v can be equivalently written as $v_i = x_i/\|x\|_p$, where $x = (x_1, \cdots, x_d)$ are i.i.d. distributed with respect to the law $f_p(t) = 1/(2\Gamma(1 + 1/p))e^{-|t|^p}$. Using the equivalent expression $v = x/\|x\|_p$, Eq. (3.160) can be re-formulated as

$$\|\ell(v)\|_q^2 = \|x\|_p^{2(p-1)}/\|x\|_{2(p-1)}^{2(p-1)}. \tag{3.161}$$

Let $\mathcal{A}_p^\omega \cap \mathcal{A}_{2(p-1)}^\omega$ denote the event in which Eq. (3.145) holds for $r \in \{2(p-1), q(p-1)\}$ with $\omega = 1 - 2/q = 2/p - 1 \in (0, 1)$ for $p \in (1, 2)$. We then have, under $\mathcal{A}_p^\omega \cap \mathcal{A}_{2(p-1)}^\omega$, that

$$\sum_{i=1}^d |v_i|^p = (1 + o(1))\kappa(p, p)) \times d; \tag{3.162}$$

$$\sum_{i=1}^d |v_i|^{2(p-1)} = (1 + o(1))\kappa(p, 2(p-1)) \times d. \tag{3.163}$$

Subsequently, under the uniform surface measure σ_p^{d-1}, we have

$$\mathbb{E}_{v\sim\sigma_p^{d-1}}[\|\ell(v)\|_q^2] \leqslant \frac{[(1 + o(1))\kappa(p, q(p-1)) \times d]^{2(p-1)/p}}{(1 + o(1))\kappa(p, 2(p-1)) \times d} \times \sigma_p^{d-1}(\mathcal{A}_p^\omega \cap \mathcal{A}_{2(p-1)}^\omega)$$
$$+ 1 \times \sigma_p^{d-1}(\neg(\mathcal{A}_p^\omega \cap \mathcal{A}_{2(p-1)}^\omega)) \tag{3.164}$$

$$\leqslant O(d^{1-2/p}) + \sigma_p^{d-1}(\neg \mathcal{A}_p^\omega) + \sigma_p^{d-1}(\neg \mathcal{A}_{2(p-1)}^\omega) \tag{3.165}$$

$$= O(d^{1/q-1/p}) + \sigma_p^{d-1}(\neg \mathcal{A}_p^\omega) + \sigma_p^{d-1}(\neg \mathcal{A}_{2(p-1)}^\omega). \tag{3.166}$$

It remains to show that $\sigma_p^{d-1}(\neg \mathcal{A}_p^\omega)$ and $\sigma_p^{d-1}(\neg \mathcal{A}_{2(p-1)}^\omega)$ are small. By Lemma 39, we know that the cone measure version of this claim is true. More specifically, applying Lemma 39 with $\omega = 1 - 2/q \in (0, 1)$, we have

$$\gamma_p^{d-1}(\neg \mathcal{A}_p^\omega) + \gamma_p^{d-1}(\neg \mathcal{A}_{2(p-1)}^\omega) = O(d^{2/q-1}) = O(d^{1/q-1/p}). \tag{3.167}$$

We then need to bound the discrepancy betwee γ_p^{d-1} and σ_p^{d-1} on $\neg \mathcal{A}_p^\omega$ and $\neg \mathcal{A}_{2(p-1)}^\omega$. Unfortunately, the $\mathrm{TV}(\sigma_p^{d-1}, \gamma_p^{d-1}) \leqslant c_p/\sqrt{d}$ bound in Naor & Romik (2003) is not strong enough to establish the desired result. We thus resort to a stronger discrepancy result proved in Naor (2007), summarized below:

Lemma 42 (Naor (2007), Theorem 7)**.** *For any $p \in (1, 2)$, there exists an absolute constant $C > 0$ such that for every measurable $A \subseteq \partial \mathbb{B}_p^d$,*

$$\left| \frac{\sigma_p^{d-1}(A)}{\gamma_p^{d-1}(A)} - 1 \right| \leqslant \frac{C}{\sqrt{d}} \cdot \sqrt{\log\left(\frac{100}{\gamma_p^{d-1}(A)} \right)}. \tag{3.168}$$

Define $\alpha := \max\{(2\omega - 1)/2\omega, 0\}$ and abbreviate $\sigma(A) = \sigma_p^{d-1}(\neg \mathcal{A}_p^\omega)$, $\gamma(A) = \gamma_p^{d-1}(\neg \mathcal{A}_p^\omega)$. Because $\omega \in (0, 1)$, we have $\alpha \in [0, 1/2]$. Invoking Lemma 42 and Eq. (3.167), we have

$$\sigma(A) \leqslant \gamma(A) + \frac{C\gamma(A)}{\sqrt{d}}\sqrt{\log\left(\frac{100}{\gamma(A)} \right)} \tag{3.169}$$

$$= \gamma(A) + \frac{C\gamma^\alpha(A)}{\sqrt{d}}\sqrt{\gamma^{2(1-\alpha)}(A)\log\left(\frac{100}{\gamma(A)} \right)} \tag{3.170}$$

$$= O(d^{-\omega}) + O(d^{-\omega}) \cdot \sqrt{\gamma^{2(1-\alpha)}(A)\log\left(\frac{100}{\gamma(A)} \right)}. \tag{3.171}$$

Here the last line holds because $\gamma(A) = O(d^{-\omega})$ and $\gamma^\alpha(A)/\sqrt{d} = O(d^{-\omega})$, because of the definitions that $\omega = 1 - 2/q = 1/p - 1/q$ and $\alpha \geqslant (2\omega - 1)/2\omega$. Furthermore, because $\gamma(A) = O(d^{-\omega}) = o(1)$ and $2(1 - \alpha) > 0$, we have $\gamma^{2(1-\alpha)}(A)\log(100/\gamma(A)) = o(1)$, and therefore $\sigma(A) = O(d^{-\omega}) = O(d^{1/q-1/p})$. The same reasoning and upper bound apply to $\sigma_p^{d-1}(\neg \mathcal{A}_{2(p-1)}^\omega)$ and $\gamma_p^{d-1}(\neg \mathcal{A}_{2(p-1)}^\omega)$ as well. Plugging both upper bounds into Eq. (3.166) we prove the desired result.

3.7 Proofs of results in Sec. 3.3

3.7.1 Proof of Proposition 10

For any $x \in \mathcal{X}^o$ and $z \in \mathbb{R}^d$ define $\|z\|_x := \sqrt{z^\top \nabla^2 \varphi(x) z}$. The *Dikin ellipsoid* $W_1(x)$ is defined as $W_1(x) := \{z \in \mathbb{R}^d : \|z - x\|_x \leqslant 1\}$ for all $x \in \mathcal{X}^o$. It is a well-known fact that $W_1(x) \subseteq \mathcal{X}$

for all $x \in \mathcal{X}^o$ (Abernethy et al., 2008; Hazan & Levy, 2014; Saha & Tewari, 2011). It remains to verify that $z = x + (\nabla^2\varphi(x) + \delta I_d)^{-1/2}u$ is in $W_1(x)$. To see this, note that

$$\|z - x\|_x^2 = u^\top(\nabla^2\varphi(x) + \delta I_d)^{-1/2}\nabla^2\varphi(x)(\nabla\varphi(x) + \delta I_d)^{-1/2}u$$
$$= \|u\|_2^2 - \delta\|(\nabla^2\varphi(x) + \delta I_d)^{-1/2}u\|_2^2 \leqslant \|u\|_2^2 = 1.$$

Hence, $z \in W_1(x) \subseteq \mathcal{X}$.

3.7.2 Proof of Theorem 11

Our proof of Theorem 11 is roughly divided into three steps. In the first step, we review existing results for the RE algorithms on upper bounding the weak regret against *stationary* benchmarks. In the second step, we present a novel local integration analysis that upper bounds the gap between regret against stationary and dynamic benchmarks using the L_p-norm difference between two smooth and strongly convex functions. Finally, we use a sequence of Hölder's inequality to analyze the restarting procedure in the meta-policy described in the previous section.

Regret against stationary benchmarks. For a sequence of convex functions $f = (f_1, \cdots, f_{T'})$, an admissible policy π, the *weak regret* against any stationary point $x^* \in \mathcal{X}$ is defined as

$$S_\phi(f; x^*) := \mathbb{E}^\pi\left[\sum_{t=1}^{T'} f_t(x_t)\right] - \sum_{t=1}^{T'} f_t(x^*). \tag{3.172}$$

Compared to the regret against dynamic solution sequence R^π defined in Eq. (4.28), in S^π the benchmark solution x^* is forced to be *stationary* among all T' epochs, resulting in smaller regret. In fact, it always holds that $S^\pi(f; x^*) \leqslant R^\pi(f)$ for any f and $x^* \in \mathcal{X}$. In the remainder of this section, we shall refer to S^π as the "weak regret" and R^π as the "strong regret".

The next lemma states existing results on upper bounding the weak regret of the RE policy for *adversarial* function sequences f. The lemma is a simple extension of the weak regret bound in Hazan & Levy (2014), with similar proofs.

Lemma 43. *Fix $1 \leqslant T' \leqslant T$. Let $f = (f_1, \cdots, f_{T'})$ be an* arbitrary *sequence of smooth and strongly convex functions satisfying (A1) through (A5). Suppose φ is a strictly convex κ-self-concordant barrier of $\mathcal{X}$, with $\kappa = O(d)$, and $\eta = d(H + 10\sigma\sqrt{\log T})/\sqrt{2T'}$. Then*

$$S^\pi(f; x^*) = O(\sqrt{T'\log T}), \qquad \text{for all } x^* \in \mathcal{X}^o_{\nu/T}. \tag{3.173}$$

Recall the definition that $\mathcal{X}^o_{\nu/T} := \{x \in \mathcal{X}^o : \forall z \in \mathbb{B}_d(\nu/T), x + z \in \mathcal{X}\}$ is the strict interior of $\mathcal{X}$ that is at least ν/T apart from $\partial\mathcal{X}$. Also, in both results we omit dependency on σ, d, D, ν, H, L and M.

We note that when using this Lemma 3.173 in our later proofs, we will replace x^* in (3.173) by x_t^*, which the is the minimizer of f_t. By Assumption (A3) and the definition of $\mathcal{X}^o_{\nu/T}$, we have $x_t^* \in \mathcal{X}^o_{\nu/T}$.
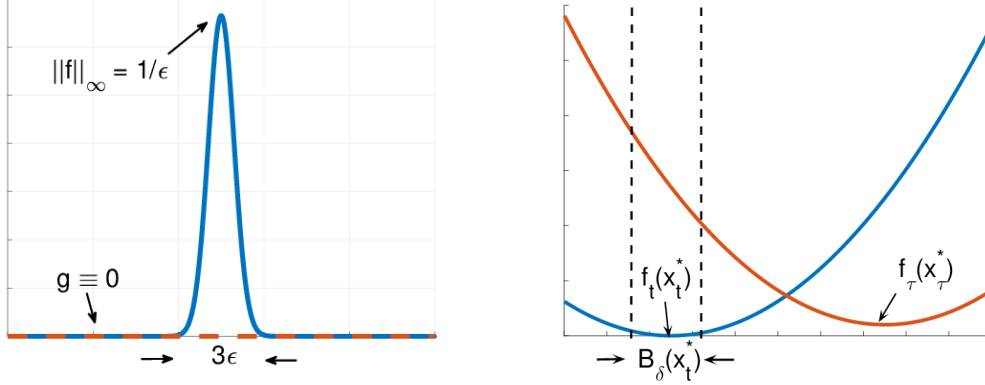
Figure 3.5: The left figure illustrates how two functions f and g can have very different L_p and L_∞ differences ($1 \leqslant p < \infty$). Both functions are defined on $\mathcal{X} = [0, 1]$, with $f(x) = \frac{1}{\sqrt{2\pi}\epsilon} \exp\left\{-\frac{(x-0.5)^2}{2\epsilon^2}\right\}$ and $g \equiv 0$. Because f is the pdf of a univariate Normal distribution with zero mean and ϵ^2 variance, f and g are essentially the same outside of $[0.5 - 3\epsilon, 0.5 + 3\epsilon]$, leading to $\|f - g\|_p \leqslant O(\epsilon^{1/p}) \cdot \|f - g\|_\infty = O(\epsilon^{(1-p)/p})$, which can be arbitrarily smaller than $\|f - g\|_\infty = \Omega(\epsilon^{-1})$ for $1 \leqslant p < \infty$ and ϵ sufficiently small. The right figure provides a graphical explanation of the key argument in the proof of Lemma 44. It shows that when x_τ^* is far away from x_t^*, f_t and f_τ would have a large difference in a neighborhood around x_t^*, because of the strong convexity of f_τ and the smoothness of f_t. Since such difference is upper bounded by $\|f_t - f_\tau\|_p$ on the entire domain $\mathcal{X}$, one can conclude that x_t^* and x_τ^* cannot be too far apart.

Gap between weak and strong regret By definition, the gap between S^π and R^π is independent of policy π:

$$R^\pi(f) - S^\pi(f; x_\tau^*) = \sum_{t=1}^{T'} f_t(x_\tau^*) - f_t(x_t^*), \quad \forall \tau \in \{1, \cdots, T'\}. \tag{3.174}$$

Eq. (3.174) shows that it is possible to upper bound the regret gap by the two-point difference of each function f_t evaluated at the optimal solution x_t^* of f_t and the optimal solution x_τ^* of f_τ, for arbitrary $\tau \in \{1, \cdots, T'\}$. Such differences, however, can be large as x_t^* could be far away from x_τ^* as the functions drift. In the special case of $p = \infty$, Besbes et al. (2015) observes

$$f_t(x_\tau^*) - f_t(x_t^*) = f_t(x_\tau^*) - f_\tau(x_\tau^*) + f_\tau(x_\tau^*) - f_t(x_t^*) \leqslant f_t(x_\tau^*) - f_\tau(x_\tau^*) + f_\tau(x_t^*) - f_t(x_t^*) \tag{3.175}$$

and further bounds both $|f_t(x_\tau^*) - f_\tau(x_\tau^*)|$ and $|f_\tau(x_t^*) - f_t(x_t^*)|$ with $\|f_t - f_\tau\|_\infty$. Such arguments, however, meet significant challenges in the more general setting when $1 \leqslant p < \infty$, because the difference between two functions at one point can be arbitrarily larger than the L_p-norm of the difference of the two functions. We give an illustrative example in Figure 3.5, where two functions f and g are presented, with $\|f - g\|_p / |f(x) - g(x)| \to 0$ for $x = 0.5$ and $p < \infty$.

In this paper we give an alternative analysis that directly upper bounds the left-hand side of Eq. (3.175), $f_t(x_\tau^*) - f_t(x_t^*)$ (i.e., the difference of the same function f_t at two points) using $\|f_t - f_\tau\|_p$, The following is our key affinity lemma:

Lemma 44. *Suppose $\mathcal{X} \subseteq \mathbb{R}^d$. Fix $1 \leqslant p < \infty$, $t \neq \tau$ and let x_t^*, x_τ^* be the minimizers of f_t and f_τ, respectively. Then under (A1) through (A5) we have that*

$$\max\left\{\left|f_t(x_t^*) - f_t(x_\tau^*)\right|, \left|f_\tau(x_t^*) - f_\tau(x_\tau^*)\right|\right\} = O\left(\|f_t - f_\tau\|_p^r\right) \quad \textit{where} \quad r = \frac{2p}{2p + d} \in (0, 1).$$

Proof. Without loss of generality we assume $f_t(x_t^*) \leqslant f_\tau(x_\tau^*)$ throughout this proof. Define $\delta = \|f_t - f_\tau\|_p^{r/2}$. We first prove that $\|x_t^* - x_\tau^*\|_2 \leqslant 2C\delta$, where $C = \max\{\sqrt{(4D^{d/p} + 2L)/M}, 1\}$.

Assume by way of contradiction that $\|x_t^* - x_\tau^*\|_2 > 2C\delta$. For any $x \in \mathcal{X}^o$ and $\alpha \in (0, 1)$ define $\mathcal{X}_\alpha(x) := \{x + \rho(y - x) : 0 \leqslant \rho \leqslant \alpha, y \in \partial\mathcal{X}\}$. It is easy to verify that $\mathcal{X}_\alpha(x) \subseteq \mathcal{X}$ and $\sup_{x' \in \mathcal{X}_\alpha(x)} \|x' - x\|_2 \leqslant \alpha D$ (recall that $D = \sup_{y,y' \in \mathcal{X}} \|y - y'\|_2$ is the diameter of $\mathcal{X}$). In addition, $\mathrm{vol}(\mathcal{X}_\alpha(x)) \geqslant \alpha^d \cdot \mathrm{vol}(\mathcal{X})$, because $\mathcal{X} - x \subseteq \alpha^{-1}[\mathcal{X}_\alpha(x) - x]$, where $\mathcal{X} - x = \{z - x : z \in \mathcal{X}\}$ is the deflation of $\mathcal{X}$ by a specific vector, and similarly $\mathcal{X}_\alpha(x) - x = \{z - x : z \in \mathcal{X}_\alpha(x)\}$. Now set $\alpha = \delta/D$, and note that $\alpha < 1/2$ because $D \geqslant \|x_t^* - x_\tau^*\|_2 > 2C\delta \geqslant 2\delta$. By strong convexity of f_τ, we have $\forall x \in \mathcal{X}_\alpha(x_t^*)$,

$$f_\tau(x) \geqslant f_\tau(x_\tau^*) + \frac{M}{2}\|x_\tau^* - x\|_2^2 \geqslant f_t(x_t^*) + \frac{M}{2}\|x_\tau^* - x\|_2^2 \tag{3.176}$$

$$\geqslant f_t(x_t^*) + \frac{M}{2}(2C\delta - \delta)^2 \geqslant f_t(x_t^*) + \frac{MC^2}{2}\delta^2. \tag{3.177}$$

Here Eq. (3.176) holds because $f_\tau(x_\tau^*) \geqslant f_t(x_t^*)$, and Eq. (3.177) is true because $\|x_t^* - x_\tau^*\|_2 > 2C\delta$ and $\|x - x_t^*\|_2 \leqslant \alpha D = \delta \leqslant C\delta$ for all $x \in \mathcal{X}_\alpha(x_t^*)$. On the other hand, by smoothness of f_t, we have that

$$f_t(x) \leqslant f_t(x_t^*) + \frac{L}{2}\|x - x_t^*\|_2^2 \leqslant f_t(x_t^*) + L\delta^2 \quad \forall x \in \mathcal{X}_\alpha(x_t^*). \tag{3.178}$$

Combining Eqs. (3.177,3.178) we have that, for arbitrary $1 \leqslant p < \infty$ and $x \in \mathcal{X}_\alpha(x_t^*)$

$$\left|f_\tau(x) - f_t(x)\right|^p \geqslant \left|\left(f_t(x_t^*) + \frac{MC^2}{2}\delta^2\right) - \left(f_t(x_t^*) + L\delta^2\right)\right|^p \geqslant (MC^2/2 - L)^p\delta^{2p}, \tag{3.179}$$

provided that $L \leqslant MC^2/2$, which holds true because $C \geqslant \sqrt{2L/M}$ by definition. Integrating both sides of Eq. (3.179) on $\mathcal{X}_\alpha(x_t^*)$ and recalling the definition of $\|f_t - f_\tau\|_p$, we have that

$$\|f_t - f_\tau\|_p^p = \frac{1}{\mathrm{vol}(\mathcal{X})}\int_{\mathcal{X}}|f_t(x) - f_\tau(x)|^p\mathrm{d}x \geqslant \frac{1}{\mathrm{vol}(\mathcal{X})}\int_{\mathcal{X}_\alpha(x_t^*)}|f_t(x) - f_\tau(x)|^p\mathrm{d}x$$

$$\geqslant \frac{\mathrm{vol}(\mathcal{X}_\alpha(x_t^*))}{\mathrm{vol}(\mathcal{X})} \cdot (MC^2/2 - L)^p\delta^{2p} \geqslant \frac{\delta^d}{D^d} \cdot (MC^2/2 - L)^p\delta^{2p}$$

$$\geqslant \frac{(MC^2/2 - L)^p}{D^d}\delta^{2p+d} = \frac{(MC^2/2 - L)^p}{D^d}\|f_t - f_\tau\|_p^p,$$

where the last equality holds because $\delta = \|f_t - f_\tau\|_p^{r/2}$ and $(2p + d) \cdot r/2 = p$. With $C \geqslant \sqrt{(4D^{d/p} + 2L)/M}$, we have that $(MC^2/2 - L)^p/D^d \geqslant 2^d > 1$ and hence the contradiction.

96

We have now established that $\|x_t^* - x_\tau^*\|_2 \leqslant 2C\delta \leqslant O(\delta)$. By smoothness of f_t and f_τ,

$$f_t(x_t^*) \leqslant f_t(x_\tau^*) \leqslant f_t(x_t^*) + \frac{L}{2}\|x_t^* - x_\tau^*\|_2^2 \leqslant f_t(x_t^*) + O(\delta^2);$$

$$f_\tau(x_\tau^*) \leqslant f_\tau(x_t^*) \leqslant f_\tau(x_\tau^*) + \frac{L}{2}\|x_t^* - x_\tau^*\|_2^2 \leqslant f_\tau(x_\tau^*) + O(\delta^2).$$

The proof of Lemma 44 is then completed by plugging in $\delta = \|f_t - f_\tau\|_p^{r/2}$. $\qquad\square$

Analysis of the re-starting procedure Recall that the T epochs are divided into J batches $B_1, \cdots, B_J$ in the meta-policy, with each batch having either Δ_T or $\Delta_T + 1$ epochs. Applying Lemmas 43, 44 together with Eq. (3.174) we have

$$R^\pi(f) \leqslant \sum_{\ell=1}^J \inf_{\tau \in B_\ell} \left\{ S^\pi(f_{\underline{b}_\ell}, \cdots, f_{\overline{b}_\ell}; x_\tau^*) + \sum_{t=\underline{b}_\ell}^{\overline{b}_\ell} f_t(x_\tau^*) - f_t(x_t^*) \right\}$$

$$\leqslant \sum_{\ell=1}^J O(\sqrt{|B_\ell| \log |B_\ell|}) + |B_\ell| \cdot \sup_{t,\tau \in B_\ell} \left| f_t(x_\tau^*) - f_t(x_t^*) \right|$$

$$\leqslant O\left(\frac{T}{\Delta_T} \cdot \sqrt{\Delta_T \log \Delta_T} \right) + O(\Delta_T) \cdot \sum_{\ell=1}^J \sup_{t,\tau \in B_\ell} \|f_t - f_\tau\|_p^r$$

$$\leqslant O\left(\frac{T \log T}{\sqrt{\Delta_T}} \right) + O(\Delta_T) \cdot \sum_{\ell=1}^J \left(\sum_{t=\underline{b}_\ell}^{\overline{b}_\ell - 1} \|f_{t+1} - f_t\|_p \right)^r. \tag{3.180}$$

Here the last inequality holds because (assuming without loss of generality that $\underline{b}_\ell \leqslant t \leqslant \tau \leqslant \overline{b}_\ell$) $\|f_t - f_\tau\|_p \leqslant \sum_{k=t}^{\tau-1} \|f_{k+1} - f_k\|_p \leqslant \sum_{k=\underline{b}_\ell}^{\overline{b}_\ell - 1} \|f_{k+1} - f_k\|_p$.

We next present another key lemma that upper bounds the critical summation term in Eq. (3.180) using J, Δ_T and $\mathrm{Var}_{p,q}(f)$. The proof is based on consecutively applying the Hölder's inequality.

Lemma 45. *Suppose* $\max_{1 \leqslant \ell \leqslant J} |B_\ell| \leqslant \Delta_T + 1$, $1 \leqslant q \leqslant \infty$ *and* $\mathrm{Var}_{p,q}(f) \leqslant V_T$. *Then*

$$\sum_{\ell=1}^J \left(\sum_{t=\underline{b}_\ell}^{\overline{b}_\ell - 1} \|f_{t+1} - f_t\|_p \right)^r \leqslant \Delta_T^{r-r/q} \cdot J^{1-r/q} \cdot T^{r/q} \cdot V_T^r.$$

Proof. By Hölder's inequality, for any d-dimensional vector x we have that

$$\|x\|_\alpha \leqslant \|x\|_\beta \leqslant d^{1/\beta - 1/\alpha}\|x\|_\alpha \qquad \forall\, 0 < \beta \leqslant \alpha \leqslant \infty. \tag{3.181}$$

Apply Eq. (3.181) with $\alpha = q$ and $\beta = 1$ on $x = (\|f_{\underline{b}_\ell+1} - f_{\underline{b}_\ell}\|_p, \cdots, \|f_{\overline{b}_\ell} - f_{\overline{b}_\ell-1}\|_p) \in \mathbb{R}^{|B_\ell|-1}$:

$$\sum_{t=\underline{b}_\ell}^{\overline{b}_\ell - 1} \|f_{t+1} - f_t\|_p = \|x\|_1 \leqslant |B_\ell - 1|^{1-1/q}\|x\|_q \leqslant \Delta_T^{1-1/q} \cdot \left(\sum_{t=\underline{b}_\ell}^{\overline{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q \right)^{1/q}.$$

Subsequently,

$$\sum_{\ell=1}^{J} \left(\sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p \right)^r \leqslant \sum_{\ell=1}^{J} \Delta_T^{r - r/q} \cdot \left(\sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q \right)^{r/q}. \tag{3.182}$$

We next consider $\widetilde{x} = (\widetilde{x}_1, \cdots, \widetilde{x}_J) \in \mathbb{R}^J$, where $\widetilde{x}_\ell = \sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q$. Apply Eq. (3.181) with $\alpha = 1$ and $\beta = r/q$ on $\widetilde{x}$ ($\beta < 1$ because $r \in (0,1)$ and $q \geqslant 1$):

$$\left[\sum_{\ell=1}^{J} \left(\sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q \right)^{r/q} \right]^{q/r} = \|\widetilde{x}\|_{r/q} \leqslant J^{1/\beta - 1/\alpha} \cdot \|\widetilde{x}\|_1 = J^{q/r - 1} \cdot \sum_{\ell=1}^{J} \sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q.$$

Raise both sides of the inequality to the power of r/q and note that $\sum_{\ell=1}^{J} \sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q = \sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_p^q \leqslant T \cdot V_T^q$. We then have

$$\sum_{\ell=1}^{J} \left(\sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p \right)^{r/q} \leqslant J^{1 - r/q} \cdot \left(\sum_{\ell=1}^{J} \sum_{t=\underline{b}_\ell}^{\bar{b}_\ell - 1} \|f_{t+1} - f_t\|_p^q \right)^{r/q} \leqslant J^{1 - r/q} T^{r/q} V_T^r. \tag{3.183}$$

Combining Eqs. (3.182, 3.183) we proved the desired lemma. $\qquad\square$

We now prove Theorem 11 by combining Lemmas 43, 44 and 45 with Eq. (3.180) and setting Δ_T appropriately. By Lemma 43, $S^\pi(f_{\underline{b}_\ell}, \cdots, f_{\bar{b}_\ell}; x_\tau^*) \leqslant O(\sqrt{|B_\ell| \log T}) \leqslant O(\sqrt{\Delta_T \log T})$ for $\phi = \phi_t^{\mathcal{F}}(x_t, f_t)$. Subsequently,

$$R_\phi^\pi(f) \leqslant O(J \sqrt{\Delta_T \log T}) + O(\Delta_T^{1 + r - r/q} J^{1 - r/q} T^{r/q} V_T^r).$$

If $V_T = O(T^{-(6p+d)/4p})$, then we set $\Delta_T = T$, $J = 1$ and obtain regret $\widetilde{O}(\sqrt{T}) + O(T^{1+r} V_T^r) = \widetilde{O}(\sqrt{T})$. Otherwise, when $V_T = \omega(T^{-(6p+d)/4p})$, one selects $\Delta_T \asymp V_T^{-2r/(2r+1)} = V_T^{-4p/(6p+d)}$ and observes that $\Delta_T = o(T)$. This yields a regret of $\widetilde{O}(T \cdot V_T^{2p/(6p+d)})$.

3.7.3 Proof of Theorem 12

Let us first consider the simpler univariate case ($d = 1$). The first step is to reduce the problem of lower bounding regret to the problem of lower bounding success probability of testing sequences of functions, for which tools from information theory such as Fano's lemma (Cover & Thomas, 2006; Ibragimov & Has'minskii, 1981; Tsybakov, 2009; Yu, 1997) could be applied. We then present a novel construction of two functions satisfying (A1) through (A5) and demonstrate that such construction leads to matching lower bounds as presented in Theorem 12. Finally, we extend the lower bound construction to multiple dimensions ($d > 1$) via a change-of-variable argument and complete the proof of general cases in Theorem 12.

Before introducing the proof we first give the definition of an important concept that measures the "discrepancy" between two functions $f, \widetilde{f} : \mathcal{X} \to \mathbb{R}$:

$$\chi(f, \widetilde{f}) := \inf_{x \in \mathcal{X}} \max \left\{ f(x) - f^*, \widetilde{f}(x) - \widetilde{f}^* \right\} \quad \text{where} \quad f^* = \inf_{x \in \mathcal{X}} f(x), \widetilde{f}^* = \inf_{x \in \mathcal{X}} \widetilde{f}(x).$$

Intuitively, $\chi(f, \widetilde{f})$ characterizes the best regret $f(x) - f^*$ one could achieve *without* knowing whether f or $\widetilde{f}$ is the underlying function. This quantity plays a central role in our reduction from regret minimization to testing problems, as well as construction of indistinguishable functions pairs.

From regret minimization to testing   Consider a *finite* subset $\Theta = \{f_1, \cdots, f_M\} \subseteq \mathcal{F}_{p,q}(V_T)$. The following lemma shows that if there exists an admissible policy π that achieves small regret over $\mathcal{F}_{p,q}(V_T)$, then it leads to a hypothesis testing procedure that identifies the true function sequence f in Θ with large probability:

Lemma 46. *Fix $1 \leqslant p < \infty$, $1 \leqslant q \leqslant \infty$ and $V_T > 0$. Let $\Theta \subseteq \mathcal{F}_{p,q}(V_T)$ be a finite subset of sequences of convex functions. Suppose there exists an admissible policy π such that*

$$\sup_{f \in \mathcal{F}_{p,q}(V_T)} R^\pi(f) \leqslant \frac{1}{9} \cdot \inf_{f, \widetilde{f} \in \Theta} \sum_{t=1}^{T} \chi(f_t, \widetilde{f}_t), \tag{3.184}$$

then there exists an estimator $\widehat{f}$ such that

$$\sup_{f \in \Theta} \Pr_f \left[f \neq \widehat{f} \right] \leqslant 1/3, \tag{3.185}$$

where $\Pr_f$ denotes the probability distribution parameterized by the underlying true function sequence $f \in \Theta$.

The proof of Lemma 67 is technical and given later. At a higher level, when there exists an admissible policy π that achieves small regret over $\mathcal{F}_{p,q}(V_T)$ (and hence small regret over $\Theta \subseteq \mathcal{F}_{p,q}(V_T)$ too), then one can correctly identify the underlying function sequence $f \in \Theta$ with large probability by searching all function sequences in Θ and selecting the one that has the smallest regret.

Reduction to testing is a standard approach for proving minimax lower bounds in stochastic estimation and optimization problems (Agarwal et al., 2012; Besbes et al., 2015; Raskutti et al., 2011). Motivations behind such reduction are a well-established class of tools that provide lower bounds on failure probability in testing problems (Ibragimov & Has'minskii, 1981; Tsybakov, 2009; Yu, 1997). Let $\mathrm{KL}(P \| Q) = \int \log \frac{\mathrm{d}P}{\mathrm{d}Q} \mathrm{d}P$ denote the Kullback-Leibler divergence between two distributions P and Q. We introduce the following version of the Fano's inequality,

Lemma 47 (Fano's inequality). *Let $\Theta = \{\theta_1, \cdots, \theta_M\}$ be a finite parameter set of size M. For each $\theta \in \Theta$, let P_θ be the distribution of observations parameterized by θ. Suppose there exists $0 < \beta < \infty$ such that $\mathrm{KL}(P_\theta \| P_{\theta'}) \leqslant \beta$ for all $\theta, \theta' \in \Theta$. Then*

$$\inf_{\widehat{\theta}} \sup_{\theta \in \Theta} \Pr_\theta \left[\widehat{\theta} \neq \theta \right] \geqslant 1 - \frac{\beta + \log 2}{\log M}. \tag{3.186}$$
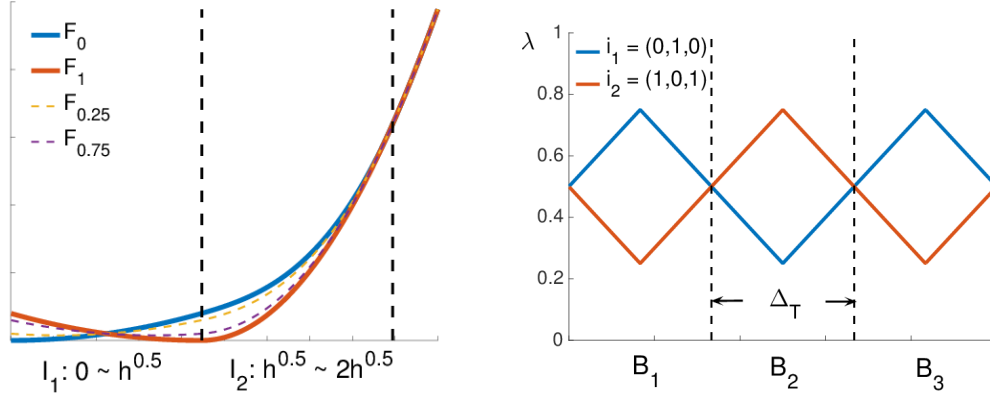
Figure 3.6: The left figure gives a graphical depiction of functions constructed in the lower bound, with thick solid lines corresponding to F_0 and F_1, and thin dashed lines corresponding to F_λ with intermediate values $\lambda = 0.25$ and $\lambda = 0.75$. For the sake of better visualization, the regions $0 \leqslant x \leqslant \sqrt{h}$ (I_1) and $\sqrt{h} \leqslant x \leqslant 2\sqrt{h}$ (I_2) are greatly exaggerated. In the actual construction both regions are very small compared to the entire domain $\mathcal{X} = [0,1]$. The right figure shows the two constructions of function sequences f on $J = 3$ batches, according to Eq. (3.190). At the beginning and the end of each batch the function is always $F_{0.5}$, while within each batch the values of λ first increase and then decrease, or vice versa, depending on the coding $i_j \in \{0,1\}$ for the particular batch. Also note that λ will never be over 0.75 nor under 0.25 throughout the entire construction of the function sequence.

With Lemmas 67 and 47, the question of proving Theorem 12 is reduced to finding a "hard" subset $\Theta \subseteq \mathcal{F}_{p,q}(V_T)$ such that the minimum discrepancy $\inf_{f,\widetilde{f} \in \Theta} \sum_{t=1}^{T} \chi(f_t, \widetilde{f_t})$ is lower bounded and the maximum KL divergence $\sup_{f,\widetilde{f} \in \Theta} \mathrm{KL}(P_f \| P_{\widehat{f}})$ is upper bounded. More precisely, the upper bound on the maximum KL divergence will provide a lower bound for right hand side of Eq. (3.186), which contradicts Eq. (3.185) in Lemma 67. Therefore, the inequality in (3.184) will not hold, which implies a lower bound on the regret. The construction of such a "worst-case example" Θ is highly non-trivial and involves complex design of cubic splines, as we explain in Figure 3.6 and the next paragraph. Below we first give such a construction for the univariate ($d = 1$) case and later extend the construction to higher dimensions.

Univariate constructions Fix $\mathcal{X} = [0,1]$ and $1/8T^2 \leqslant h \leqslant 1/8$. Define $F_0, F_1 : \mathcal{X} \to \mathbb{R}$ as follows:

$$F_0(x) := \begin{cases} x^2, & 0 \leqslant x < \sqrt{h}; \\ \frac{4}{\sqrt{h}}x^3 - 11x^2 + 12\sqrt{h}x - 4h, & \sqrt{h} \leqslant x < 2\sqrt{h}; \\ 8(x - \sqrt{h})^2, & 2\sqrt{h} \leqslant x \leqslant 1. \end{cases} \tag{3.187}$$

$$F_1(x) := \begin{cases} (x - \sqrt{h})^2, & 0 \leqslant x < \sqrt{h}; \\ 8(x - \sqrt{h})^2, & \sqrt{h} \leqslant x \leqslant 1. \end{cases} \tag{3.188}$$

Further define

$$F_\lambda := F_0 + \lambda(F_1 - F_0), \qquad \lambda \in [0,1] \tag{3.189}$$

100

as a convex combination of F_0 and F_1. Figure 3.6 gives a graphical sketch of F_0, F_1 and F_λ. The key insight in the constructions of F_0 and F_1 is to use a cubic function to connect two quadratic functions of different curvatures, and hence allow F_λ to be the same on a wide region of $\mathcal{X}$ (in particular $[2\sqrt{h}, 1]$) and produce small L_p difference $\|F_0 - F_1\|_p$. In contrast, the lower bound construction in existing work (Besbes et al., 2015) uses quadratic functions only, which are not capable of producing smooth functions that differ locally and therefore only applies to the special case of $p = \infty$.

The following lemma lists some properties of F_λ.

Lemma 48. *The following statements are true for all* $\lambda, \mu \in [1/4, 3/4]$.

1. F_λ *satisfies (A1) through (A5) with* $D = 2$, $\nu = 1/64$, $H = 16$, $L = 26$ *and* $M = 2$.
2. $\|F_\lambda - F_\mu\|_\infty \leqslant |\lambda - \mu| \cdot O(h)$ *and* $\|F'_\lambda - F'_\mu\|_\infty \leqslant |\lambda - \mu| \cdot O(\sqrt{h})$.
3. $\|F_\lambda - F_\mu\|_p \leqslant |\lambda - \mu| \cdot O(h^{(2p+1)/2p})$ *for all* $1 \leqslant p < \infty$.
4. $\chi(F_\lambda, F_{1-\lambda}) = |1/2 - \lambda|^2 \cdot h/4$.

We are now ready to describe our construction of a "hard" subset $\Theta \subseteq \mathcal{F}_{p,q}(V_T)$. Note that $\mathcal{F}_{p,\infty}(V_T) \subseteq \mathcal{F}_{p,q}(V_T)$ for all $1 \leqslant q \leqslant \infty$ due to the monotonicity of $\mathrm{Var}_{p,q}(f)$. Therefore we shall focus solely on the $q = \infty$ case, whose construction is automatically valid for all $1 \leqslant q \leqslant \infty$.

Let $1 \leqslant J \leqslant T$ be a parameter to be determined later, and define $\Delta_T = \lfloor T/J \rfloor$. Again partition the entire T time epochs into J disjoint batches $B_1, \cdots, B_J$, where each batch consists of either Δ_T or $\Delta_T + 1$ consecutive epochs. Let $\{0, 1\}^J$ be the class of all binary vectors of length J and let $\mathcal{I} \subseteq \{0, 1\}^J$ be a certain subset of $\{0, 1\}^J$ to be specified later. The subset $\Theta \in \mathcal{F}_{p,\infty}(V_T)$ is constructed so that each function sequence $f_i \in \Theta$ is indexed by a unique J-dimensional binary vector $i \in \mathcal{I}$, with $f_i = (f_{i,1}, \cdots, f_{i,T})$ defined as

$$f_{i,(j-1)\Delta_T + \ell} = \begin{cases} F_{0.5 + 0.5\ell/|B_j|}, & i_j = 0, 1 \leqslant \ell \leqslant \lfloor |B_j|/2 \rfloor; \\ F_{0.75 - 0.5\ell/|B_j|}, & i_j = 0, \lfloor |B_j|/2 \rfloor < \ell \leqslant |B_j|; \\ F_{0.5 - 0.5\ell/|B_j|}, & i_j = 1, 1 \leqslant \ell \leqslant \lfloor |B_j|/2 \rfloor; \\ F_{0.25 + 0.5\ell/|B_j|}, & i_j = 1, \lfloor |B_j|/2 \rfloor < \ell \leqslant |B_j|. \end{cases} \qquad 1 \leqslant j \leqslant J. \qquad (3.190)$$

Figure 3.6 gives a visual illustration of the change pattern of f_i and $f_{i'}$ by plotting the values of λ for each function in the constructed sequences. For a particular batch B_j, when $i_j = i'_j$ then f_i and $f_{i'}$ are exactly the same within B_j; on the other hand, if $i_j = 0$ then f_i will drift towards the function F_0 and if $i'_j = 1$ the functions $f_{i'}$ will drift towards F_1, creating gaps between f_i and $f_{i'}$ within batch B_j. For regularity reasons, we constrain the λ value to be within the range of $(0.25, 0.75)$ regardless of i_j values. We also note that f_i and $f_{i'}$ always agree on the first and the last epochs within each batch. This property makes repetition of constructions across all J batches possible. The following lemma lists some key quantities of interest between f_i and $f_{i'}$:

Lemma 49. *Suppose* $\xi_t \overset{i.i.d.}{\sim} \mathcal{N}(0, 1)$ *For any* $i, i' \in \{0, 1\}^J$ *consider* f_i *and* $f_{i'}$ *as defined in Eq. (3.190). Then the following statements are true:*

1. *(Variation).* $\mathrm{Var}_{p,q}(f) \leqslant \mathrm{Var}_{p,\infty}(f) \leqslant O(h^{(2p+1)/2p}/\Delta_T)$, *for all* $1 \leqslant p < \infty$ *and* $1 \leqslant q \leqslant \infty$.
2. *(Discrepancy).* $\sum_{t=1}^{T} \chi(f_{i,t}, f_{i',t}) \geqslant \Delta_H(i, i') \cdot \Omega(h\Delta_T)$, *where* $\Delta_H(i, i') = \sum_{j=1}^{J} \mathbb{I}[i_j \neq i'_j]$ *is the Hamming distance between* i *and* i'.

3. *(KL divergence).* *Let $P_{f_i}^{\pi}$ be the distribution of $\{f_{i,t}(x_t) + \xi_t\}_{t=1}^T$, with $\{x_t\}_{t=1}^T \subseteq \mathcal{X}$ selected by an admissible policy π. Then for any such policy π we have that*

$$\mathrm{KL}(P_{f_i}^{\pi} \| P_{f_{i'}}^{\pi}) \leqslant \Delta_H(i, i') \cdot O(h^2 \Delta_T).$$

The proof of Lemma 49 is deferred later.

Finally, we describe the construction of $\mathcal{I} \subseteq \{0,1\}^J$ and the choices of J, Δ_T and h that give rise to matching lower bounds. For simplicity we restrict ourselves to J being an even number. The construction of $\mathcal{I}$ is based on the concept of *constant-weight codings*, where each code $i \in \mathcal{I}$ has *exactly* $J/2$ ones and $J/2$ zeros, and each pair of codes $i, i' \in \mathcal{I}$ have large Hamming distance $\Delta_H(i, i') \geqslant J/16$. The construction of constant-weight codings originates from (Graham & Sloane, 1980), and Wang & Singh (2016) gave an explicit lower bound on the size of $\mathcal{I}$, which we cite below:

Lemma 50 (Wang & Singh (2016), Lemma 9)**.** *Suppose $J \geqslant 2$ and J is even. There exists a subset $\mathcal{I} \subseteq \{0,1\}^J$ such that $\forall i \in \mathcal{I}$, $\sum_{j=1}^J i_j = J/2$, and $\forall i, i' \in \mathcal{I}$, $\Delta_H(i, i') \geqslant J/16$. Furthermore, $\log |\mathcal{I}| \geqslant 0.0625J$.*

The univariate case of Theorem 12 can then be proved by appropriately setting the scalings of h, Δ_T and invoking Lemmas 48, 49 and 50. Because an $\Omega(\sqrt{T})$ regret lower bound for *stationary* stochastic online optimization is known (see, for example, (Hazan & Kale, 2014; Jamieson et al., 2012)), we only need to prove the lower bound with the additional assumption that $V_T = \Omega(T^{-(6p+d)/4p})$. More specifically, we set $h \asymp V_T^{2p/(6p+1)}$ and $\Delta_T \asymp V_T^{-4p/(6p+1)}$. It is easy to verify that with the additional lower bound on V_T, $\Delta_T = o(T)$ and $h \gtrsim 1/T^2$, and therefore the constructions are valid. A complete proof is given later after we introduce our adversarial construction of $d > 1$, which includes the univariate setting ($d = 1$) as a special case.

Extension to higher dimensions The lower bound construction can be extended to higher dimensions $d > 1$ to obtain a matching lower bound of $V_T^{2p/(4p+d)} \cdot T$ for noisy gradient feedback and $V_T^{2p/(6p+d)} \cdot T$ for noisy function value feedback. Let $1 = (1, \cdots, 1) \in \mathbb{R}^d$ be a d-dimensional vector with all components equal to 1. We consider $\mathcal{X} = \{x \in \mathbb{R}^d : x \geqslant 0, 1^{\top}x \leqslant 1\}$. Define $\overline{F}_\lambda : \mathcal{X} \to \mathbb{R}$ as follows:

$$\overline{F}_\lambda(x) := F_\lambda(1^{\top}x) + \|x\|_2^2, \qquad \lambda \in [0,1], x \in \mathcal{X}. \tag{3.191}$$

Here F_λ is the univariate function defined in Eq. (3.189). Intuitively, the multi-variate function $\overline{F}$ is constructed by "projecting" a d-dimensional vector x onto a 1-dimensional axis supported on $[0,1]$, and subsequently invoking existing univariate construction of adversarial functions. An additional quadratic term $\|x\|_2^2$ is appended to ensure the strong convexity of $\overline{F}_\lambda$ without interfering with the structure in F_λ. The following lemma lists the properties of $\overline{F}$, which are rigorously verified later this section.

Lemma 51. *Suppose $1/8T^2 \leqslant h \leqslant 1/8$. The following statements are true for any fixed $d \in \mathbb{N}$ and all $\lambda, \mu \in [1/4, 3/4]$.*

1. *$\overline{F}_\lambda$ satisfies (A1) through (A5) with $D = 2$, $\nu = 1/16\sqrt{d+1}$, $H = 16\sqrt{d} + 2$, $L = 26\sqrt{d} + 2$ and $M = 2$.*
2. *$\|\overline{F}_\lambda - \overline{F}_\mu\|_\infty \leqslant |\lambda - \mu| \cdot O(h)$ and $\sup_{x \in \mathcal{X}} \|\nabla \overline{F}_\lambda(x) - \nabla \overline{F}_\mu(x)\|_2 \leqslant |\lambda - \mu| \cdot O(\sqrt{h})$.*

3. $\|\overline{F}_\lambda - \overline{F}_\mu\|_p \leqslant |\lambda - \mu| \cdot O(h^{(2p+d)/2p})$ *for all* $1 \leqslant p < \infty$.

4. $\chi(\overline{F}_\lambda, \overline{F}_{1-\lambda}) = \frac{d}{d+1} \left(\frac{1}{2} - \lambda\right)^2 \cdot h$.

The third property in Lemma 51 deserves special attention, which is a key property that is significantly different from Lemma 48 for the univariate case, because the dependency of $\|\overline{F}_\lambda - \overline{F}_\mu\|_p$ on h has an extra term involving the domain dimension d in the exponent. At a higher level, the presence of the $O(h^{2p/(2p+d)})$ term comes from the *concentration of measure* phenomenon in high dimensions.

We then have the next corollary, by following the same construction of $\Theta \subseteq \mathcal{F}_{p,q}(V_T)$ in the univariate case and invoking Lemma 51:

Corollary 3. *Suppose* $1 \leqslant J \leqslant T$ *is even,* $\Delta_T = \lfloor T/J \rfloor$ *and* $1/8T^2 \leqslant h \leqslant 1/8$. *Let* $\mathcal{I} \subseteq \{0,1\}^J$ *be constructed according to Lemma 50, and* $\Theta = \{f_i : i \in \mathcal{I}\}$, *where* f_i *is defined in Eq. (3.190) except that* F_λ *is replaced with its high-dimensional version* $\overline{F}_\lambda$ *defined in Eq. (3.191). Then the following holds:*

1. *(Variation).* $\sup_{f \in \Theta} \mathrm{Var}_{p,q}(f) \leqslant O(h^{(2p+d)/2p}/\Delta_T)$ *for all* $1 \leqslant p < \infty$, $1 \leqslant q \leqslant \infty$.

2. *(Discrepancy).* $\inf_{f,\widetilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \widetilde{f}_t) \geqslant \Omega(hT)$.

3. *(KL-divergence).* *For all admissible policy* π, $\sup_{f,\widetilde{f} \in \Theta} \mathrm{KL}(P_f^\pi \| P_{\widetilde{f}}^\pi) \leqslant O(h^2 T)$.

We now prove the multi-dimensional case based on Corollary 3. Set $h \asymp V_T^{2p/(6p+d)}$ and Δ_T accordingly such that $\mathrm{Var}_{p,\infty}(f) \leqslant O(h^{(2p+d)/2p}/\Delta_T) = V_T$. This yields $\Delta_T \asymp V_T^{-4p/(6p+d)}$ and $J = T/\Delta_T \asymp TV_T^{4p/(6p+d)}$. The KL divergence is then upper bounded by $O(h^2 T) = O(TV_T^{4p/(6p+d)})$ and $\log|\Theta| = \Omega(J) = \Omega(TV_T^{4p/(6p+d)})$. By carefully selecting constants in the asymptotic notations, one can make the right-hand side of Eq. (3.186) to be lower bounded by $1/2$. Subsequently invoking Lemma 67, we conclude that there does not exist an admissible policy π such that $\sup_{f \in \mathcal{F}_{p,\infty}(V_T)} R^\pi(f) \leqslant 1/9 \cdot \inf_{f,\widetilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \widetilde{f}_t)$. The lower bound proof is then completed by the discrepancy claim in Corollary 3 that $\inf_{f,\widetilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \widetilde{f}_t) \geqslant \Omega(h^2 T) = \Omega(TV_T^{2p/(6p+d)})$.

Proof of Lemma 67 Let π be a policy that attains the minimax rate. By Markov's inequality, with probability $2/3$ it holds that

$$\frac{1}{T}\sum_{t=1}^T f_t(x_t) - f_t(x_t^*) \leqslant \frac{1}{3} \cdot \inf_{f,\widetilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \widetilde{f}_t), \quad \forall f, \widetilde{f} \in \Theta. \tag{3.192}$$

Define $\widehat{f} := \arg\min_{\widehat{f} \in \Theta} \sum_{t=1}^T \widehat{f}_t(x_t) - \widehat{f}_t(\widehat{x}_t^*)$, where $\widehat{x}_t^*$ is the (unique) minimizer of $\widehat{f}_t$. Let $\widehat{f}_t^* = \inf_{x \in \mathcal{X}} \widehat{f}_t(x)$ and $f_t^* = \inf_{x \in \mathcal{X}} f_t(x)$. Because $\widehat{f}$ minimizes the "empirical" regret on $\{x_t\}_{t=1}^T$, it holds that

$$\sum_{t=1}^T \widehat{f}_t(x_t) - \widehat{f}_t^* \leqslant \sum_{t=1}^T f_t(x_t) - f_t^*.$$

Subsequently,

$$\sum_{t=1}^T \chi(\widehat{f}_t, f_t) = \sum_{t=1}^T \inf_{x \in \mathcal{X}} \max\left\{\widehat{f}_t(x) - \widehat{f}_t^*, f_t(x) - f_t^*\right\}$$

103

$$\leqslant \sum_{t=1}^{T} \max\left\{ \widehat{f}_t(x_t) - \widehat{f}_t^*, f_t(x_t) - f_t^* \right\}$$

$$\leqslant \sum_{t=1}^{T} \widehat{f}_t(x_t) - \widehat{f}_t^* + f_t(x_t) - f_t^*$$

$$\leqslant 2\left(\sum_{t=1}^{T} f_t(x_t) - f_t^* \right) \leqslant \frac{2}{3} \cdot \inf_{f,\widetilde{f}\in\Theta} \sum_{t=1}^{T} \chi(f_t, \widetilde{f}_t).$$

Therefore, we must have $\widehat{f} = f$ conditioned on Eq. (3.192), which completes the proof.

Proof of Lemma 48 We verify the properties separately.

Verification of property 1: (A1) is obvious because $\mathcal{X} = [0,1]$. We next focus (A3), (A4) and (A5). It is easy to check that if two functions f and g satisfy (A3) through (A5), then their convex combination $f + \lambda(g - f)$ for $\lambda \in [0,1]$ also satisfies (A3) through (A5). Therefore we only need to verify these conditions for F_0 and F_1, respectively. We first prove that both F_0 and F_1 are differentiable. Because both F_0 and F_1 are differentiable within each piece, to prove the global differentiablity we only need to show that the left and right function values and derivatives of F_0 and F_1 at $x = \sqrt{h}$ and $x = 2\sqrt{h}$ are equal. Define $F(x^+) = \lim_{t\to 0^+} F(x+t)$, $F(x^-) = \lim_{t\to 0^-} F(x+t)$, $F'(x^+) = \lim_{t\to 0^+} \frac{F(x+t)-F(x)}{t}$ and $F'(x^-) = \lim_{t\to 0^-} \frac{F(x+t)-F(x)}{t}$. We then have that $F_1(\sqrt{h}^+) = F_0(\sqrt{h}^-) = h$, $F_0(2\sqrt{h}^+) = F_0(2\sqrt{h}^-) = 8h$, $F_0'(\sqrt{h}^+) = F_0'(\sqrt{h}^-) = 2\sqrt{h}$, $F_0'(2\sqrt{h}^+) = F_0'(2\sqrt{h}^-) = 16\sqrt{h}$, $F_2(\sqrt{h}^+) = F_1(\sqrt{h}^-) = 0$, $F_1'(\sqrt{h}^+) = F_1'(\sqrt{h}^-) = 0$. Therefore, both F_0 and F_1 are differentiable on $[0,1]$. It is then easy to check that $\sup_{0\leqslant x\leqslant 1} \max\{|F_0(x)|, |F_1(x)|\} \leqslant 8$ and $\sup_{0\leqslant x\leqslant 1} \max\{|F_0'(x)|, |F_1'(x)|\} \leqslant 16$. Therefore (A3) is satisfied with $H = 16$.

To verify (A4) and (A5) we need to compute the second-order derivatives of F_0 and F_1. By construction, $F_0''(x) = F_1''(x) = 2$ for $x \in [0, \sqrt{h}]$, $F_0''(x) = F_1''(x) = 8$ for $x \in [\sqrt{h}, 1]$, and $2 \leqslant F_0''(x) \leqslant 26$ for $x \in [\sqrt{h}, 2\sqrt{h}]$. Therefore, F_0 and F_1 satisfy (A4) and (A5) with $L = 26$ and $M = 2$. Note that F_0 and F_1 are *not* twice differentiable at $x = \sqrt{h}$ and $x = 2\sqrt{h}$: however, this does not affect the smoothness and strong convexity of both functions.

Finally we check (A2). Let x_λ^* be the unique minimizer of $F_\lambda = F_0 + \lambda(F_1 - F_0)$. Elementary algebra yields that $x_\lambda^* = \lambda\sqrt{h}$. Because $h \geqslant 1/8T^2$, we know that F_λ satisfies (A2) with $\nu = 1/32$ for $\lambda \in [1/4, 3/4]$.

Verification of property 2: It is easy to see that $\|F_\lambda - F_\mu\|_p = |\lambda - \mu| \cdot \|F_0 - F_1\|_p$ and $\|F_\lambda' - F_\mu'\|_p \leqslant |\lambda - \mu| \cdot \|F_0' - F_1'\|_p$ for all $1 \leqslant p \leqslant \infty$. Thus we only need to consider $\lambda = 0$ and $\mu = 1$. It is easy to verify that $\|F_0 - F_1\|_\infty = |F_0(0) - F_1(0)| = \sqrt{h}$ and $\|F_0' - F_1'\|_\infty = |F_0(0)' - F_1(1)'| = 2h$.

Verification of property 3: Similarly we only need to consider $\lambda = 0$ and $\mu = 1$. Because F_0 and F_1 only differ on $[0, 2\sqrt{h}]$, we have that

$$\|F_0 - F_1\|_p = \left(\int_0^{2\sqrt{h}} |F_0(x) - F_1(x)|^p \mathrm{d}x \right)^{1/p} = O(h^{1/2p}) \cdot \|F_0 - F_1\|_\infty = O(h^{(2p+1)/2p}).$$

Verification of property 4: We have that $x_\lambda^* = \lambda\sqrt{h}$ and $F_\lambda^* = F_\lambda(x_\lambda^*) = \lambda(1 - \lambda)h$. Subsequently, $\chi(F_\lambda, F_{1-\lambda}) = F_\lambda(\sqrt{h}/2) - F_\lambda^* = (1/2 - \lambda)^2 \cdot h/4$.

Proof of Lemma 49 Fix an arbitrary interval I_j for some $j \in \{1, \cdots, J\}$. Without loss of generality assume $|I_j| = \Delta_T$ (the extra one function in some intervals can be safely neglected as both T and Δ_T are large). Then

$$\sup_{t \in I_j} \|f_{t+1} - f_t\|_p = \frac{1}{\Delta_T} \cdot O(h^{(2p+1)/2p}).$$

Subsequently,

$$\mathrm{Var}_{p,\infty}(f) = \sup_{1 \leqslant t \leqslant T-1} \|f_{t+1} - f_t\|_p = O(h^{(2p+1)/2p}/\Delta_T).$$

For the discrepancy term, again fix I_j for some $j \in \{1, \cdots, J\}$ such that $i_j \neq i'_j$. We then have,

$$\sum_{t \in I_j} \chi(f_{i,t}, f_{i',t}) \geqslant 2 \sum_{t=0}^{\lfloor \Delta_T/2 \rfloor} \chi(F_{0.5+t/\Delta_T}, F_{0.5-t/\Delta_T}) \geqslant 2 \sum_{t=0}^{\lfloor \Delta_T/2 \rfloor} \left(\frac{t}{\Delta_T}\right)^2 \cdot \Omega(h) = \Omega(h\Delta_T).$$

Subsequently, summing over all intervals with $i_j \neq i'_j$ we have that $\sum_{t=1}^{T} \chi(f_{i,t}, f_{i',t}) \geqslant \Delta_H(i, i') \cdot \Omega(h\Delta_T)$.

Finally we compute the KL divergence $\mathrm{KL}(P^\pi_{f_i} \| P^\pi_{f_{i'}})$. Let $y_t = f_t(x_t) + \xi_t$ be the random variables of the feedbacks and denote $x^t = (x_1, \cdots, x_t)$ and $y^t = (y_1, \cdots, y_t)$. For any admissible policy π, we have that

$$
\begin{aligned}
\mathrm{KL}(P^\pi_{f_i} \| P^\pi_{f_{i'}}) &= \mathbb{E}_{f_i, \pi} \left[\log \frac{P^\pi_{f_i}(x^T, y^T)}{P^\pi_{f_{i'}}(x^T, y^T)} \right] \\
&= \mathbb{E}_{f_i, \pi} \left[\log \frac{P_{f_i}(y^T|x^T) \cdot \prod_{t=1}^{T} P_\pi(x_t|y^{t-1}, x^{t-1})}{P_{f_{i'}}(y^T|x^T) \cdot \prod_{t=1}^{T} P_\pi(x_t|y^{t-1}, x^{t-1})} \right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{f_i, \pi} \left[\log \frac{P_{f_i,t}(y_t|x_t)}{P_{f_{i'},t}(y_t|x_t)} \right] \\
&\leqslant \sum_{t=1}^{T} \sup_{x \in \mathcal{X}} \mathrm{KL}(P_{f_i,t}(\cdot|x) \| P_{f_{i'},t}(\cdot|x)).
\end{aligned}
$$

Here the third identity holds because ε_t are independent. For $y_t = f_t(x_t) + \xi_t \sim \mathcal{N}(f_t(x_t), 1)$, it holds that

$$\sup_{x \in \mathcal{X}} \mathrm{KL}(P^\phi_{f_i,t}(\cdot|x) \| P^\phi_{f_{i'},t}(\cdot|x)) = \sup_{x \in \mathcal{X}} \left| f_{i,t}(x) - f_{i',t}(x) \right|^2 = \|f_{i,t} - f_{i',t}\|_\infty^2 = O(h^2). \quad (3.193)$$

where in the last inequality we invoke Lemma 48. Summing over $t = 1$ to T we have that $\mathrm{KL}(P^\pi_{f_i} \| P^\pi_{f_{i'}}) = O(h^2 T)$.

Proof of Lemma 51 We verify the properties separately. **Verification of property 1**: Because $\forall x \in \mathcal{X}$, $\|x\|_1 \leqslant 1$, we have that $\|x - y\|_2 \leqslant \|x - y\|_1 \leqslant 2$ for all $x, y \in \mathcal{X}$ and therefore $\mathcal{X}$ satisfies (A1) with $D = 2$.

We next verify (A3). Because F_λ is convex differentiable, it holds that $\overline{F}_\lambda(x) = F_\lambda(1^\top x) + \|x\|_2^2$ is also convex differentiable because convexity is preserved with affine transform. In particular, $\sup_{x \in \mathcal{X}} \overline{F}_\lambda(x) \leqslant \|F_\lambda\|_\infty + 1 \leqslant 9$ and $\sup_{x \in \mathcal{X}} \|\overline{F}_\lambda(x)\|_2 \leqslant \sup_{x \in \mathcal{X}} |F_\lambda(1^\top x)| \cdot \|1\|_2 + 2\|x\|_2 \leqslant 16\sqrt{d} + 2$. Therefore, (A3) is satisfied with $H = 16\sqrt{d} + 2$.

To verify (A4) and (A5), note that $\overline{F}_\lambda$ is twice differentiable except at points $\{x : 1^\top x = \sqrt{h}\} \cup \{x : 1^\top x = 2\sqrt{h}\}$. Furthermore, $\nabla^2 \overline{F}_\lambda(x) = F_\lambda''(1^\top x) \cdot 11^\top + 2I_d$. Subsequently, on points $x \in \mathcal{X}$ where $\overline{F}_\lambda$ is twice differentiable, we have that $\nabla^2 \overline{F}_\lambda(x) \preceq (\|F_\lambda''\|_\infty \sqrt{d} + 2)I = (26\sqrt{d} + 2)I$ and $\nabla^2 \overline{F}_\lambda(x) \succeq 2I$. Therefore, (A4) and (A5) are satisfied with $L = 26\sqrt{d} + 2$ and $M = 2$.

Finally we check (A2). Let x_λ^* be the unique minimizer of $\overline{F}_\lambda$ on $\mathcal{X}$. It is clear that x_λ^* must take the form of $x_\lambda^* = (\overline{x}_\lambda^*, \cdots, \overline{x}_\lambda^*)$, which gives the smallest $\|x\|_2^2$ without changing the value of $F_\lambda(1^\top x)$. Completing the squares in $\overline{F}_\lambda$ we have that

$$\overline{F}_\lambda(\overline{x}_\lambda) = d(d+1)\left[\overline{x} - \frac{\lambda\sqrt{h}}{d+1}\right]^2 + \lambda h\left[1 - \frac{\lambda d}{d+1}\right]. \tag{3.194}$$

Subsequently, $\overline{x}_\lambda^* = \frac{\lambda\sqrt{h}}{d+1}$. It is easy to verify that for $h \leqslant 1/8$, $\inf\{t \geqslant 0 : \overline{x}_\lambda^* + tu \in \mathcal{X} \forall u \in \mathbb{B}_d(1)\} \geqslant \|\overline{x}_\lambda^*\|_2 \geqslant \lambda\sqrt{h/(d+1)}$. Therefore, for all $\lambda \in [1/4, 3/4]$ and $1/8T^2 \leqslant h1 \leqslant 1/8$ the condition (A2) holds with $\nu = 1/16\sqrt{d+1}$.

Verification of property 2: $\|\overline{F}_\lambda - \overline{F}_\mu\|_\infty = \|F_\lambda - F_\mu\|_\infty = O(h)$. In addition, $\sup_{x \in \mathcal{X}} \|\nabla \overline{F}_\lambda(x) - \nabla \overline{F}_\mu(x)\|_2 = \|F_\lambda' - F_\mu'\|_\infty \cdot \|1\|_2 = O(\sqrt{hd})$. Omitting the dependency on d we obtain property 2.

Verification of property 3: Define $\widetilde{\mathbb{B}}_d(r) := \{x \in \mathbb{R}^d : x \geqslant 0, \|x\|_1 \leqslant r\}$. It is easy to verify that $\mathrm{vol}(\widetilde{\mathbb{B}}_d(r_1))/\mathrm{vol}(\widetilde{\mathbb{B}}_d(r_2)) = (r_1/r_2)^d$. Subsequently, for any $1 \leqslant p < \infty$ we have that

$$\|\overline{F}_0 - \overline{F}_1\|_p \leqslant \left[\frac{\mathrm{vol}(\widetilde{\mathbb{B}}_d(2\sqrt{h}))}{\mathrm{vol}(\widetilde{\mathbb{B}}_d(1))} \cdot \|\overline{F}_0 - \overline{F}_1\|_\infty^p\right]^{1/p} = O(h^{(2p+d)/2p}).$$

Verification of property 4: From previous derivations we know that $x_\lambda^* = (\overline{x}_\lambda^*, \cdots, \overline{x}_\lambda^*)$ with $\overline{x}_\lambda^* = \frac{\lambda\sqrt{h}}{d+1}$ and $\overline{F}_\lambda^* = \inf_{x \in \mathcal{X}} \overline{F}_\lambda(x) = \lambda h(1 - \frac{\lambda d}{d+1})$. Subsequently,

$$\chi(\overline{F}_\lambda, \overline{F}_{1-\lambda}) = \overline{F}_\lambda\left(\frac{1}{2}\frac{\sqrt{h}}{d+1}\right) = \frac{d}{d+1}\left|\frac{1}{2} - \lambda\right|^2 \cdot h.$$

Chapter 4

Dynamic assortment planning

Assortment planning has a wide range of applications in e-commerce and online advertising. Given a large number of substitutable products, the assortment planning problem refers to the selection of a subset of products (a.k.a., an assortment) offering to a customer such that the expected revenue is maximized (Agrawal et al., 2017a,b; Golrezaei et al., 2014; Kök et al., 2008; Rusmevichientong & Topaloglu, 2012). Given N items, each associated with a revenue parameter [1] $r_i \in [0, 1]$ representing the revenue a retailer collects once a customer purchases the i-th item. The revenue parameters $\{r_i\}_{i=1}^N$ are typically known to the retailer, who has full knowledge of each item's prices/costs.

Usually, the customer's purchasing choice i_t is governed by a probabilisitic model

$$i_t \ \sim \ p_{\theta_0}(\cdot | S_t),$$

where θ_0 is an underlying parameter characterizing the customer's *preferences* of items. Examples include independent preference parameters v_i for each $i \in [N]$, or contextual models $v_i = \exp\{x_i^\top \theta_0\}$. Traditionally in the operations management literature, the parameters of the customers' choice model are fully known (assumed to be estimated from historical data), and the assortment planning problem is merely a combinatorial optimization question. The readers are referred to (Anderson et al., 1992; Kök et al., 2008) for some excellent surveys.

In many scenarios, customers' choice behavior (e.g., mean utilities of products) may not be given as *a priori* and cannot be easily estimated well due to the insufficiency of historical data (e.g., fast fashion sale or online advertising). To address this challenge, dynamic assortment planning that simultaneously learns choice behavior and makes decisions on the assortment has received a lot of attentions (Agrawal et al., 2017a,b; Caro & Gallien, 2007; Rusmevichientong et al., 2010; Saure & Zeevi, 2013). More specifically, in a dynamic assortment planning problem, the seller offers an assortment to each arriving customer in a finite time horizon of length T. The goal of the seller is to maximize the cumulative expected revenue over T periods, or equivalently, to minimize the *regret*, which is defined as the gap between the expected revenue generated by the policy and the oracle expected revenue when the mean utility for each product is known as *a priori*.

In this chapter, we consider the dynamic assortment planning problem under variants of discrete choice models, including the plain multinomial logit model (Sec.. 4.1), the nested logit

---

[1] The constraint $r_i \leqslant 1$ is without loss of generality, because it is only a normalization of revenues.

choice model (Sec. 4.2) and a contextual choice model with linear regression modeling of contextual information (Sec. 4.3). For all works presented, algorithmic policy development and theoretical regret analysis are the primarily focused aspects.

4.1 The plain multinomial logit model

Perhaps the simplest discrete choice model is the plain (or vanilla) logit choice model, which associates with each item i an independent and unknown preference parameter $v_i > 0$. Given assortment $S \subseteq [N]$, the logit choice model posits that (define $v_0 := 0$ for notational simplicity)

$$\Pr[i|S] = \frac{v_i}{1 + \sum_{i' \in S} v_{i'}}, \qquad i \in S \cup \{0\}. \tag{4.1}$$

The logit choice model is a cornerstone model in economics decision theory (Börsch-Supan, 1990; McFadden, 1980; Williams, 1977). It is also simple that comprehensive theoretical analysis can be carried out. Under model Eq. (4.1), the expected revenue the retailer could collect by presenting an assortment $S \subseteq [N]$ to an incoming customer can be calculated as

$$R(S) = \mathbb{E}_{i \sim p(\cdot|S)}[r_i] = \frac{\sum_{i \in S} r_i v_i}{1 + \sum_{i \in S} v_i}. \tag{4.2}$$

Suppose a policy π produces a sequence of assortment selections $\{S_t\}_{t=1}^T$ over T time periods, with sequentially arriving customers. The (cumulative) *regret* of the assortment sequence $\{S_t\}_{t=1}^T$ can be subsequently defined as

$$\text{Regret}(\{S_t\}_{t=1}^T) := \sum_{t=1}^T R(S^*) - \mathbb{E}^\pi[R(S_t)] \quad \text{where} \quad S^* \in \arg\max_S R(S). \tag{4.3}$$

We also impose the condition $\max_i r_i \leqslant 1$ throughout this section, which is only for normalization purposes as the units of revenue measurement can be arbitrarily changed.

4.1.1 Popular assortments, level sets, and a potential function

For the MNL assortment selection model without capacity constraints, it is a classical result that the optimal assortment must consist of items with the largest revenue parameters (see, e.g., Kök et al. (2008)):

Proposition 14. *There exists $\theta \in [0,1]$ such that $\mathcal{L}_\theta := \{i \in [N] : r_i \geqslant \theta\}$ satisfies $R(\mathcal{L}_\theta) = R(S^*)$.*

Proposition 14 suggests that it suffices to consider "level-set" type assortments $\mathcal{L}_\theta = \{i \in [N] : r_i \geqslant \theta\}$ and finds $\theta \in [0,1]$ that gives rises to the largest $R(\mathcal{L}_\theta)$.

Intuitively, $F(\theta)$ is the expected revenue obtained by providing the assortment consisting of all items whose revenues exceed or are equal to θ. The potential function plays a central role in the development of our dynamic trisection search algorithm and item-independent regret bounds. Similar idea of studying the expected revenue of revenue-ordered items was also considered
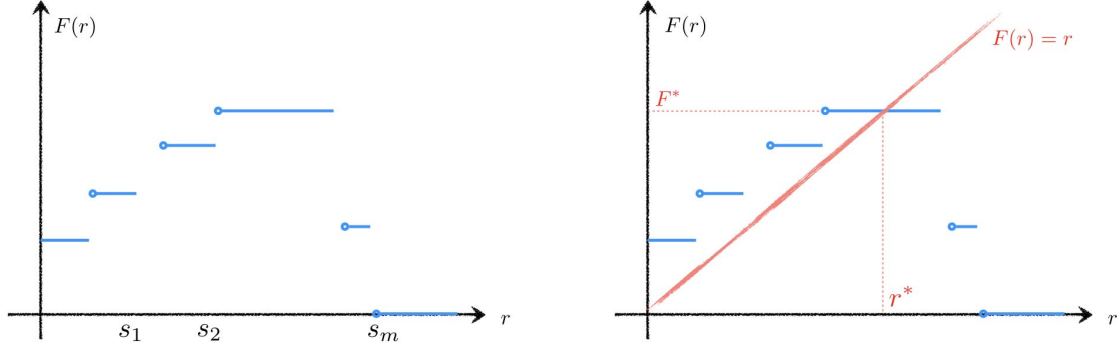
Figure 4.1: Illustration of the potential function $F(\theta)$, the important quantities F^* and θ^*, and their properties.

in Rusmevichientong & Topaloglu (2012). But we will derive a more comprehensive list of properties of the potential function F to facilitate our algorithmic development and analysis. The derived properties in this section could also be potentially useful for solving other assortment planning problems under the MNL.

Because item revenues r_i are discrete, F is a piecewise-constant function as illustrated in the left picture in Fig. 4.1, where $\mathcal{S} = \{s_1, \cdots, s_m\}$ are the changing points of F. More specifically, we have the following proposition and its verification is easy from the definition and the discretized nature of F.

Proposition 15. *There exists $c_0, \cdots, c_m \geqslant 0$ satisfying $c_i \neq c_{i+1}$ for all $i = 0, \cdots, m-1$, and $\mathcal{S} = \{s_1, \cdots, s_m\} \subseteq \{r_i\}_{i=1}^N$, such that*

$$F(\theta) = c_0 \cdot \mathbb{I}[\theta \leqslant s_1] + \sum_{i=1}^{m-1} c_i \cdot \mathbb{I}[s_i < \theta \leqslant s_{i+1}] + c_m \cdot \mathbb{I}[\theta > s_m], \qquad (4.4)$$

where $c_m = 0$.

Define $F^* := \max_{0 \leqslant i \leqslant m} c_i = \sup_{\theta \geqslant 0} F(\theta)$ as the maximum value of F. By Proposition 14, we have the following corollary saying that F^* equals the expected revenue of the optimal assortment.

Corollary 4. $F^* = R(S^*)$.

We further establish some more refined structural properties of F. For notational simplicity, let $F(x^+) := \lim_{y \to x^+} F(y)$ and $F(x^-) := \lim_{y \to x^-} F(y)$.

Lemma 52. *There exists $\theta^* > 0$ such that $\theta^* = F(\theta^*) = F^*$.*

Lemma 53. *For any $\theta \geqslant \theta^*$, $F(\theta) \leqslant \theta$ and $F(\theta) \geqslant F(\theta^+)$.*

Lemma 54. *For any $\theta \leqslant \theta^*$, $F(\theta) \geqslant \theta$ and $F(\theta) \leqslant F(\theta^+)$.*

The proofs of the above lemmas are given later. Lemmas 52, 53 and 54 provide a complete picture of the structure of the potential function F, and most importantly the relationship between F and the central straight line $F(\theta) = \theta$, as depicted in the right picture of Fig. 4.1. In particular, F intersects with the $y = x$ line at θ^* that attains the maximum function value F^*,

109

and monotonically decreases as one moves away from θ^*, meaning that F is *uni-modal*. Furthermore, Lemmas 53 and 54 show that (1) F is left-continuous; (2) F^* lies below the $y = x$ line to the right of θ^* and above the $y = x$ line to the left of θ^*. This helps us judge the positioning of a particular revenue level θ by simply comparing the expected revenue of $R(\mathcal{L}_\theta)$ with θ itself, motivating an asymmetric trisection algorithm which we describe in the next section.

4.1.2   The trisection algorithm and its regret

We propose an algorithm based on trisections of the potential function F in order to locate level θ^* at which the maximum expected revenue $F^* = F(\theta^*)$ is attained. Our algorithm avoids explicitly estimating individual items' mean utilities $\{v_i\}_{i=1}^N$, and subsequently yields a regret independent of the number of items N.

To assist with readability, below we list notations used in the algorithm description together with their meanings:

- a_τ and b_τ: left and right boundaries that contain θ^*; it is guaranteed that $a_\tau \leqslant \theta^* \leqslant b_\tau$ with high probability, and the regret incurred on failure events is strictly controlled;

- x_τ and y_τ: trisection points; x_τ is closer to a_τ and y_τ is closer to b_τ;

- $\ell_t(y_\tau)$ and $u_t(y_\tau)$: lower and upper confidence bands for $F(y_\tau)$ established at iteration t; it is guaranteed that $\ell_t(y_\tau) \leqslant F(y_\tau) \leqslant u_t(y_\tau)$ with high probability, and the regret incurred on failure events is strictly controlled;

- $\rho_t(y_\tau)$: accumulated reward by exploring level set $\mathcal{L}_{y_\tau}$ up to iteration t.

With these notations in place, we provide a detailed description of Algorithm 8 to facilitate the understanding. The algorithm operates in epochs (outer iterations) $\tau = 1, 2, \cdots$ until a total of T assortment selections are made. The objective of each outer iteration τ is to find the relative position between trisection points (x_τ, y_τ) and the "reference" location θ^*, after which the algorithm either moves a_τ to x_τ or b_τ to y_τ, effectively shrinking the length of the interval $[a_\tau, b_\tau]$ that contains θ^* to its two thirds. Furthermore, to avoid a large cumulative regret, level set corresponding to the left endpoint a_τ is exploited in each time period within the epoch τ to offset potentially large regret incurred by exploring y_τ.

In Steps 9 and 10 of Algorithm 8, lower and upper confidence bands $[\ell_t(y_\tau), u_t(y_\tau)]$ for $F(y_\tau)$ are constructed using concentration inequalities (e.g. Hoeffding's inequality (Hoeffding, 1963)). These confidence bands are updated until the relationship between y_τ and $F(y_\tau)$ is clear, or a pre-specified number of inner iterations for outer iteration τ has been reached (set to $n_\tau := \lceil 16(y_\tau - x_\tau)^{-2} \ln(T^2) \rceil$ in Step 8). Algorithm 9 gives detailed descriptions on how such confidence intervals are built, based on repeated exploration of level set $\mathcal{L}_{y_\tau}$.

After sufficiently many explorations of $\mathcal{L}_{y_\tau}$, a decision is made on whether to advance the left boundary (i.e., $a_{\tau+1} \leftarrow x_\tau$) or the right boundary (i.e., $b_{\tau+1} \leftarrow y_\tau$). Below we give high-level intuitions on how such decisions are made, with rigorous justifications presented later as part of the proof of the main regret theorem for Algorithm 8.

1. If there is sufficient evidence that $F(y_\tau) < y_\tau$ (e.g., $u_t(y_\tau) < y_\tau$), then y_τ must be to the right of θ^* (i.e., $y_\tau \geqslant \theta^*$) due to Lemma 53. Therefore, we will shrink the value of right boundary by setting $b_{\tau+1} \leftarrow y_\tau$.

Algorithm 8 The trisection algorithm.

1: **Input**: revenue parameters $r_1, \cdots, r_n \in [0, 1]$, time horizon T
2: **Output**: sequence of assortment selections $S_1, S_2, \cdots, S_T \subseteq [N]$
3: Initialization: $a_0 = 0$, $b_0 = 1$;
4: **for** $\tau = 0, 1, \cdots$ **do**
5:      $x_\tau = \frac{2}{3} a_\tau + \frac{1}{3} b_\tau$, $y_\tau = \frac{1}{3} a_\tau + \frac{2}{3} b_\tau$;   ▷ trisection
6:      $\ell_0(x_\tau) = \ell_0(y_\tau) = 0$, $u_0(x_\tau) = u_0(y_\tau) = 1$;   ▷ initialization of confidence intervals
7:      $\rho_0(x_\tau) = \rho_0(y_\tau) = 0$;   ▷ initialization of accumulated rewards
8:      **for** $t = 1$ to $16 \lceil (y_\tau - x_\tau)^{-2} \ln(T)) \rceil$ [3] **do**
9:          **if** $\ell_{t-1}(y_\tau) \leqslant y_\tau \leqslant u_{t-1}(y_\tau)$ **then** $\rho_t(y_\tau), \ell_t(y_\tau), u_t(y_\tau) \leftarrow$ EXPLORE$(y_\tau, t, 1/T^2)$;
                                                   ▷ Explore the right midpoint y_τ
10:          **else** $\rho_t(y_\tau), \ell_t(y_\tau), u_t(y_\tau) \leftarrow \rho_{t-1}(y_\tau), \ell_{t-1}(y_\tau), u_{t-1}(y_\tau)$
11:          **end if**
12:      **end for**
13:      Exploit the left endpoint a_τ: pick assortment $S = \mathcal{L}_{a_\tau}$;
14: **end for**
                                                      ▷ Update trisection parameters
15: **if** $u_t(y_\tau) < y_\tau$ **then** $a_{\tau+1} = a_\tau$, $b_{\tau+1} = y_\tau$;
16: **else** $a_{\tau+1} = x_\tau$, $b_{\tau+1} = b_\tau$.
17: **end if**

Algorithm 9 EXPLORE Subroutine: exploring a certain revenue level θ

1: **Input**: revenue level θ, time t, confidence level δ
2: **Output**: accumulated revenue $\rho_t(\theta)$, confidence intervals $\ell_t(\theta)$ and $u_t(\theta)$
3: Pick assortment $S = \mathcal{L}_\theta(\mathcal{N})$ and observe purchasing action $j \in S \cup \{0\}$;
4: Update accumulated reward: $\rho_t(\theta) = \rho_{t-1}(\theta) + r_j$ ▷ $r_0 := 0$
5: Update confidence intervals: $[\ell_t(\theta), u_t(\theta)] = \frac{\rho_t(\theta)}{t} \pm \sqrt{\frac{\log(1/\delta)}{2t}}$.
---

2. On the other hand, when $u_t(y_\tau) \geqslant y_\tau$, we can conclude that x_τ *must be to the left of θ^* (i.e.,*
$x_\tau \leqslant \theta^*$). We show this by contradiction. Assuming that $x_\tau > \theta^*$, since y_τ is always greater
than x_τ (and thus $y_\tau > \theta^*$) and the gap between y_τ and $F(y_\tau)$ is at least $y_\tau - x_\tau$ [2], the gap will
be detected by the confidence bands and thus we will have $u_t(y_\tau) < y_\tau$ with high probability.
This leads to a contradiction.

Therefore, since x_τ is to the left of θ^*, we should increase the value of the left boundary by
setting $a_{\tau+1} \leftarrow x_\tau$.

    The following theorem is our main upper bound result for the (worst-case) regret incurred by
Algorithm 8.

Theorem 13. *There exists a universal constant $C_1 > 0$ such that for all parameters $\{v_i\}_{i=1}^N$ and*

---

[2]By Lemma 53, we have $y_\tau - F(y_\tau) \geqslant y_\tau - F(x_\tau) \geqslant y_\tau - x_\tau$
[3]Stop whenever the maximum number of iterations T is reached.

$\{r_i\}_{i=1}^N$ *satisfying* $r_i \in [0,1]$*, the regret incurred by Algorithm 8 satisfies*

$$\text{Regret}(\{S_t\}_{t=1}^T) = \mathbb{E} \sum_{t=1}^T R(S^*) - R(S_t) \leqslant C_1 \sqrt{T \log T}. \tag{4.5}$$

4.1.3 Improved regret with adaptive confidence intervals

In this section we consider a variant of Algorithm 8 that achieves an improved regret of $O(\sqrt{T})$. The key idea is to use an adaptive allocation of confidence levels, by allowing larger failure probability as more data are collected. This is because later failures result in smaller accumulated regret. Such a strategy is motivated by the MOSS algorithm (Audibert & Bubeck, 2009) for multi-armed bandits. However, our analysis is quite different from (Audibert & Bubeck, 2009), involving new concentration inequalities and induction arguments tailored specifically to our model and proposed policy.

We start with a new uniform concentration inequality for adaptively chosen confidence levels.

Lemma 55. *Let* $X_1, \cdots, X_L$ *be i.i.d. random variables with mean* μ *and satisfy* $a \leqslant X_i \leqslant b$ *almost surely for all* $\ell \in [L]$*. For any* $\delta \in (0,1]$*, it holds that*

$$\Pr\left[\forall \ell \in [L], \left| \frac{1}{\ell} \sum_{i=1}^\ell X_i - \mu \right| \leqslant \sqrt{\frac{2(b-a)^2 \ln(8/(\delta\ell))}{\ell}} \right] \geqslant 1 - L\delta. \tag{4.6}$$

The proof of Lemma 55 is deferred later, based on a careful doubling argument with Hoeffding's maximal inequality (Hoeffding (1963), re-phrased in Lemma 93). Compared to the classical Hoeffding's inequality (Lemma 89) with the union bound, one notable difference is the increasing "failure probability" as ℓ increases (effectively $\ell\delta$ in $\sqrt{\frac{2\ln(8/(\delta\ell))(b-a)^2}{\ell}}$ instead of δ). This allows the confidence intervals to be much shorter for large ℓ.

With Lemma 55, we are ready to describe the variant of Algorithm 8, which attains the tight regret bound. Most steps in Algorithms 8 and 9 remain unchanged, and the changes are summarized below:

- Step 5 in Algorithm 9 is replaced with

$$[\ell_t(\theta), u_t(\theta)] = \frac{\rho_t(\theta)}{t} \pm \sqrt{\frac{2\ln[8/(\delta t)]}{t}}. \tag{4.7}$$

- Step 9 in Algorithm 8 is replaced with $\text{EXPLORE}(y_\tau, t, 1/T)$; correspondingly, the number of inner iterations is changed to $n_\tau = 8[(y_\tau - x_\tau)^{-2} \ln(8T(y_\tau - x_\tau)^2)]$.

The first change for improving the regret is the way how confidence intervals $[\ell_t(\theta), u_t(\theta)]$ of $F(\theta)$ is constructed. Instead of using fixed confidence level $1/T^2$ as in the baseline policy, in the revised policy *varying* confidence levels are employed, with "effective" failure probabilities increase as the algorithm collects more data.

We also remark that similar confidence parameter choices were also adopted in (Audibert & Bubeck, 2009) to remove additional $\log(T)$ factors in multi-armed bandit problems.

The following theorem shows that the algorithm variant presented above achieves an asymptotic regret of $O(\sqrt{T})$, considerably improving Theorem 13 with an $O(\sqrt{T \log T})$ regret bound.

Theorem 14. *There exists a universal constant $C_2 > 0$ such that for all parameters $\{v_i\}_{i=1}^N$ and $\{r_i\}_{i=1}^N$ satisfying $r_i \in [0, 1]$, the regret incurred by the variant of Algorithm 8 described above satisfies*

$$\text{Regret}(\{S_t\}_{t=1}^T) = \mathbb{E} \sum_{t=1}^T R(S^*) - R(S_t) \leqslant C_2\sqrt{T}. \tag{4.8}$$

4.1.4 Lower bound: the uncapacitated setting

In this section we prove a matching lower bound of the worst-case regret attainable by *any* policy π, under the "uncapacitated setting" in which there is no capacity constraint imposed on the provided assortments $\{S_t\}$.

Theorem 15. *Let N and T be the number of items and the time horizon that can be arbitrary. There exists revenue parameters $r_1, \cdots, r_N \in [0, 1]$ such that for any policy π,*

$$\sup_{v_1, \cdots, v_N \geqslant 0} \text{Regret}(\{S_t\}_{t=1}^T) \geqslant \frac{\sqrt{T}}{384}. \tag{4.9}$$

Theorem 15 shows that our regret upper bounds in Theorems 13 and 14 are tight up to an $O(\sqrt{\log T})$ term and/or numerical constants.

4.1.5 Lower bound: the capacitated setting

In the *capacitated setting*, capacity constraints are imposed on the supplied assortments. More specifically, the provided assortments $\{S_t\}_{t=1}^T$ must satisfy $|S_t| \leqslant K$ for all t, for some pre-specified capacity limit $K \leqslant N$. The uncapacitated setting would then be the special case of $K = N$.

In the case of $K < N$, the trisection algorithms we considered in the previous section will no longer be valid, as the key popular set structure (displayed in Proposition 14) is violated when $K < N$. The works of Agrawal et al. (2017a,b) considered alternative UCB or Thompson sampling based approaches, and established regret upper bounds on the order of $\widetilde{O}(\sqrt{NT})$, which incurs an extra $O(\sqrt{N})$ term compared to Theorems 13 and 14.

In this section, we shall prove the following result, showing that such worsened regret upper bounds cannot be improved when the capacity constraint parameter K is much smaller than the total number of products N.

Theorem 16. *Suppose $K \leqslant N/4$. There exists an absolute constant $C \geqslant 10^{-3}$ independent of N, T and K such that for all policy π,*

$$\sup_{v_1, \cdots, v_N \geqslant 0} \text{Regret}(\{S_t\}_{t=1}^T) \geqslant C \cdot \min\{\sqrt{NT}, T\}. \tag{4.10}$$

Remark 26. When the revenue parameters $\{r_i\}_{i=1}^N$ are uniformly bounded (i.e., $r_i \leqslant 1$ for all i), a trivial policy that outputs an arbitrary fixed assortment attains regret $O(T)$, meaning that the $\Omega(\sqrt{NT})$ regret cannot be optimal when $T \ll N$. In the more common scenario of $T = \Omega(N)$, the $\sqrt{NT}$ term in Eq. (4.10) dominates, leading to an $\Omega(\sqrt{NT})$ regret lower bound.

Remark 27. In the work of Agrawal et al. (2017b), a regret lower bound of $\Omega(\sqrt{NT/K})$ is established, which matches Theorem 16 when K is a small constant but deteriorates to $\Omega(\sqrt{T})$ when K is smaller but on the same order of N. In contrast, our result in Theorem 16 establishes an $\Omega(\sqrt{NT})$ even if K is as large as $N/4$.

We also remark that the "capacity constraint" $K \leqslant N/4$ in Theorem 15 is essential. In the case of $K = N$, Theorems 13, 14, as well as previous works (Rusmevichientong et al., 2010), establish regret upper bounds that depend logarithmically or even independent of N. In the case of $N/4 < K < N$, we *conjecture* that the lower bound in Theorem 16 remains valid provided that $K/N \to \gamma$ for some constant $\gamma < 1/2$, by selecting constants in Eq. (4.85) more carefully. It is, however, unclear to us how the regret will behave for $\gamma \geqslant 1/2$ and we leave it as an interesting technical open problem. We remark that for capacitated problems the $K \leqslant N/4$ condition is very weak and could be easily satisfied in practice, because at each time an incoming customer can only be offered an assortment with much fewer items (as compared to the entire commodity pool).

Finally, there is still a gap of $O(\log T)$ between our Theorem 16 and the regret upper bounds established in the work of Agrawal et al. (2017a). We leave this as another interesting open question.

4.1.6 Numerical results

We present numerical results of our proposed trisection (and its improved variant) algorithm and compare their performance with several competitors on synthetic data.

Experimental setup. We generate each of the revenue parameters $\{r_i\}_{i=1}^N$ independently and identically from the uniform distribution on $[.4, .5]$. For the preference parameters $\{v_i\}_{i=1}^N$, they are generated independently and identically from the uniform distribution on $[10/N, 20/N]$, where N is the total number of items available.

To motivate our parameter setting, consider the following three types of assortments: the "single assortment" $S = \{i\}$ for some $i \in \mathcal{N}$, the "full assortment" $S = \{1, 2, \cdots, N\}$, and the "appropriate" assortment $S = \{i \in \mathcal{N} : r_i \geqslant 0.42\}$. For the single assortment $S = \{i\}$, because the preference parameter for each item is rather small ($v_i \leqslant 20/N$), no single assortment can produce an expected revenue exceeding $0.5 \times (20/N)/(1 + 20/N) = 10/(20 + N)$. For the full assortment $S = \{1, 2, \cdots, N\}$, because $\sum_{i=1}^N r_i v_i \xrightarrow{p} 0.45 \times 15/N \times N = 6.75$ and $\sum_{i=1}^N v_i \xrightarrow{p} 15$ by the law of large numbers, the expected revenue of S is around $6.75/(1 + 15) = 0.422$. Finally, for the "appropriate" assortment $S = \{i \in \mathcal{N} : r_i \geqslant 0.42\}$, we have $\sum_{i \in S} r_i v_i \xrightarrow{p} 0.46 \times 15/N \times 0.8N = 5.52$ and $\sum_{i \in S} v_i \xrightarrow{p} 15/N \times 0.8N = 12$. Therefore, the expected revenue of S is around $5.52/(1 + 12) = 0.425 > 0.422$. The above discussion shows that a revenue threshold $r^* \in (0.4, 0.5)$ is mandatory to extract a portion of the items $\{i \in \mathcal{N} : r_i \geqslant r^*\}$ that attain the optimal expected revenue, which is highly non-trivial for a dynamic assortment selection algorithm to identify.

Comparative methods. Our trisection algorithm with $O(\sqrt{T \log T})$ regret is denoted as TRISEC, and its improved adaptive variant (with regret $O(\sqrt{T})$) is denoted as ADAP-TRISEC. The other

Table 4.1: Average (mean) and worst-case (max) regret of our trisection (TRISEC.) and adaptive trisection (ADAP-TRISEC.) algorithms and their competitors on synthetic data. N is the number of items and T is the time horizon.

(N, T)	UCB		THOMPSON		GRS		TRISEC.		ADAP-TRISEC.	
	mean	max	mean	max	mean	max	mean	max	mean	max
(100,500)	34.9	38.1	1.28	2.97	10.9	22.4	7.68	7.68	1.99	1.99
(250,500)	54.3	56.2	2.81	4.95	7.93	34.2	7.57	7.57	2.23	2.23
(500,500)	73.4	75.5	4.90	4.95	7.02	43.4	7.43	7.43	2.23	2.23
(1000,500)	90.3	93.5	8.17	10.7	5.34	45.1	7.44	7.44	2.25	2.25
(100,1000)	73.1	78.2	1.36	2.79	139.9	175.0	8.69	8.69	3.90	3.90
(250,1000)	113.7	119.3	3.36	5.17	90.1	110.1	8.69	8.69	4.13	4.14
(500,1000)	136.8	140.3	5.65	7.64	65.7	113.9	9.38	9.38	3.80	3.80
(1000, 1000)	160.8	165.4	9.31	12.4	8.43	22.8	9.77	9.77	3.97	3.97

methods we compare against include the *Upper Confidence Bound* algorithm of Agrawal et al. (2017a) (denoted as UCB), the *Thompson sampling* algorithm of Agrawal et al. (2017b) (denoted as THOMPSON), and the *Golden Ratio Search* algorithm of Rusmevichientong et al. (2010) (denoted as GRS). Note that both UCB and THOMPSON proposed in Agrawal et al. (2017a,b) were initially designed for the *capacitated* MNL model, in which the number of items each assortment contains is restricted to be at most $K < N$. In our experiments, we operate both the UCB and THOMPSON algorithms under the uncapacitated setting, simply by removing the constraint set when performing each assortment optimization.

Most hyper-parameters (such as constants in confidence bounds) are set directly using the theoretical values. One exception is our improved adaptive trisection algorithm (ADAP-TRISEC), in which we replace the $\sqrt{\frac{2\ln(8/(\delta\ell))}{\ell}}$ confidence interval configuration with $\sqrt{\frac{0.1\ln(8/(\delta\ell))}{\ell}}$. We observe that a smaller constant value leads to better empirical performance. Another is the GRS algorithm: in Rusmevichientong et al. (2010) the number of exploration iterations is set to $34\ln(2N)/\beta^2$ where $\beta = \min_{j \neq j'} |R(\mathcal{L}_{r_j}) - R(\mathcal{L}_{r_{j'}})|$, which is inappropriate for our "gap-free" synthetical setting in which $\beta = 0$. Instead, we use the common choice of $\sqrt{T}$ exploration iterations in typical gap-independent bandit problems for GRS.

Results. In Table 4.1 we report the mean and maximum regret from 20 independent runs of each algorithm on our synthetic data, with different settings of N (number of items) and T (time horizon length). We observe that as the number of items (N) becomes large, our algorithms (TRISEC and ADAP-TRISEC) achieve smaller mean and maximum regret compared to their competitors, and ADAP-TRISEC consistently outperforms TRISEC in all settings. Unlike UCB and THOMPSON whose regret depend polynomial on N, our TRISEC and ADAP-TRISEC algorithms have no dependency on N and hence their regret does not increase with N. Moreover, the separate exploration and exploitation structure in GRS makes its performance somewhat unstable, which leads to a larger gap between mean and maximum regrets.
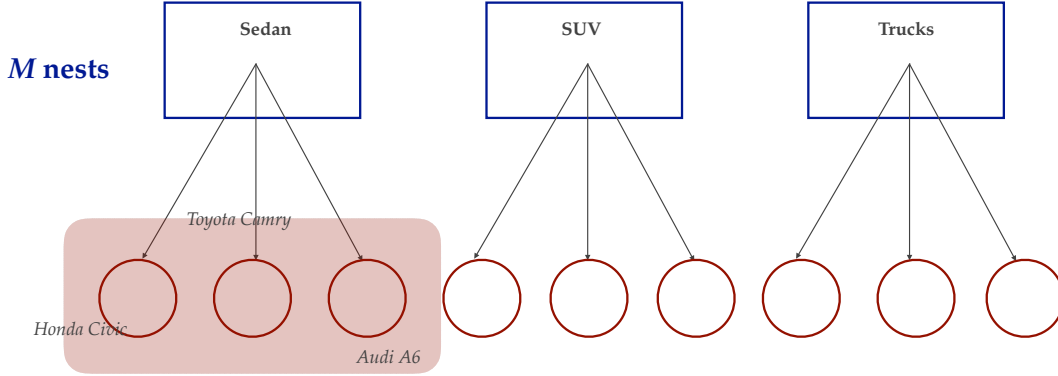
Figure 4.2: A simple illustration of the nested logit model with a car sales example. In this example, 9 cars are organized into $M = 3$ nests (Sedan, SUV and Trucks), and within each nest there are $N = 3$ car makes. An incoming customer would first choose a nest for purchase (if she makes a purchase at all), and then chooses one particular make within the nest she chooses in the first stage. The retailer has full knowledge of the nested tree structure and the profit margins for selling each item, but will need to learn the utility parameters of the nests and items.

4.2   The nested multinomial logit model

In a nested multinomial logit model (Train, 2009, Chapter 9), items are organized into *nests*, as depicted in Figure 4.2. We use $[M] = \{1, 2, \cdots, M\}$ to denote M nests. For each nest $i \in [M]$, denote the items in nest i by $[N_i] = \{1, 2, \cdots, N_i\}$. Each item $j \in [N_i]$ is associated with a *known* revenue parameter r_{ij} and an *unknown* mean utility parameter v_{ij}. Without loss of generality, we assume each nest has an equal number of items, i.e., $N_1 = \cdots = N_M = N$, because one can always add items with zero utility and revenue parameters. Let $\mathbb{S}_i = 2^{[N]}$ be the set of all possible assortments for nest i. Further, let $\{\gamma_i\}_{i\in[M]} \subseteq [0, 1]$ be a collection of *unknown* correlation parameters for different nests. Each parameter γ_i is a measure of the degree of independence among the items in nest i: a larger value of γ_i indicates less correlation (see Chapter 4 of Train (2009)).

At each time period $t \in \{1, 2, \cdots, T\}$, the retailer offers the arriving customer an assortment $S_i^{(t)} \in \mathbb{S}_i$ for every nest $i \in [M]$, conveniently denoted as $S^{(t)} = (S_1^{(t)}, \cdots, S_M^{(t)})$. The retailer then observes a nest-level purchase option $i_t \in [M] \cup \{0\}$. If $i_t \in [M]$, an item $j_t \in [N]$ is purchased within the nest i_t. On the other hand, $i_t = 0$ means no purchase occurs at time t. The

probabilistic model for the purchasing option (i_t, j_t) can be formulated below:

$$\Pr\left[i_t = i | S^{(t)}\right] = \frac{V_i(S_i^{(t)})^{\gamma_i}}{V_0 + \sum_{i'=1}^{M} V_{i'}(S_{i'}^{(t)})^{\gamma_{i'}}}, \quad \text{where} \ \ V_0 \equiv 1, \ \ V_i(S_i^{(t)}) = \sum_{j \in S_i^{(t)}} v_{ij} \ \text{ for } i \in [M];$$

(4.11)

$$\Pr\left[j_t = j | i_t = i, S^{(t)}\right] = \frac{v_{ij}}{\sum_{j' \in S_i^{(t)}} v_{ij'}} \quad \text{for} \ \ i \in [M], \ \ j \in S_i^{(t)}.$$

(4.12)

Note that when $\gamma_i = 1$ for all $i \in [M]$, the nested logit model reduces to the standard MNL model.

The retailer then collects revenue r_{i_t, j_t} provided that $i_t \neq 0$. The expected revenue $R(S^{(t)})$ given the assortment combination $S^{(t)}$ can then be written as

$$
\begin{aligned}
R(S^{(t)}) &= \sum_{i=1}^{M} \Pr\left[i_t = i | S^{(t)}\right] \sum_{j \in S_i^{(t)}} r_{ij} \Pr\left[j_t = j | i_t = i, S^{(t)}\right] \\
&= \frac{\sum_{i=1}^{M} R_i(S_i^{(t)}) V_i(S_i^{(t)})^{\gamma_i}}{1 + \sum_{i=1}^{M} V_i(S_i^{(t)})^{\gamma_i}}; \qquad \text{where} \ \ R_i(S_i^{(t)}) = \frac{\sum_{j \in S_i^{(t)}} r_{ij} v_{ij}}{\sum_{j \in S_i^{(t)}} v_{ij}}. \ \ (4.13)
\end{aligned}
$$

The objective of the seller to minimize *expected (accumulated) regret*, which is defined as follows:

$$\text{Regret}(\{S^{(t)}\}_{t=1}^{T}) := \sum_{t=1}^{T} R^* - \mathbb{E}^{\pi}\left[R(S^{(t)})\right], \quad \text{where} \quad R^* = \max_{S \in \mathbb{S} = \mathbb{S}_1 \times \cdots \times \mathbb{S}_M} R(S). \quad (4.14)$$

Throughout this section, we make the following boundedness assumptions on revenue and utility parameters:

(A1) $0 \leqslant r_{ij} \leqslant 1$ for all $i \in [M]$ and $j \in [N]$.

(A2) $0 < v_{ij} \leqslant C_V$ for all $i \in [M]$ and $j \in [N]$ with some constant $C_V \geqslant 1$.

The first boundedness assumption on revenue parameters is standard in the literature (see e.g., Theorem 1 in Agrawal et al. (2017a)). It is also worthwhile noting that assumption (A2) is different from and weaker than the common assumption that no purchase (with $V_0 = 1$) is the most frequent outcome (see e.g., Agrawal et al. (2017a,b)).

4.2.1 Assortment space reductions

For nested logit models, the complete assortment selection space (a.k.a. action space) $\mathbb{S} = \mathbb{S}_1 \times \mathbb{S}_2 \times \cdots \times \mathbb{S}_M$ is extremely large, consisting of an exponential number of candidate assortment selections (on the order of $(2^N)^M$). Existing bandit learning approaches treating each assortment set in $\mathbb{S}$ independently would easily incur a regret also exponentially large. It is thus mandatory to reduce the number of candidate assortment sets in $\mathbb{S}$.

Fortunately, existing results on the structure of optimal S show that it suffices to consider *level sets* $\mathcal{L}_i(\theta_i) := \{j \in [N] : r_{ij} \geqslant \theta_i\}$ for each nest i. In other words, $\mathcal{L}_i(\theta_i)$ is the set

of products in nest i with revenue larger than or equal to a given threshold $\theta_i \geqslant 0$. Define $\mathbb{P}_i := \{\mathcal{L}_i(\theta_i) : \theta_i \geqslant 0\} \subseteq \mathbb{S}_i$ to be all the possible level sets of $\mathbb{S}_i$ and let

$$\mathbb{P} := \mathbb{P}_1 \times \mathbb{P}_2 \times \cdots \times \mathbb{P}_M \subseteq \mathbb{S}. \tag{4.15}$$

The following lemma from Davis et al. (2014) and Li et al. (2015) shows that one can restrict the assortment selections to $\mathbb{P}$ without loss of any optimality in terms of expected revenue.

Lemma 56 (Davis et al. (2014); Li et al. (2015)). *There exists level set threshold parameters* $(\theta_1^*, \ldots, \theta_M^*)$ *and* $S^* = (\mathcal{L}_1(\theta_1^*), \cdots, \mathcal{L}_M(\theta_M^*)) \in \mathbb{P}$ *such that the following hold:*

1. $R(S^*) = \max_{S \in \mathbb{S}} R(S) = R^*$;
2. $\theta_i^* \geqslant \gamma_i R^* + (1 - \gamma_i) R_i(S_i^*)$ *for all* $i \in [M]$, *where* $S_i^* = \mathcal{L}_i(\theta_i^*)$.

The first item in Lemma 56 is an important structural result showing that the optimal assortments are "revenue-ordered" within each nest. The second item is a technical result, which will be used in the proof. Compared to the original action space $\mathbb{S}$, the reduced "level set" space $\mathbb{P}$ is much smaller, with each $\mathbb{P}_i$ consisting of N instead of 2^N assortment candidates.

4.2.2 A nested singleton model

To facilitate the illustration of our idea, we introduce a "singleton" description of the original nested model which we name *nested singleton models*. In our singleton model, we treat each *level set* as a "singleton item" in the nested model with an aggregate random revenue (which corresponds to the nest-level revenue in (4.13)). The introduced "nested singleton model" not only helps simplify our algorithms' descriptions and their analysis but also highlights our main idea of "aggregated estimation" on a nested level. Moreover, this nested singleton model will provide a unified description of a more sophisticated policy based on a discretization technique, which will be introduced later.

Recall that for each nest $i \in [M]$, there will be only $(N + 1)$ distinct level sets $\{\mathcal{L}_i(\theta_i) : i \in [N]\} \cup \{\mathcal{L}_i(\infty)\}$, where $\mathcal{L}_i(\infty) = \varnothing$ corresponds to the empty assortment set. To simplify the problem, we shall consider each level set as a *singleton item*, associated with a preference parameter and a *mean* revenue parameter. More specifically, each nest $i \in [M]$ consists of $N + 1$ "singleton items", each labeled as $\theta \in \mathcal{K}_i := \{r_{ij} : j \in [N]\} \cup \{\infty\}$, where the singleton θ in nest i corresponds to the assortment level set $\mathcal{L}_i(\theta)$. It should also be noted that $\theta = \infty$ corresponds to the empty assortment. With this notation, each assortment combination $S = (S_1, \cdots, S_M) \in \mathbb{P}$ can be equivalently written as

$$\theta = (\theta_1, \cdots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M. \tag{4.16}$$

Here $\theta_i \in \mathcal{K}_i$ corresponds to level set $S_i = \mathcal{L}_i(\theta_i) = \{j \in [N] : r_{ij} \geqslant \theta_i\}$ being offered in nest i. When necessary, we will also write $S_i(\theta_i)$ to emphasize that the assortment in nest i depends on the singleton θ_i.

After presenting the customer with assortment combination $S \in \mathbb{P}$ (or equivalently parameterized by θ), the retailer observes $i_t \in [M] \cup \{0\}$ indicating which nest is chosen ($i_t = 0$ means no purchase is made at time t) and collects revenue $r_t \in [0, 1]$, which is a random variable

corresponding to the revenue r_{i_t,j_t} of item j_t in nest i_t (if $i_t = 0$, then $r_t = 0$ almost surely). In the nested singleton model, we discard the item choice j_t within nest i_t, and only record the nest-level selection i_t and revenue collected r_t.

For any $i \in [M]$ and $\theta_i \in \mathcal{K}_i$, we define $u_{i,\theta_i} := V_i(S_i(\theta_i))^{\gamma_i}$ and $\phi_{i,\theta} := R_i(S_i(\theta_i))$, where $V_i(S_i(\theta_i))$ and $R_i(S_i(\theta_i))$ are nest-level utility parameter and expected revenue associated with the level set $S_i(\theta_i) = \mathcal{L}_i(\theta_i)$ (see definitions of V_i and R_i in Eq (4.11) and (4.13), respectively). Given an assortment combination $\theta^{(t)} = (\theta_1, \cdots, \theta_M)$ at time epoch t, it is easy to verify that the random choice $i_t \in [M] \cup \{0\}$, and corresponding random revenue $r_t \in [0, 1]$ satisfy the following:

$$\Pr[i_t = i|\theta^{(t)}] = \frac{u_{i,\theta_i}}{1 + \sum_{i'=1}^{M} u_{i',\theta_{i'}}}; \qquad \mathbb{E}[r_t|i_t = i] = \phi_{i,\theta_i}; \;\; r_t = 0 \;\; a.s. \;\; \text{if } i_t = 0. \qquad (4.17)$$

In fact, Eq. (4.17) resembles the classical *plain* MNL model, with two important differences. First, the revenues collected on each purchased singleton i_t are *random* instead of fixed, and the mean revenue parameters $\{\phi_{i,\theta}\}$ are unknown and have to be estimated from random revenues collected from purchasing events. Second, it is constrained in that at most one singleton $\theta_i \in \mathcal{K}_i$ can be offered within each nest $i \in [M]$, where in the classical plain MNL model with capacity constraints, any M items can be combined as an assortment.

Define the expected revenue for an assortment combination $\theta^{(t)}$ as,

$$R'(\theta^{(t)}) := \sum_{i=1}^{M} \Pr[i_t = i|\theta^{(t)}] \cdot \mathbb{E}[r_t|i_t = i] = \frac{\sum_{i=1}^{M} \phi_{i,\theta_i} u_{i,\theta_i}}{1 + \sum_{i=1}^{M} u_{i,\theta_i}}.$$

The objective is to minimize the regret:

$$\text{Regret}(\{\theta^{(t)}\}_{t=1}^{T}) := \mathbb{E} \sum_{t=1}^{T} R'(\theta^*) - R'(\theta^{(t)}) \qquad \text{where } R'(\theta^*) = \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} R'(\theta). \qquad (4.18)$$

By our assumptions (A1) and (A2), it is easy to verify that $\phi_{i,\theta_i} \in [0, 1]$ and $u_{i,\theta_i} \in [0, (NC_V)^{\gamma_i}] \subseteq [0, NC_V]$ for all $i \in [M]$ and $\theta_i \in \mathcal{K}_i$. We also note that $(NC_V)^{\gamma_i} \leqslant NC_V$, since $\gamma_i \in [0, 1]$ and $C_V \geqslant 1$.

For a given dynamic assortment selection policy π' under this nested singleton model, it is easy to construct a policy π under the original nested logit model simply by converting $\{\theta^{(t)}\}$ to their corresponding assortment combinations $\{S^{(t)}\}$. Please see Algorithm 10 and the following Proposition 16 for more details.

Proposition 16. *Suppose there exists a policy π' that attains a regret of at most Δ on any instance of the nested singleton model with $|\mathcal{K}_i| = K = N + 1$ for all $i \in [M]$ and $u_{i,\theta_i} \in [0, U]$ with $U \leqslant NC_V$ for all $i \in [M]$ and $\theta_i \in \mathcal{K}_i$. Then there is a meta-policy π (see Algorithm 10) that produces an assortment combination sequence $\{S^{(t)}\}_{t=1}^{T}$ under the original nested choice model with regret (defined in Eq. (4.14)) at most Δ.*

Proposition 16 is a simple consequence of Lemma 56 and we omit its proof.

Algorithm 10 The meta-policy π built upon policy π' under the nested singleton model.

1: **Input**: a dynamic assortment planning policy π' under the nested singleton model.
2: **Output**: a dynamic assortment planning policy π for the original nested logit model.
3: **for** $t = 1, 2, \cdots, T$ **do**
4:     Let $\theta^{(t)} = (\theta_1^{(t)}, \cdots, \theta_M^{(t)}) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ be the output of π';
5:     Produce assortment combination $S^{(t)} = (S_1^{(t)}, \cdots, S_M^{(t)})$, where $S_i^{(t)} = \mathcal{L}_i(\theta_i^{(t)})$;
6:     Observe the nest-selection $i_t \in [M] \cup \{0\}$ and revenue r_t, and pass i_t, r_t to policy π';
7: **end for**

4.2.3 UCB-based dynamic assortment planning policies

We design dynamic planning policies under the nested logit model using an *upper-confidence-bound (UCB)* approach. We focus entirely on the simpler assortment model specified in Eq. (4.17), since it is (approximately) equivalent to the original nested logit model as shown in Proposition 16.

Our main policy is based on the idea of UCB from classical bandit algorithms (Bubeck & Cesa-Bianchi, 2012) and repeated exploration of the same action until no-purchase happens, which was found to be very useful for assortment planning problems (Agrawal et al., 2017a) because it provides *unbiased* estimates of model parameters.

The pseudo-code of our proposed policy is given in Algorithm 11. We first explain a few notations used in the algorithm and then describe the details of the algorithm.

- $\mathcal{E}_\tau$: all iterations in epoch τ where the same assortment combination θ is provided. Each epoch (corresponding to Steps 7-9 in Algorithm 11) terminates whenever the no-purchase action is observed. In other words, one and only one "no-purchase" action $i_t = 0$ appears at the last iteration of each epoch $\mathcal{E}_\tau$.

- $\mathcal{T}(i, \theta)$: the indices of epochs in which $\theta \in \mathcal{K}_i$ is supplied in nest i; $T(i, \theta) = |\mathcal{T}(i, \theta)|$ denotes the cardinality of $\mathcal{T}(i, \theta)$;

- $\widehat{n}_{i,\tau}$: the number of iterations in the epoch τ (i.e., $\mathcal{E}_\tau$) in which an item in nest i is purchased;

- $\widehat{r}_{i,\tau}$: the total revenue collected for all iterations in $\mathcal{E}_\tau$ in which an item in nest i is purchased;

- $\widehat{u}_{i,\theta}, \widehat{\phi}_{i,\theta}, \overline{u}_{i,\theta}, \overline{\phi}_{i,\theta}$: estimates of $u_{i,\theta}, \phi_{i,\theta}$, and their upper confidence bands.

The high-level idea of Algorithm 11 can be described as follows. The algorithm operates in "epochs" $\mathcal{E}_1, \mathcal{E}_2, \cdots$. For all iterations in each epoch $\mathcal{E}_\tau$, the *same* assortment combination $\theta = (\theta_1, \cdots, \theta_M)$ are offered and customers' purchasing actions are observed. The θ offered in each epoch $\mathcal{E}_\tau$ is computed by maximizing upper confidence bands of expected regrets over all assortments. An epoch terminates whenever a "no-purchase" action is made by the arriving customer. This epoch-based strategy (i.e., offering the same assortment until no-purchase is observed) was first introduced by Agrawal et al. (2017a) and enjoys the favorable properties stated in the next lemma.

Lemma 57. *For each epoch $\mathcal{E}_\tau$ and nest $i \in [M]$, let $\widehat{\theta}_i \in \mathcal{K}_i$ be the singleton provided in nest i. The expectations of the number of iterations and total revenues collected in which nest*

Algorithm 11 The upper confidence bound (UCB) policy for dynamic assortment planning under the nested singleton model in Eq. (4.17).

1: **Input**: singleton sets $\mathcal{K}_1, \cdots, \mathcal{K}_M$, upper bound U on $\{u_{i,\theta}\}$, time horizon T.
2: **Output**: assortment sequences $\theta^{(1)}, \cdots, \theta^{(T)} \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$.
3: Initialization: $\tau = 1$, $\{\mathcal{E}_\tau\}_{\tau=1}^\infty = \varnothing$, $t = 1$; for every $i \in [M]$ and $\theta \in \mathcal{K}_i$, set $\mathcal{T}(i,\theta) = \varnothing$, $T(i,\theta) = 0$, $\widehat{\phi}_{i,\theta} = \overline{\phi}_{i,\theta} = 1$, $\widehat{u}_{i,\theta} = \overline{u}_{i,\theta} = U$; for all $i \in [M]$ and $\theta \in \mathcal{K}_i$ corresponding to the empty assortment (i.e., $\mathcal{L}_i(\theta) = \varnothing$), set $\overline{\phi}_{i,\theta} = \phi_{i,\theta} = \overline{u}_{i,\theta} = u_{i,\theta} = 0$;
4: **while** $t \leqslant T$ **do**
5: $\quad\widehat{\theta}^{(\tau)} = \widehat{\theta} \leftarrow \arg\max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} \overline{R}'(\theta)$, where $\overline{R}'(\theta) = [\sum_{i=1}^M \overline{\phi}_{i,\theta_i} \overline{u}_{i,\theta_i}] / [1 + \sum_{i=1}^M \overline{u}_{i,\theta_i}]$
6: $\qquad\qquad\qquad\qquad \triangleright$ This optimization problem can be solved in polynomial time; see Sec. 4.2.3;
7: $\quad$ **repeat**
8: $\qquad$ Pick $\theta^{(t)} = \widehat{\theta}$ and observe i_t, r_t in Eq. (4.17) and update $\mathcal{E}_\tau \leftarrow \mathcal{E}_\tau \cup \{t\}$, $t = t + 1$;
9: $\quad$ **until** $i_{t-1} = 0$ or $t > T$
10: $\quad$ **for** each $i \in [M]$ **do**
11: $\qquad$ Compute $\widehat{n}_{i,\tau} = \sum_{t' \in \mathcal{E}_\tau} \mathbb{I}[i_{t'} = i]$ and $\widehat{r}_{i,\tau} = \sum_{t' \in \mathcal{E}_\tau} r_{t'} \mathbb{I}[i_{t'} = i]$;
12: $\qquad$ Let $\theta = \widehat{\theta}_i$ and update: $\mathcal{T}(i,\theta) \leftarrow \mathcal{T}(i,\theta) \cup \{\tau\}$, $\quad T(i,\theta) \leftarrow T(i,\theta) + 1$;
13: $\qquad$ Update the utility and mean revenue estimates and their associated confidence bands:

$$\widehat{u}_{i,\theta} = \frac{1}{T(i,\theta)} \sum_{\tau' \in \mathcal{T}(i,\theta)} \widehat{n}_{i,\tau'}, \quad \widehat{\phi}_{i,\theta} = \frac{\sum_{\tau' \in \mathcal{T}(i,\theta)} \widehat{r}_{i,\tau'}}{\sum_{\tau' \in \mathcal{T}(i,\theta)} \widehat{n}_{i,\tau'}};$$

14: $\qquad$ **if** $T(i,\theta) \geqslant 96 \ln(2MTK)$ **then**
15: $\qquad\qquad \overline{u}_{i,\theta} = \min\{U, \widehat{u}_{i,\theta} + \sqrt{\frac{96 \max(\widehat{u}_{i,\theta}, \widehat{u}_{i,\theta}^2) \ln(2MTK)}{T(i,\theta)}} + \frac{144 \ln(2MTK)}{T(i,\theta)}\}$,
16: $\qquad\qquad \overline{\phi}_{i,\theta} = \min\{1, \widehat{\phi}_{i,\theta} + \sqrt{\frac{\ln(2MTK)}{T(i,\theta)\widehat{u}_{i,\theta}}}\}$;
17: $\qquad$ **else**
18: $\qquad\qquad \overline{u}_{i,\theta} = U, \quad \overline{\phi}_{i,\theta} = 1$.
19: $\qquad$ **end if**
20: $\quad$ **end for**
21: $\quad \tau \leftarrow \tau + 1$;
22: **end while**

---

i is purchased (denoted by $\widehat{n}_{i,\tau}$ and $\widehat{r}_{i,\tau}$, respectively, in Algorithm 11) satisfies the following regardless of the other offered assortments $\widehat{\theta}_{i'}$ for $i' \neq i$:

1. $\mathbb{E}[\widehat{n}_{i,\tau}] = u_{i,\widehat{\theta}_i}$;
2. $\mathbb{E}[\widehat{r}_{i,\tau}|\widehat{n}_{i,\tau}] = \widehat{n}_{i,\tau} \phi_{i,\widehat{\theta}_i}$.

Proof. Simple calculations show that (see for example Corollary A.1 of Agrawal et al. (2017a))

$$\Pr[\widehat{n}_{i,\tau} = k] = \left(\frac{u_{i,\widehat{\theta}_i}}{1 + u_{i,\widehat{\theta}_i}}\right)^k \left(\frac{1}{1 + u_{i,\widehat{\theta}_i}}\right) \qquad \text{for } k = 0, 1, 2, \cdots \qquad (4.19)$$

That is, $\widehat{n}_{i,\tau}$ is a geometric random variable with parameter $1/(1 + u_{i,\widehat{\theta}_i})$. Hence, $\widehat{n}_{i,\tau}$ is an *unbiased* estimator of $u_{i,\widehat{\theta}_i}$, meaning that $\mathbb{E}\widehat{n}_{i,\tau} = u_{i,\tau}$.

The distribution and expectation of $\widehat{r}_{i,\tau}$ can be similarly derived, using the property that $\mathbb{E}[\widehat{r}_{i,\tau}|i_t = i] = \phi_{i,\widehat{\theta}_i}$. $\qquad\square$

The above properties motivate intuitive parameter estimators $\widehat{u}_{i,\theta}, \widehat{\phi}_{i,\theta}$ of $u_{i,\theta}$ and $\phi_{i,\theta}$ for $\theta = \widehat{\theta}_i$, which are taken to be the sample averages of $\widehat{n}_{i,\tau}$ and $\widehat{r}_{i,\tau}$ over all epochs $\mathcal{E}_\tau$ in which the item $\widehat{\theta}_i$ in nest i is offered. It is worth noting that in those epochs, the offered singletons in nests other than the i-th nest (i.e., the nests i' for $i' \neq i$) can be arbitrary since the distributions of $\widehat{n}_{i,\tau}$ and $\widehat{r}_{i,\tau}$ are *independent* of $\widehat{\theta}_{i'}$ for $i' \neq i$. This key independence property enables us to combine purchasing information of vastly different assortment combinations (provided that $\widehat{\theta}_i$ remains the same), which forms an important *aggregation strategy* that avoids exponentially large regret if assortment combinations are treated independently.

Efficient computation of $\widehat{\theta}$

Our policy in Algorithm 11 involves a combinatorial optimization problem over all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ (see Step 5 in Algorithm 11). A brute-force algorithm that enumerates all such θ takes $O(K^M)$ time and is computationally intractable even for small M values. In this section we introduce a computationally efficient procedure to compute $\widehat{\theta}$ by using a binary search technique. The idea behind our procedure is similar to the one (Rusmevichientong et al., 2010) introduced for dynamic assortment optimization in capacitated MNL models.

For any $\lambda \in [0, 1]$ and $\theta = (\theta_1, \cdots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ define potential function

$$\psi_\lambda(\theta) := \sum_{i=1}^{M} (\overline{\phi}_{i,\theta_i} - \lambda)\overline{u}_{i,\theta_i}. \tag{4.20}$$

Recall the definition of $\overline{R}'(\theta) = \frac{\sum_{i=1}^{M} \overline{\phi}_{i,\theta_i}\overline{u}_{i,\theta_i}}{1+\sum_{i=1}^{M} \overline{u}_{i,\theta_i}}$ in Step 5 of Algorithm 11. The following lemma characterizes the properties of $\psi_\lambda(\theta)$ and its relationship with $\overline{R}^* = \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} \overline{R}'(\theta)$:

Lemma 58. *The following holds for all $\lambda \in [0, 1]$:*

1. *If $\overline{R}^* \geq \lambda$, then there exists a $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ such that $\psi_\lambda(\theta) \geq \lambda$; furthermore if $\overline{R}^* > \lambda$, then the inequality is strict;*
2. *If $\overline{R}^* \leq \lambda$, then for all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$, $\psi_\lambda(\theta) \leq \lambda$; furthermore if $\overline{R}^* < \lambda$, then the inequalities are strict.*

Proof. Let $\theta^* = (\theta_1^*, \cdots, \theta_M^*) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ be a maximizer of $\overline{R}'$ (i.e., $\overline{R}^* = \overline{R}'(\theta^*)$). By definition, $\sum_{i=1}^{M} (\overline{\phi}_{i,\theta_i^*} - \overline{R}^*)\overline{u}_{i,\theta_i^*} = \overline{R}^*$. If $\overline{R}^* \geq \lambda$, then $\sum_{i=1}^{M} (\overline{\phi}_{i,\theta_i^*} - \lambda)\overline{u}_{i,\theta_i^*} \geq \sum_{i=1}^{M} (\overline{\phi}_{i,\theta_i^*} - \overline{R}^*)\overline{u}_{i,\theta_i^*} = \overline{R}^* \geq \lambda$. Therefore $\psi_\lambda(\theta^*) \geq \lambda$. Furthermore, if $\overline{R}^* > \lambda$ then the last inequality in the chain of inequalities is strict. The first property is thus proved.

We next prove the second property. Assume by way of contradiction that there exists $\theta = (\theta_1, \cdots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ such that $\psi_\lambda(\theta) > \lambda$, meaning that $\sum_{i=1}^{M} (\overline{\phi}_{i,\theta_i} - \lambda)\overline{u}_{i,\theta_i} > \lambda$. Rearranging terms and dividing both sides by $(1+\sum_{i=1}^{M} \overline{u}_{i,\theta_i})$ we have $\overline{R}'(\theta) = [\sum_{i=1}^{M} \overline{\phi}_{i,\theta_i}\overline{u}_{i,\theta_i}]/[1+\sum_{i=1}^{M} \overline{u}_{i,\theta_i}] > \lambda$. This contradicts the assumption that $\overline{R}^* = \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_m} R'(\theta) \leq \lambda$. To prove the second half of the second property, simply replace all occurrences of $>$ by $\geq$. $\qquad\square$

Based on Lemma 58, an efficient optimization algorithm computing the maximizer $\widehat{\theta}^{(\tau)}$ can be designed by a *binary search* over $\lambda \in [0,1]$. In particular, for each fixed value of λ, the $\theta^*(\lambda) = (\theta_1^*(\lambda), \cdots, \theta_M^*(\lambda)) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ that maximizes $\psi_\lambda(\theta)$ can be found by setting $\theta_i^*(\lambda) \in \arg\max_{\theta_i \in \mathcal{K}_i}(\overline{\phi}_{i,\theta_i} - \lambda)\overline{u}_{i,\theta_i}$. If $\psi_\lambda(\theta^*(\lambda)) > \lambda$, then $\overline{R}^* > \lambda$, because otherwise it violates the second property in Lemma 58. Similarly, if $\psi_\lambda(\theta^*(\lambda)) \leqslant \lambda$, then $\overline{R}^* \leqslant \lambda$, because otherwise it violates the second part of the first property in Lemma 58 (note that since $\theta^*(\lambda)$ is the maximizer of $\psi_\lambda(\theta)$, $\psi_\lambda(\theta^*(\lambda)) \leqslant \lambda$ implies that $\psi_\lambda(\theta) \leqslant \lambda$ for all θ). Thus, whether $\overline{R}^* > \lambda$ or $\overline{R}^* \leqslant \lambda$ can be determined by solely comparing $\psi_\lambda(\theta^*(\lambda))$ with λ.

We remark that each evaluation of $\psi_\lambda(\theta^*(\lambda))$ takes $O(MK)$ time, and the entire binary search procedure takes time $O(MK\log(1/\epsilon))$ to approximate $\overline{R}^*$ up to arbitrarily small error ϵ. This is much faster than the brute force algorithm that takes $O(K^M)$ time.

Regret analysis

Below is our main regret theorem for Algorithm 11.

Theorem 17. *The assortment sequence $\{\theta^{(t)}\}_{t=1}^T$ produced by Algorithm 11 has regret (defined in Eq. (4.18)) upper bounded as*

$$\text{Regret}(\{\theta^{(t)}\}_{t=1}^T) \lesssim \sqrt{MKT\log(MKT)} + MKU\log^2(MKT) + O(1), \quad (4.21)$$

where $K = |\mathcal{K}|$ and $U = \max_{i \in [M]} \max_{\theta \in \mathcal{K}_i} u_{i,\theta}$.

Corollary 5. *If $K = |\mathcal{K}_i| = N+1$ (for any $i \in [M]$) and the meta-policy in Algorithm 10 is used to convert Algorithm 11 into a dynamic assortment planning algorithm for the original nested model, then*

$$\begin{aligned} \text{Regret}(\{S^{(t)}\}_{t=1}^T) &\lesssim \sqrt{MNT\log(MNT)} + MN^2 C_V \log^2(MNT) + O(1) \\ &= \widetilde{O}(\sqrt{MNT} + MN^2) \end{aligned} \quad (4.22)$$

We make several remarks on the regret upper bound in Corollary 5. First, when $T > M$ and the number of items per nest N is small, the dominating term in Eq. (4.22) is $\widetilde{O}(\sqrt{MNT})$. This matches the lower bound result $\Omega(\sqrt{MT})$ in Theorem 18 when N is a constant. When the number of items per nest N is large compared to the time horizon T, the dominating term in Eq. (4.22) is $\widetilde{O}(MN^2)$. We will show later in Sec. 4.2.4 how regret can be improved by considering a "discretization" approach under such large N settings.

4.2.4 Policies with an improved N dependency

Although we are not able to provide a lower bound on the dependence of N and derive an optimal policy, we provide a class of policies based on a discretizing technique. This class of policies generalizes our first policy since the first policy simply corresponds to a special case by setting the discretization granularity to zero. In addition, by choosing an appropriate non-zero discretization granularity, we obtain another policy with an improved regret dependence on N while sacrificing the dependence on T.

Discretizing the singleton sets

In this section, we introduce a *discretization* technique to further reduce the size of the level set space $\mathbb{P}_i$ in (4.15) (or equivalently, $\mathcal{K}_i$ in the nested singleton model introduced in Sec. 4.2.2). Instead of considering level sets defined for thresholds $\theta = r_{ij}$ for all $j \in [N]$ so that $|\mathcal{K}_i| = N + 1$, we only include level sets whose thresholds are on a finite grid.

More specifically, let $\delta \in (0, 1)$ be a granularity parameter to be optimized later. Recall the definition of the level set $\mathcal{L}_i(\theta) = \{j \in [N] : r_{ij} \geqslant \theta\}$ and we only consider level set threshold parameters θ that are multiples of $1/\delta$. Let $\mathbb{N}$ be the set of non-negative integers and define

$$\widetilde{\mathcal{K}}_i^\delta := \{\theta : 0 \leqslant \theta \leqslant 1, \;\; \theta/\delta \in \mathbb{N}, \;\; \mathcal{L}_i(\theta)'s \;\; \text{are distinct}\} \cup \{\infty\}, \quad \text{for} \;\; i \in [M] \qquad (4.23)$$

where each $\theta \in \widetilde{\mathcal{K}}_i^\delta$ corresponds to a *unique* level set $\mathcal{L}_i(\theta)$. When there are multiple θ's leading to the same level set, we keep any one of these θ's in $\widetilde{\mathcal{K}}_i^\delta$ and thus the level sets induced by $\widetilde{\mathcal{K}}_i^\delta$ (i.e., $\{\mathcal{L}_i(\theta) : \theta \in \widetilde{\mathcal{K}}_i^\delta\}$) are unique. Since duplicate assortment sets are removed in $\widetilde{\mathcal{K}}_i^\delta$, we have $\widetilde{\mathcal{K}}_i^\delta \subseteq \mathcal{K}_i$ and thus $|\widetilde{\mathcal{K}}_i^\delta| \leqslant |\mathcal{K}_i| = K = N + 1$. On the other hand, we also have $|\widetilde{\mathcal{K}}_i^\delta| \leqslant \lfloor 1/\delta \rfloor + 2$ because level set thresholds in $\widetilde{\mathcal{K}}_i^\delta$ must be an integer multiple of δ. On one hand, when δ is not too small, the size of $\widetilde{\mathcal{K}}_i^\delta$ could be significantly smaller than N. On the other hand, when $\delta \to 0$, we recover the original singleton set $\mathcal{K}_i$, which gives the full level sets. We shall thus define $\widetilde{\mathcal{K}}_i^\delta := \mathcal{K}_i$ when $\delta = 0$.

The following key discretized reduction lemma shows that by restricting ourselves to $\widetilde{\mathcal{K}}_i^\delta$ instead of $\mathcal{K}_i$, the approximation error in terms of expected revenue can be upper bounded by δ, which goes to zero as we take $\delta \to 0$.

Lemma 59 (Discretized reduction lemma). *Fix an arbitrary $\delta \in (0, 1)$. Then*

$$\max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} R'(\theta) - \max_{\theta \in \widetilde{\mathcal{K}}_1^\delta \times \cdots \times \widetilde{\mathcal{K}}_M^\delta} R'(\theta) \leqslant \delta,$$

where $R'(\theta) := [\sum_{i=1}^M \phi_{i,\theta_i} u_{i,\theta_i}]/[1 + \sum_{i=1}^M u_{i,\theta_i}]$.

A discretization based meta-policy and regret analysis

We first present a meta-policy using the discretization technique in Algorithm 12, which connects a dynamic assortment planning policy under the singleton nested model to the original nested logit model. The following proposition upper bounds the regret of the proposed meta-policy, as consequences of Lemma 59 and the fact that $|\widetilde{\mathcal{K}}_i^\delta| \leqslant \min\{N, \lfloor 1/\delta \rfloor + 2\}$ for all $i \in [M]$ and $\delta \in [0, 1)$.

Proposition 17. *Suppose there exists a policy π' that attains a regret of at most Δ on any instance of the nested singleton model with $|\widetilde{\mathcal{K}}_i^\delta| \leqslant \min\{N, \lfloor 1/\delta \rfloor + 2\}$ for all $i \in [M]$ and $u_{i,\theta_i} \in [0, U]$ with $U \leqslant NC_V$ for all $i \in [M]$ and $\theta_i \in \widetilde{\mathcal{K}}_i^\delta$. Then there is a meta-policy π (see Algorithm 12) that produces an assortment combination sequence $\{S^{(t)}\}_{t=1}^T$ under the original nested choice model with regret (defined in Eq. (4.14)) at most $\Delta + \delta T$.*

We note that the extra regret δT comes from the loss of the discretization in Lemma 59.

Now for the nested singleton model, we invoke Algorithm 11 with the discretized singletons $\widetilde{\mathcal{K}}_1^\delta, \cdots, \widetilde{\mathcal{K}}_M^\delta$ as input to construct the policy π'. Then we obtain a class of policies parameterized

---
Algorithm 12 The meta-policy π built upon policy π' and the discretization argument under the nested singleton model.

1: **Input**: a dynamic assortment planning policy π' under the nested singleton model, discretization granularity δ.
2: **Output**: a dynamic assortment planning policy π for the original nested logit model.
3: Construct discretized singleton sets $\widetilde{\mathcal{K}}_1^\delta, \cdots, \widetilde{\mathcal{K}}_M^\delta$ in Eq. (4.23);
4: **for** $t = 1, 2, \cdots, T$ **do**
5:     Let $\theta^{(t)} = (\theta_1^{(t)}, \cdots, \theta_M^{(t)}) \in \widetilde{\mathcal{K}}_1^\delta \times \cdots \times \widetilde{\mathcal{K}}_M^\delta$ be the output of π';
6:     Produce assortment combination $S^{(t)} = (S_1^{(t)}, \cdots, S_M^{(t)})$, where $S_i^{(t)} = \mathcal{L}_i(\theta_i^{(t)})$;
7:     Observe the nest-selection $i_t \in [M] \cup \{0\}$ and revenue r_t, and pass i_t, r_t to policy π'.
8: **end for**
---

by δ. By replacing $K = |\mathcal{K}_i|$ in Corollary 5 with $K = |\widetilde{\mathcal{K}}_i^\delta| \leqslant \min\{N, \lfloor 1/\delta \rfloor + 2\}$, and combing it with Proposition 17, we obtain the following corollary on the regret under the original nested logit model.

Corollary 6. *If the meta-policy in Algorithm 12 with discretization granularity δ is used to convert Algorithm 11 into a dynamic assortment planning algorithm for the original nested model, then*

$$
\begin{aligned}
\mathrm{Regret}(\{S^{(t)}\}_{t=1}^T) &\lesssim \sqrt{\min\{N, \delta^{-1}\} MT \log(MNT)} \\
&\quad + \min\{N, \delta^{-1}\} MN C_V \log^2(MNT) + \delta T + O(1). \quad (4.24)
\end{aligned}
$$

For example, by choosing $\delta \asymp T^{-1/3}$, we have $\mathrm{Regret}(\{S^{(t)}\}_{t=1}^T) = \widetilde{O}(\sqrt{M}T^{2/3} + MNT^{1/3})$.

Remark 28. In cases with many items per nest (i.e., N is large), a *small* value of δ is desired to balance the terms and achieve small overall regret in Eq. (4.24). Similarly, in instances with few items per nest (i.e., N is small), a *large* value of δ is desired to achieve the best overall regret.

We first note that Corollary 6 is a more general result, which includes Corollary 5 as a special case. Indeed, when $\delta = 0$ ($\min\{N, \delta^{-1}\} = N$), then $\widetilde{\mathcal{K}}_i^\delta = \mathcal{K}_i$ and Eq. (4.24) automatically reduces to Eq. (4.22), which upper bounds the regret of our UCB policy without discretization. On the other hand, when N is large compared with T, it is beneficial to set the discretization granularity parameter δ to be a non-zero value. In particular, by setting $\delta \asymp T^{-1/3}$, we obtain a regret upper bound of $\widetilde{O}(\sqrt{M}T^{2/3} + MNT^{1/3})$, which is smaller than the regret bound in Corollary 5 whenever $N > T^{1/3}$. (Note that when $N > T^{1/3}$, the dominating term in Corollary 5 is $\widetilde{O}(MN^2)$.)

4.2.5 A regret lower bound

We establish the following lower bound on the regret of any dynamic assortment planning policy under nested logit models.

Theorem 18. *Suppose the number of nests M is divisible by 4 and $\gamma_1 = \cdots = \gamma_M = 0.5$. Assume also that (A1) and (A2) hold. Then there exists a numerical constant $C_0 > 0$ such that*
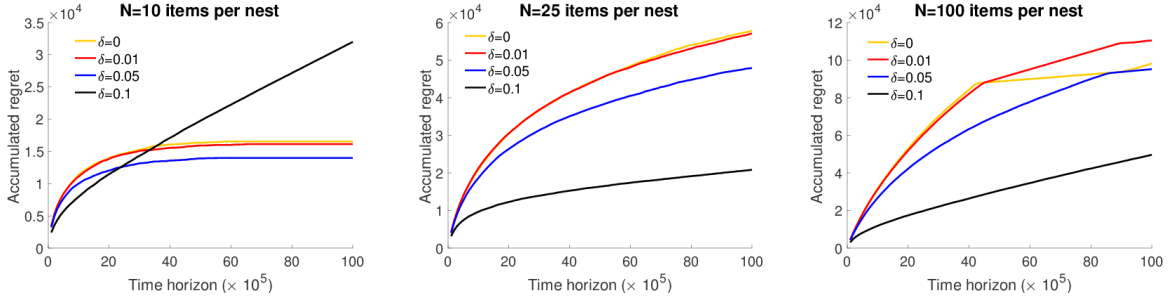
Figure 4.3: Accumulated regret of our policy with $M = 5$ nests, varying the number of items per nest N and the granularity parameter δ.

for any dynamic assortment planning policy π,

$$\sup_{\{r_{ij}, v_{ij}\}} \sum_{t=1}^{T} R^* - \mathbb{E}^{\pi}\left[R(S^{(t)})\right] \geqslant C_0\sqrt{MT} \quad \textit{where } R^* = \max_{S \in \mathbb{S}} R(S). \quad (4.25)$$

We note the condition that M is divisible by 4 is only a technical condition and does not affect the main message delivered in Theorem 18, which shows necessary dependency on M asymptotically when M is large. Our lower bound construction treats N as a constant. In particular, in our constructions of adversarial model parameters $\{r_{ij}, v_{ij}\}_{i,j=1}^{M,N}$, each nest consists of $N = 3$ items. Since N is a constant, Eq. (4.25) cannot possibly be tight in terms of dependence on N. The optimal dependence on N is a technically very challenging problem and we leave it as an open problem.

4.2.6 Numerical results

We present numerical studies of our proposed policies for dynamic nested assortment planning on synthetic data. The main focus of our simulation is the regret of our policies under various model parameter settings of M, N, and T, as well as the effect of the discretization granularity $\delta \in [0, 1]$ on the regret.

For each nest $i \in [M]$, we generate the revenue parameters $\{r_{ij}\}_{j=1}^{N}$ independently and identically from the uniform distribution on $[0.2, 0.8]$ and the preference parameters $\{v_{ij}\}_{j=1}^{N}$ independently and identically from the uniform distribution on $[10/N(M-1), 20/N(M-1)]$, where N is the number of items in each nest. The nest discounting parameters $\{\gamma_i\}_{i=1}^{M}$ are generated independently and identically from the uniform distribution on $[0.5, 1]$.

We consider the different combinations of parameters in terms of M (i.e., the number of nests), N (i.e., the number of items per nest), T (i.e., time horizon length), and δ (i.e., the granularity parameter in the discretized policy). We note that $\delta = 0$ means that no discretization is carried out, which corresponds to the policy in Algorithms 10 and 11. For each (M, N) settings, we generate model parameters $\{r_{ij}, v_{ij}, \gamma_i\}_{i,j=1}^{M,N}$ as described in the previous paragraph, and then run the dynamic assortment policy for 100 independent trials. The median and maximum accumulated regret over T periods are reported.

126

Table 4.2: Median (MED) and Maximum (MAX) accumulated regret (summation over T periods) under various model and parameter settings. The minimum regret for each case is highlighted using the bold font.

(M, N)	$\delta = 0$		$\delta = 10^{-3}$		$\delta = 5 \times 10^{-3}$		$\delta = 10^{-2}$		$\delta = 5 \times 10^{-2}$	
	MED	MAX	MED	MAX	MED	MAX	MED	MAX	MED	MAX
$T = 100$:										
(5,100)	5.5	6.4	5.5	6.0	3.8	**4.1**	**3.2**	4.3	5.4	8.5
(10,100)	4.8	6.2	5.4	5.5	4.7	6.5	**2.3**	**3.9**	5.8	7.0
(5,250)	10.4	14.1	9.8	12.0	5.7	6.5	**3.3**	**3.4**	7.0	8.3
(10,250)	10.8	12.0	9.7	12.3	5.5	7.4	**3.0**	**4.4**	5.1	8.7
(5,1000)	22.0	25.3	16.0	18.2	6.2	7.5	**3.2**	**5.0**	6.9	10.9
(10,1000)	21.5	24.1	15.1	17.7	5.1	6.4	**3.1**	**4.9**	6.2	9.4
$T = 500$:										
(5,100)	**14.3**	**18.5**	18.3	22.6	26.8	30.9	31.9	35.3	33.3	34.3
(10,100)	**15.7**	23.0	16.5	**22.1**	28.4	28.9	35.4	36.5	35.0	36.5
(5,250)	14.2	17.3	**12.7**	**14.9**	16.4	18.4	29.1	36.8	32.6	34.2
(10,250)	13.8	**15.9**	**13.0**	17.4	16.6	19.6	29.2	35.0	35.8	38.6
(5,1000)	41.1	46.1	22.7	25.7	**14.1**	**17.3**	29.4	37.4	33.0	35.8
(10,1000)	39.3	44.2	21.0	27.2	**13.7**	**18.7**	28.0	37.0	35.7	41.5
$T = 10000$:										
(5,100)	491.5	505.5	**489.4**	**496.5**	494.5	500.8	503.1	511.8	513.4	525.2
(10,100)	548.4	558.0	548.6	552.9	**529.3**	**534.7**	538.2	544.3	554.3	565.2
(5,250)	534.4	543.7	529.7	543.9	523.4	536.1	**519.7**	**525.5**	526.1	532.2
(10,250)	551.0	560.5	554.5	563.3	**547.4**	555.2	548.6	**555.1**	571.6	578.4
(5,1000)	669.0	704.4	570.5	584.8	538.8	552.7	**532.9**	**541.3**	535.8	558.4
(10,1000)	703.5	738.2	613.1	633.6	555.7	566.3	**549.9**	**559.5**	567.2	578.6

In Table 4.2, we compare the accumulated regret of our proposed policies with different granularity parameters δ, under a range of different parameter settings of number of nests M, number of items per nest N, and time horizon T. In Figure 4.3, we further plot the accumulated regret of our policies for time horizon when T is large (T between 10^5 and 10^7). From both Table 4.2 and Figure 4.3, one can see a clear pattern of sub-linear accumulated regret. Moreover, when N is small as compared to T, a smaller discretization granularity leads to better empirical performance; while when N is large, a larger discretization granularity is better.

4.3 The linear contextual logit model

While many discrete choice models associate each item for sale with a single utility parameter v_i, in many practical scenarios additional *contextual* information is available, such as their color, brand, size, texture, as well as customers' evolving demands. The objective of this section is to develop principled dynamic assortment recommendation methods incorporating such contextual information of items.

We assume that there are N items, conveniently labeled as $1, 2, \cdots, N$. At each time t, a set of time-sensitive "feature vectors" $v_{t1}, v_{t2}, \cdots, v_{tN} \in \mathbb{R}^d$ and revenues $r_{t1}, \cdots, r_{tN} \in [0, 1]$ are observed, reflecting time-varying changes of items' revenues and customers' preferences. A retailer, based on the features $\{v_{ti}\}_{i=1}^N$ and previous purchasing actions, picks an assortment $S_t \subseteq [N]$ under the cardinality constraint $|S_t| \leqslant K$ to present to an incoming customer; the retailer then observes a purchasing action $i_t \in S_t \cup \{0\}$ and collects the associated revenue r_{i_t} of the purchased item (if $i_t = 0$ then no item is purchased and zero revenue is collected).

We use an MNL model with features to characterize how a customer makes choices. Let $\theta_0 \in \mathbb{R}^d$ be an *unknown* time-invariant coefficient. For any $S \subseteq [N]$, the choice model $p_{\theta_0, t}(\cdot|S)$ is specified as (let $r_0 = 0$ and $v_{t0} = 0$)

$$p_{\theta_0, t}(j|S) = \frac{\exp\{v_{tj}^\top \theta_0\}}{1 + \sum_{k \in S} \exp\{v_{tk}^\top \theta_0\}} \qquad \forall j \in S \cup \{0\}. \tag{4.26}$$

For simplicity, in the rest of the paper we use $p_{\theta, t}(\cdot|S)$ to denote the law of the purchased item i_t conditioned on given assortment S at time t, parameterized by the coefficient $\theta \in \mathbb{R}^d$. The expected revenue $R_t(S)$ of assortment $S \subseteq [N]$ at time t is then given by

$$R_t(S) := \mathbb{E}_{\theta_0, t}[r_{tj}|S] = \frac{\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta_0\}}{1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta_0\}}. \tag{4.27}$$

Note that throughout the paper, we use $\mathbb{E}_{\theta_0, t}[\cdot|S]$ to denote the expectation with respect to the choice probabilities $p_{\theta_0, t}(j|S)$ defined in Eq. (4.26).

Our objective is to design policy π such that the regret

$$\text{Regret}(\{S_t\}_{t=1}^T) = \mathbb{E}^\pi \sum_{t=1}^T R_t(S_t^*) - R_t(S_t) \quad \text{where} \quad S_t^* = \arg \max_{S \subseteq [N], |S| \leqslant K} R_t(S) \tag{4.28}$$

is minimized. Here, S_t^* is an optimal assortment chosen when the full knowledge of choice probabilities is available (i.e., θ_0 is known).

Algorithm 13 The MLE-UCB policy for dynamic assortment optimization

---

1: **Input**: Number of pure explorations T_0, constraint radius τ.

2: **Output**: Assortment selections $\{S_t\}_{t=1}^T \subseteq [N]$ satisfying $|S_t| \leqslant K$.

3: Pure exploration: for $t = 1, \cdots, T_0$, pick $S_t = \{\ell_t\}$ for a single product ℓ_t sampled uniformly at random from $\{1, \cdots, N\}$ and record purchasing actions $(i_1, \cdots, i_{T_0})$;

4: Compute a pilot estimator using global MLE: $\theta^* \in \arg\max_{\theta \in \mathbb{R}^d} \sum_{t'=1}^{T_0} \log p_{\theta,t}(i_{t'}|S_{t'})$;

5: **for** $t = T_0 + 1$ to T **do**

6:     Observe revenue parameters $\{r_{tj}\}_{j=1}^N$ and preference features $\{v_{tj}\}_{j=1}^N$ at time t;

7:     Compute local MLE $\widehat{\theta}_{t-1} \in \arg\max_{\|\theta - \theta^*\|_2 \leqslant \tau} \sum_{t'=1}^{t-1} \log p_{\theta,t}(i_{t'}|S_{t'})$;

8:     For every assortment $S \subseteq [N]$, $|S| \leqslant K$, compute its upper confidence bound

$$\overline{R}_t(S) := \mathbb{E}_{\widehat{\theta}_{t-1},t}[r_{tj}|S] + \min\left\{1, \omega\sqrt{\|\widehat{I}_{t-1}^{-1/2}(\widehat{\theta}_{t-1})\widehat{M}_t(\widehat{\theta}_{t-1}|S)\widehat{I}_{t-1}^{-1/2}(\widehat{\theta}_{t-1})\|_{\mathrm{op}}}\right\};$$

$$\widehat{I}_{t-1}(\theta) := \sum_{t'=1}^{t-1} \widehat{M}_{t'}(\theta|S_{t'}); \quad \widehat{M}_t(\theta|S) := \mathbb{E}_{\theta,t}[v_{tj}v_{tj}^\top|S] - \{\mathbb{E}_{\theta,t}[v_{tj}|S]\}\{\mathbb{E}_{\theta,t}[v_{tj}|S]\}^\top;$$

$$\omega \asymp \sqrt{d\log(\rho\nu TK)};$$

9:     Pick $S_t \in \arg\max_{S \subseteq [N], |S| \leqslant K} \overline{R}_t(S)$ and observe purchasing action $i_t \in S_t \cup \{0\}$;

10: **end for**

    Remark: the expectations admit the following closed-form expressions:

$\mathbb{E}_{\theta,t}[r_{tj}|S] = \sum_{j \in S} p_{\theta,t}(j|S)r_{tj} = \frac{\sum_{j \in S} r_{tj}\exp\{v_{tj}^\top\theta\}}{1+\sum_{j \in S}\exp\{v_{tj}^\top\theta\}}$;

$\mathbb{E}_{\theta,t}[v_{tj}|S] = \sum_{j \in S} p_{\theta,t}(j|S)v_{tj} = \frac{\sum_{j \in S} v_{tj}\exp\{v_{tj}^\top\theta\}}{1+\sum_{j \in S}\exp\{v_{tj}^\top\theta\}}$;

$\mathbb{E}_{\theta,t}[v_{tj}v_{tj}^\top|S] = \sum_{j \in S} p_{\theta,t}(j|S)v_{tj}v_{tj}^\top = \frac{\sum_{j \in S} v_{tj}v_{tj}^\top\exp\{v_{tj}^\top\theta\}}{1+\sum_{j \in S}\exp\{v_{tj}^\top\theta\}}$.

4.3.1 An MLE-UCB policy and its regret

We propose an MLE-UCB policy, described in Algorithm 13.

The policy can be roughly divided into two phases. In the first *pure exploration* phase, the policy selects assortments uniformly at random, consisting of only *one item*. The objective of the pure exploration is to establish a "pilot" estimator of the unknown coefficient θ_0, i.e., a good initial estimator for θ_0. For the simplicity of the analysis, we choose one item for each assortment in this phase, which facilitates us to adapt existing analysis in the works of Filippi et al. (2010); Li et al. (2017b) as the MNL-logit choice model reduces to a *generalized linear model* when only one item is present in the assortment. In the second phase, we use a UCB-type approach that selects S_t as the assortment maximizing an *upper bound* $\overline{R}_t(S_t)$ of the expected revenue $R_t(S_t)$. Such upper bounds are built using a *local Maximum Likelihood Estimation (MLE)* of θ_0. In particular, in Step 5, instead of computing an MLE, we compute a local MLE, where the estimator $\widehat{\theta}_{t-1}$ lies in a ball centered at the pilot estimator θ^* with a radius τ. This localization also simplifies the technical analysis based on Taylor expansion, which benefits from the constraint

that $\widehat{\theta}_{t-1}$ is not too far away from θ^*.

To construct the confidence bound, we introduce the matrices $\widehat{M}_t(\widehat{\theta}_{t-1}|S)$ and $\widehat{I}_{t-1}(\widehat{\theta}_{t-1})$ in Step 6 of Algorithm 13, which are empirical estimates of the *Fisher's information* matrices $-\mathbb{E}[\nabla_\theta^2 \log p(\cdot|\theta)]$ corresponding to the MNL choice model $p(\cdot|S_t)$. The population version of the Fisher's information matrices are presented in Eq. (**??**). These quantities play an essential role in classical statistical analysis of maximum likelihood estimators (see, e.g., (Van der Vaart, 1998)).

The proposed MLE-UCB policy has three hyper-parameters: the coefficient $\omega > 0$ that controls the lengths of confidence intervals of $R_t(S)$, the number of pure exploration iterations T_0, and the radius τ_0 in the local MLE formulation. While theoretical values of ω, T_0 and τ are given in Theorem 19, which potentially depend on several unknown problem parameters, in practice we recommend the usage of $T_0 = \max\{d \log T, T^{1/4}\}$, $\omega = \sqrt{d \log T}$ and $\tau = 1/K$.

To establish rigorous regret upper bounds on Algorithm 13, we impose the following assumptions:

(A1) There exists a constant ν such that $\|v_{tj}\|_2 \leqslant \nu$ for all t and j. Moreover, for all $t \leqslant T_0$ and $j \in [N]$, v_{tj} are i.i.d. generated from an unknown distribution with the density μ satisfying that $\lambda_{\min}(\mathbb{E}_\mu vv^\top) \geqslant \lambda_0$ for some constant $\lambda_0 > 0$;

(A2) There exists a constant $\rho < \infty$ such that for all $t \in [T]$ and $S \subseteq [N]$ with $|S| \leqslant K$, $\frac{p_{\theta_0,t}(j|S)}{p_{\theta_0,t}(j'|S)} \leqslant \rho$ for all $j, j' \in S \cup \{0\}$.

The item (A1) assumes that the contextual information vectors $\{v_{tj}\}$ in the pure-exploration phase with $t \leqslant T_0$ are *randomly generated* from a non-degenerate density. It also places a standard boundedness condition on $\{v_{tj}\}$ for all time periods t. Note that after the pure-exploration phase, we allow the contextual vectors $\{v_{tj}\}$ to be adversarially chosen, only subject to boundedness conditions. (A2) additionally assumes a bounded ratio between the probability of choosing any two different items in an arbitrary assortment set. We remark that if $\|\theta_0\|_2 \leqslant C$, then the boundedness assumption in (A1) implies (A2) with $\rho \leqslant e^{2\max\{1,C\nu\}}$.

We are now ready to state our main result that upper bounds the worst-case accumulated regret of the proposed MLE-UCB policy in Algorithm 13.

Theorem 19. *Suppose that $T_0 \asymp \max\{\nu^2 d \log T/\lambda_0^2, \rho^2(d+\log T)/(\tau^2 \lambda_0)\}$ and $\tau \asymp 1/\sqrt{\rho^2 \nu^2 K^2}$ in Algorithm 13, then the regret of the MLE-UCB policy is upper bounded by*

$$C_1 \left[d\sqrt{T} \cdot \log(\lambda_0^{-1} \rho \nu T K) + d^2 \lambda_0^{-2} \rho^4 \nu^2 K^2 \log T \right] + C_2, \tag{4.29}$$

where $C_1, C_2 > 0$ are universal constants.

In addition to universal constants, the regret upper bound established in Theorem 19 has two terms. The first term, $d\sqrt{T} \cdot \log(\lambda_0^{-1} \rho \nu T K)$, is the main regret term that scales as $\widetilde{O}(d\sqrt{T})$ dropping logarithmic dependency. The second $d^2 \lambda_0^{-2} \rho^4 \nu^2 K^2 \log T$ term is a minor term, because it only scales logarithmically with the time horizon T. One remarkable aspect of Theorem 19 is the fact that the regret upper bound has *no* dependency on the total number of items N (even in a logarithmic term). This is an attractive property of the proposed policy, which allows N to be very large, even exponentially large in d and K.

While the computational task in Step 8 is quite challenging, approximation algorithms can be developed with rigorous performance guarantees. Interested readers should refer to (Chen et al.,

, Sec. 5) for further details.

4.3.2 A regret lower bound

To complement our regret analysis in the previous section, in this section we prove a *lower* bound for worst-case regret. Our lower bound is information theoretical, and therefore applies to *any* policy for dynamic assortment optimization with changing contextual features.

Theorem 20. *Suppose d is divisible by 4. There exists a universal constant $C_0 > 0$ such that for any sufficiently large T and policy π, there is a worst-case problem instance with $N \asymp K \cdot 2^d$ items and uniformly bounded feature and coefficient vector (i.e., $\|v_{ti}\|_2 \leqslant 1$ and $\|\theta_0\|_2 \leqslant 1$ for all $i \in [N]$, $t \in [T]$) such that the regret of π is lower bounded by $C_2 \cdot d\sqrt{T}/K$.*

Theorem 20 essentially implies that the $\widetilde{O}(d\sqrt{T})$ regret upper bound established in Theorem 19 is tight (up to logarithmic factors) in T and d. Although there is an $O(K)$ gap between the upper and lower regret bounds, in practical applications K is usually small and can be generally regarded as a constant. It is an interesting technical open problem to close this gap of $O(K)$.

We also remark that an $\Omega(d\sqrt{T})$ lower bound was established in the works of Dani et al. (2008) for contextual linear bandit problems. However, in assortment selection, the reward function is *not* coordinate-wise decomposable, making techniques in the works of Dani et al. (2008) not directly applicable.

4.3.3 Numerical results

In this section, we present numerical results of our proposed MLE-UCB algorithm. We use the greedy swapping heuristics (proposed in (Chen et al., 2018b, Algorithm 4)) as the subroutine to solve the combinatorial optimization problem in Step 8 of Algorithm 13.

Experiment setup. The unknown model parameter $\theta_0 \in \mathbb{R}^d$ is generated as a uniformly random unit d-dimensional vector. The revenue parameters $\{r_{tj}\}$ for $j \in [N]$ are independently and identically generated from the uniform distribution $[0.5, 0.8]$. For the feature vectors $\{v_{tj}\}$, each of them is independently generated as a uniform random vector v such that $\|v\| = 2$ and $v^\top \theta_0 < -0.6$. Here we set an upper bound of -0.6 for the inner product so that the utility parameters $u_{tj} = \exp\{v_{tj}^\top \theta_0\}$ are upper bounded by $\exp(-0.6) \approx 0.55$. We set such an upper bound because if the utility parameters are uniformly large, the optimal assortment is likely to pick very few items, leading to degenerated problem instances. In the implementation of our MLE-UCB algorithm, we simply set $T_0 = \lfloor\sqrt{T}\rfloor$ and $\omega = \sqrt{d\ln(TK)}$.

Performance of the MLE-UCB algorithm. In Figure 4.4a we plot the average regret (i.e. regret/T) of MLE-UCB algorithm with $N = 1000, K = 10, d = 5$ for the first $T = 10000$ time periods. For each experiment (in both Figure 4.4a and other figures), we repeat the experiment for 100 times and report the average value. In Figure 4.4b we compare our algorithm with the UCB algorithm for multinomial logit bandit (MNL-UCB) from the works of Agrawal et al. (2017a) without utilizing the feature information. Since the MNL-UCB algorithm assumes fixed item utilities that do not change over time, in this experiment we randomly generate one feature
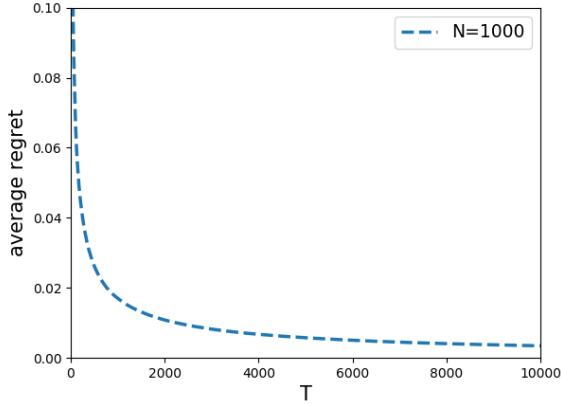
(a) Average regret of MLE-UCB
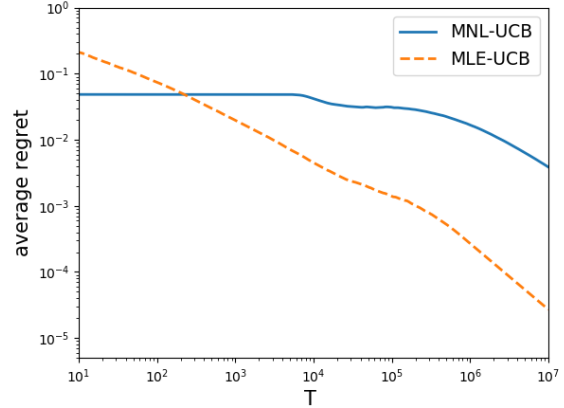
(b) Comparison of MLE-UCB and MNL-UCB

Figure 4.4: Illustration of the performance of MLE-UCB.
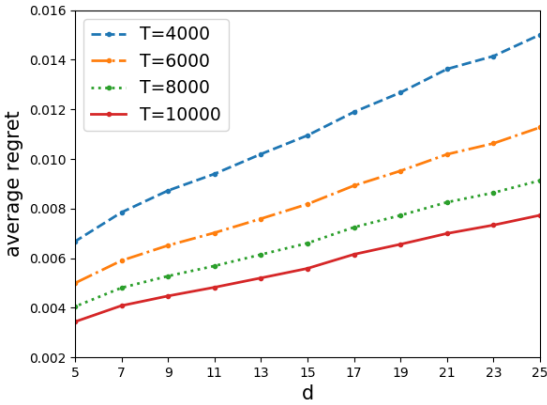
Figure 4.5: Average regret of MLE-UCB for various d's.
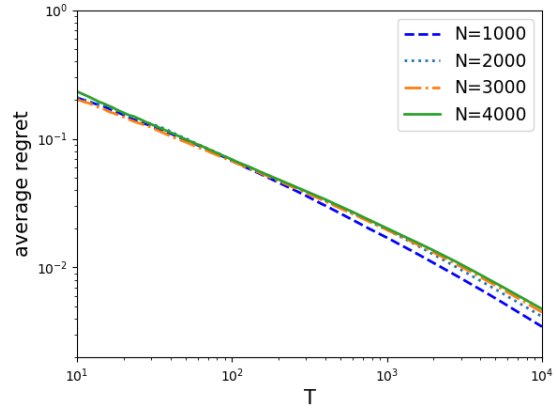
Figure 4.6: Average regret of MLE-UCB for various N's.

vector for each of the $N = 1000$ items and this feature vector will be fixed for the entire time span. We can observe that our MLE-UCB algorithm performs much better than MNL-UCB, which suggests the importance of taking the advantage of the contextual information.

Impact of the dimension size d. We study how the dimension of the feature vector impacts the performance of our MLE-UCB algorithm. We fix $N = 1000$ and $K = 10$ and test our algorithm for dimension sizes in $5, 7, 9, 11, \ldots, 25$. In Figure 4.5, we report the average regret at times $T \in \{4000, 6000, 8000, 10000\}$. We can see that the average regret increases approximately linearly with d. This phenomenon matches the linear dependency on d of the main term of the regret Eq. (4.29) of the MLE-UCB.

132

Impact of the number of items N. We compare the performance of our MLE-UCB algorithm for the varying number of items N. We fix $K = 10$ and $d = 5$, and test MLE-UCB for $N \in \{1000, 2000, 3000, 4000\}$. In Figure 4.6, we report the average regret for the first $T = 10000$ time periods. We observe that the regret of the algorithm is almost not affected by a bigger N. This confirms the fact that the regret Eq. (4.29) of MLE-UCB is totally independent of N.

4.4 Summary and related works

Assortment optimization plays a central role in revenue and recommendation management systems, with dynamic modeling and planning receiving much recent attention from the operations research and operations management society, which combines statistical modeling and sequential decision making at the same time. In my works, my collaborators and I extend the seminal works of Agrawal et al. (2017a); Rusmevichientong et al. (2010) by sharpening their regret upper bounds (Sec. 4.1), and considering more complex and practical assortment choice models (Secs. 4.2, 4.3).

Below we summarize some major literature on the stationary and dynamic assortment optimization problem. We also review relevant literature of online learning and bandit optimization.

Stationary and dynamic assortment planning Static assortment planning with known choice behavior has been an active research area since the seminal work by Mahajan & van Ryzin (2001); van Ryzin & Mahajan (1999). When the customer makes the choice according to the MNL model, Gallego et al. (2004); Talluri & van Ryzin (2004) prove the the optimal assortment will belong to revenue-ordered assortments. An alternative proof is provided in the work of Liu & van Ryzin (2008). This important structural result enables efficient computation of static assortment planning under the MNL model, which reduces the number of candidate assortments from 2^N to N and will also be used in our policy development.

Motivated by the large-scale online retailing, researchers start to relax the assumption on prior knowledge of customers' choice behavior. The question of *dynamic* optimization of assortments has received increasing attention in both the machine learning and operations management society Agrawal et al. (2017a,b); Caro & Gallien (2007); Rusmevichientong et al. (2010); Saure & Zeevi (2013), where the mean utilities of products are unknown and have to be learnt on the fly. Motivated by fast-fashion retailing, the work by Caro & Gallien (2007) was the first to study dynamic assortment planning problem, which assumes that the demand for product is independent of each other. The work Rusmevichientong et al. (2010) and Saure & Zeevi (2013) incorporate choice models of MNL into dynamic assortment planning and formulate the problem into a online regret minimization problem.

Assortment planning under nested logit models The nested logit model is considered as "the most widely used member of the GEV (generalized extreme value) family" and "has been applied by many researchers in a variety of situations" (see Chapter 4 from Train (2009)). It is well known that the standard MNL suffers from the independence of irrelevant alternatives (IIA), which implies proportional substitution across alternatives (see Chapter 4 from Train (2009)).

Davis et al. (2014) proved an important structural result that the optimal assortment within each nest is revenue-ordered, which will also be used in designing our dynamic policies. Assuming that there are M nests and N products within each nest, Li & Rusmevichientong (2014) further proposed an efficient greedy algorithm to find an optimal assortment set with $O(NM \log M)$ time complexity. Kök & Xu (2011) considered the joint assortment optimization and pricing problems with a restricted number of nests. There are several recent works on static assortment planning under variants of nested logit models. For example, Gallego & Topaloglu (2014) studied the constrained nested logit model, Li et al. (2015) extended the popular two-level nested logit model to a d-level nested logit model with $d \geqslant 2$. In addition, there are extensive research on static assortment optimization for more complex choice models, e.g., a robust version of MNL (Rusmevichientong & Topaloglu, 2012), the mixture of logit models (Bront et al., 2009; Méndez-Díaz et al., 2014; Rusmevichientong et al., 2014), Markov chain-based choice models (Blanchet et al., 2016), the generalized attraction model (Wang, 2013), Mallows-based choice models (Désir et al., 2016), a multiple attempt model (Chung et al., 2019), and a general class of choice models based on a distribution over permutations (Farias et al., 2013).

Assortment planning under contextual models   Personalized assortment optimization has attracted much research effort recently. By incorporating the feature information of each arriving customer, both the static and dynamic assortment optimization problems are studied in the works of Chen et al. (2015b) and Cheung & Simchi-Levi (2017), respectively. Other research studies personalized assortment optimization in an adversarial setting rather than stochastic setting. For example, Chen et al. (2016); Golrezaei et al. (2014) assumed that each customer's choice behavior is known, but that the customers' arriving sequence (or customers' types) can be adversarially chosen and took the inventory level into consideration. Since the arriving sequence can be arbitrary, there is no learning component in the problem and both the works of Golrezaei et al. (2014) and Chen et al. (2016) adopted the competitive ratio as the performance evaluation metric.

Unimodal bandits   The assortment optimization problem with the plain logit model under the uncapacitated setting is closely related to multi-armed bandit problems with unimodal constraints (Agarwal et al., 2013; Combes & Proutiere, 2014; Cope, 2009; Yu & Mannor, 2011), in which discrete or continuous multi-armed bandit problems are considered with additional unimodality constraints on the means of the arms. Apart from unimodality, additional structures such as "inverse Lipschitz continuity" (e.g., $|\mu(i) - \mu(j)| \geqslant L|i-j|$ for some constant L, where $\mu(i)$ denotes the mean reward of the i-th arm) or convexity are imposed to ensure smaller regret compared to unstructured multi-armed bandits. However, both conditions fail to hold for the revenue potential function arising from uncapacitated MNL-based assortment planning problems. In addition, under the gap-free setting where an $O(\sqrt{T})$ regret is to be expected, most previous works have additional $\log T$ terms in their regret upper bounds (except for the work of Cope (2009) which introduces additional strong regularity conditions on the underlying functions). In the work of Cohen-Addad & Kanade (2017), a more general problem of optimizing piecewise-constant function is considered, without assuming a unimodal structure of the function. Consequently, a weaker $\widetilde{O}(T^{2/3})$ regret is derived.

Contextual bandits The assortment optimization problem with linear contextual modeling is related to the contextual bandit problem in the bandit online learning literature, and is particularly connected to the linear and generalized linear bandits (Abbasi-Yadkori et al., 2011; Abe et al., 2003; Agrawal & Goyal, 2013; Auer, 2002; Chu et al., 2011; Dani et al., 2008; Filippi et al., 2010; Li et al., 2017b; Rusmevichientong & Tsitsiklis, 2010). The assortment optimization problem is technically *not* a generalized linear model and is therefore more challenging. Moreover, in contextual bandit problems, only one arm is selected by the decision-maker at each time period. In contrast, each action in an assortment optimization problem involves a set of items, which makes the action space more complicated.

4.5 Proofs of results in Sec. 4.1

4.5.1 Proof of Lemma 52

Let $s < s'$ be the two endpoints such that $F(s^+) = F(s') = F^*$ (if there are multiple such s, s' pairs, pick any one of them). We will prove that $s < F^* \leqslant s'$, which then implies Lemma 52.

We first prove $s < F^*$. Assume by contradiction that $F^* \leqslant s$. Clearly $s \neq 0$ because $F^* > 0$. By definition of F and F^*, we have

$$F^* = F(s') = \frac{\sum_{r_i \geqslant s'} r_i v_i}{1 + \sum_{r_i \geqslant s'} v_i} \implies \sum_{r_i \geqslant s'} (r_i - F^*) v_i = F^*. \tag{4.30}$$

Because $F^* \leqslant s$, adding we have that

$$\sum_{r_i \geqslant s} (r_i - F^*) v_i \geqslant F^* \implies F(s) \geqslant F^*. \tag{4.31}$$

This contradicts with the fact that $F(s) \neq F(s^+)$ and that F^* is the maximum value of F.

We next prove $F^* \leqslant s'$. Assume by contradiction that $F^* > s'$. Removing all items corresponding to $r_i = s'$ in Eq. (4.30), we have

$$\sum_{r_i > s'} (r_i - F^*) v_i \geqslant F^* \implies F(s'^+) \geqslant F^*. \tag{4.32}$$

This contradicts with the fact that $F(s'^+) \neq F(s')$ and that F^* is the maximum value of F.

4.5.2 Proof of Lemma 53

Because $F(\theta^*) = \theta^* = F^*$ and F^* is the maximum value of F, we have $F(\theta) \leqslant \theta$ for all $\theta \geqslant \theta^*$. In addition, for any $\theta \geqslant \theta^*$, by definition of F we have

$$F(\theta) - F(\theta^+) = R(\{i \in \mathcal{N} : r_i \geqslant \theta\}) - R(\{i \in \mathcal{N} : r_i > \theta\}) \tag{4.33}$$

$$= \frac{\sum_{r_i \geqslant \theta} r_i v_i}{1 + \sum_{r_i \geqslant \theta} v_i} - \frac{\sum_{r_i > \theta} r_i v_i}{1 + \sum_{r_i > \theta} v_i} \tag{4.34}$$

$$= \frac{(1 + \sum_{r_i > \theta} v_i)(\sum_{r_i \geqslant \theta} r_i v_i) - (1 + \sum_{r_i \geqslant \theta} v_i)(\sum_{r_i > \theta} r_i v_i)}{(1 + \sum_{r_i \geqslant \theta} v_i)(1 + \sum_{r_i > \theta} v_i)} \tag{4.35}$$

$$= \frac{(1 + \sum_{r_i > \theta} v_i)(\sum_{r_i = \theta} r_i v_i) - (\sum_{r_i = \theta} v_i)(\sum_{r_i > \theta} r_i v_i)}{(1 + \sum_{r_i \geqslant \theta} v_i)(1 + \sum_{r_i > \theta} v_i)} \tag{4.36}$$

$$= \frac{\sum_{r_i = \theta} v_i}{1 + \sum_{r_i \geqslant \theta} v_i} \left[\theta - F(\theta^+) \right]. \tag{4.37}$$

Because $\theta \geqslant F(\theta)$ holds for all $\theta \geqslant \theta^*$, we conclude that $\theta \geqslant F(\theta^+)$ also holds for all $\theta \geqslant \theta^*$. Subsequently, the right-hand side of Eq. (4.37) is non-negative and therefore $F(\theta) \geqslant F(\theta^+)$.

4.5.3 Proof of Lemma 54

If $F(\theta) \equiv F^*$ for all $\theta \leqslant \theta^*$ then the lemma clearly holds. In the rest of the proof we shall assume that there is at least one jumping point strictly smaller than θ^*. Formally, we let $0 < s_1 < s_2 < \cdots < s_t < \theta^*$ be all jumping points that are strictly smaller than θ^*. To prove Lemma 54, it suffices to show that $F(s_j) \geqslant s_j$ and $F(s_j) \geqslant F(s_j^+)$ for all $j = 1, \cdots, t$.

We use induction to establish the above claims. The base case is $j = t$. Because F^* is the maximum value of F, we conclude that $F(s_t) \leqslant F^* = F(s_t^+)$. In addition, because $s_t \leqslant \theta^* = F^* = F(s_t^+)$, invoking Eq. (4.37) we have that $F(s_t) \leqslant F(s_t^+)$. The base case is then proved.

We next prove the claim for s_j, assuming it holds for s_{j+1} by induction. By inductive hypothesis, $F(s_{j+1}) \geqslant s_{j+1} \geqslant s_j$. Also, $F(s_j^+) = F(s_{j+1})$ because there is no jump points between s_j and s_{j+1}, and subsequently $F(s_j^+) \geqslant s_j$. Invoking Eq. (4.37) we proved $F(s_j) \leqslant F(s_j^+)$.

To prove $F(s_j) \geqslant s_j$, define $\gamma_j := (\sum_{r_i = s_j} v_i)/(1 + \sum_{r_i \geqslant s_j} v_i)$. It is clear that $0 \leqslant \gamma_j \leqslant 1$. By Eq. (4.37), we have

$$F(s_j) - s_j = F(s_j) - F(s_j^+) + F(s_j^+) - s_j \tag{4.38}$$

$$= \gamma_j \left[s_j - F(s_j^+) \right] + F(s_j^+) - s_j \tag{4.39}$$

$$= (1 - \gamma_j) \left[F(s_j^+) - s_j \right]. \tag{4.40}$$

As we have already proved $F(s_j^+) \geqslant s_j$, the right-hand side of the above inequality is non-negative and therefore $F(s_j) \geqslant s_j$.

4.5.4 Proof of Theorem 13

We first state a simple lemma showing that the confidence bound $\ell_t(y_\tau)$ and $u_t(y_\tau)$ constructed in Algorithm 8 contains $F(y_\tau)$ with high probability.

Lemma 60. *With probability* $1 - O(T^{-1})$, $\ell_t(\theta) \leqslant F(\theta) \leqslant u_t(\theta)$ *for all* t.

Proof. Let $\delta = 1/T^2$ be the confidence parameter in Algorithm 9. By Hoeffding's inequality (Lemma 89) and the fact that $0 \leqslant F(\theta) \leqslant 1$ for all θ, we have

$$\Pr\left[F(\theta) \notin [\ell_t(\theta), u_t(\theta)] \right] = \Pr\left[\left| \frac{\rho_t(\theta)}{t} - F(\theta) \right| > \sqrt{\frac{\ln(1/\delta)}{2t}} \right] \tag{4.41}$$

$$\leqslant 2 \exp \left\{ -2t \cdot \frac{\ln(1/\delta)}{2t} \right\} \leqslant 2\delta = 2/T^2. \tag{4.42}$$

Subsequently, by union bound the probability of $F(\theta) \notin [\ell_t(\theta), u_t(\theta)]$ for *at least* one t is at most $O(T^{-1})$. $\square$

The following lemma, based on properties of the potential function F and Lemma 60, establishes that (with high probability) the shrinkage of a_τ or b_τ are "consistent"; i.e., θ^* is always contained in $[a_\tau, b_\tau]$. Its proof is based on the intuitive two-case analysis discussed before Theorem 13 and will be provided later.

Lemma 61. *With probability $1 - O(T^{-1})$, $a_\tau \leqslant \theta^* \leqslant b_\tau$ for all $\tau = 1, 2, \cdots, \tau_0$, where τ_0 is the last outer iteration of Algorithm 8.*

Using Lemmas 60 and 61, we are able to prove the following lemma that upper bounds the regret incurred at each outer iteration τ using the distance between the trisection points x_τ and y_τ.

Lemma 62. *For $\tau = 0, 1, \cdots$ let $\mathcal{T}(\tau)$ denote the set of all indices of inner iterations at outer iteration τ. Conditioned on the success events in Lemmas 60 and 61, it holds that*

$$\mathbb{E} \sum_{t \in \mathcal{T}(\tau)} R(S^*) - R(S_t) \lesssim \varepsilon_\tau^{-1} \log T. \tag{4.43}$$

We are now ready to prove Theorem 13.

Proof. Recall the definition that $\varepsilon_\tau = y_\tau - x_\tau$ for outer iterations $\tau = 0, 1, \cdots$. Because after each outer iteration we either set $b_{\tau+1} = y_\tau$ or $a_{\tau+1} = x_\tau$, it is easy to verify that $\varepsilon_\tau = (2/3) \cdot \varepsilon_{\tau-1}$. Subsequently, invoking Lemma 61 and using summation of geometric series we have

$$\mathbb{E} \sum_{t=1}^{T} R(S^*) - R(S_t) \lesssim \sum_{\tau=0}^{\tau_0} \varepsilon_\tau^{-1} \log T \lesssim \varepsilon_{\tau_0}^{-1} \log T, \tag{4.44}$$

where τ_0 is the total number of outer iterations executed by Algorithm 8. On the other hand, because at each outer iteration τ the revenue level a_τ is exploited for exactly $n_\tau = 16\lceil (y_\tau - x_\tau)^{-2} \ln(T^2) \rceil$ times, we have

$$T \geqslant n_{\tau_0} \gtrsim \varepsilon_{\tau_0}^{-2} \log T. \tag{4.45}$$

Combining Eqs. (4.44) and (4.45) we conclude that $\text{Regret}(\{S_t\}_{t=1}^{T}) \lesssim \sqrt{T \log T}$. $\square$

Proof of Lemma 61 We use induction to prove this lemma. We also conditioned on the fact that $\ell_t(x_\tau) \leqslant F(x_\tau) \leqslant u_t(x_\tau)$ and $\ell_t(y_\tau) \leqslant F(y_\tau) \leqslant u_t(y_\tau)$ for all t and τ, which happens with probability at least $1 - O(T^{-1})$ by Lemma 60.

We first prove the lemma for the base case of $\tau = 0$. According to the initialization step in Algorithm 8, we have $a_\tau = 0$ and $b_\tau = 1$. On the other hand, for any $\theta \geqslant 0$ it holds that $0 \leqslant F(\theta) \leqslant F^* \leqslant 1$. Therefore, $0 \leqslant \theta^* \leqslant 1$ and hence $a_\tau \leqslant \theta^* \leqslant b_\tau$ for $\tau = 0$.

We next prove the lemma for outer iteration τ, assuming the lemma holds for outer iteration $\tau - 1$ (i.e., $a_{\tau-1} \leqslant r^* \leqslant b_{\tau-1}$). According to the trisection parameter update step in Algorithm 8, the proof can be divided into two cases:

137

Case 1: $u_t(y_{\tau-1}) < y_{\tau-1}$. Because $\ell_t(y_{\tau-1}) \leqslant F(y_{\tau-1}) \leqslant u_t(y_{\tau-1})$ always holds, we conclude in this case that $F(y_{\tau-1}) < y_{\tau-1}$. Invoking Lemma 54 we conclude that $b_\tau = y_{\tau-1} > \theta^*$. On the other hand, by inductive hypothesis $a_\tau = a_{\tau-1} \leqslant \theta^*$. Therefore, $a_\tau \leqslant r^* \leqslant b_\tau$.

Case 2: $u_t(y_{\tau-1}) \geqslant y_{\tau-1}$. In this case, the revenue level $y_{\tau-1}$ must be explored at every inner iteration in Algorithm 8 at outer iteration $\tau - 1$, because $u_t(y_{\tau-1})$ is a non-increasing function of t. Denote $\varepsilon_\tau = y_\tau - x_\tau$ and $n_\tau = 16\lceil \varepsilon_\tau^{-2} \ln(T^2) \rceil$ as the number of inner iterations in outer iteration τ. Subsequently, the length of the confidence intervals on $y_{\tau-1}$ at the end of all inner iterations can be upper bounded by

$$|u_t(y_{\tau-1}) - \ell_t(y_{\tau-1})| \leqslant 2\sqrt{\frac{\ln(T^2)}{n_\tau}} \leqslant \frac{1}{2}\varepsilon_\tau^{-1}. \tag{4.46}$$

Invoking Lemma 60 we then have

$$F(y_{\tau-1}) \geqslant \ell_t(y_{\tau-1}) \geqslant u_t(y_{\tau-1}) - \frac{y_{\tau-1} - x_{\tau-1}}{2} \geqslant y_{\tau-1} - \frac{y_{\tau-1} - x_{\tau-1}}{2}. \tag{4.47}$$

We now establish that $x_{\tau-1} \leqslant \theta^*$, which implies $a_\tau \leqslant \theta^* \leqslant b_\tau$ because $a_\tau = x_{\tau-1}$ and $b_\tau = b_{\tau-1} \geqslant \theta^*$ by the inductive hypothesis. Assume by contradiction that $x_{\tau-1} > \theta^*$. By Lemma 53, $F(x_{\tau-1}) \leqslant x_{\tau-1}$ and $F(x_{\tau-1}) \geqslant F(y_{\tau-1})$. Subsequently,

$$F(y_{\tau-1}) \leqslant x_{\tau-1} = y_{\tau-1} - (y_{\tau-1} - x_{\tau-1}) < y_{\tau-1} - \frac{y_{\tau-1} - x_{\tau-1}}{2}, \tag{4.48}$$

which contradicts Eq. (4.47).

Proof of Lemma 62 This lemma upper bounds the expected regret incurred at each outer iteration τ, conditioned on the success events in Lemmas 60 and 61.

We analyze the regret incurred at outer iteration τ from exploration of y_τ and exploitation of a_τ separately.

1. *Regret from exploring y_τ:* suppose the level set $\mathcal{L}_{y_\tau}(\mathcal{N})$ is explored for $m_\tau \leqslant n_\tau$ times at outer iteration τ. Then we have $u_{m_\tau}(y_\tau) \geqslant y_\tau$. In addition, by Lemma 60 and widths in the constructed confidence bands $\ell_{m_\tau}(y_\tau)$ and $u_{m_\tau}(y_\tau)$, we have with probability $1 - O(T^{-1})$ that $\ell_{m_\tau}(y_\tau) \leqslant F(y_\tau) \leqslant u_{m_\tau}(y_\tau)$ and $|u_{m_\tau}(y_\tau) - \ell_{m_\tau}(y_\tau)| \leqslant 2\sqrt{(\ln(T^2)/2m_\tau}$. Subsequently,

$$F(y_\tau) \geqslant \ell_{m_\tau}(y_\tau) \geqslant u_{m_\tau}(y_\tau) - 2\sqrt{\frac{\ln(T^2)}{2m_\tau}} \geqslant y_\tau - 2\sqrt{\frac{\ln T}{m_\tau}}. \tag{4.49}$$

Note also that $y_\tau \geqslant a_\tau \geqslant \theta^* - 3\varepsilon_\tau = F^* - 3\varepsilon_\tau$; we have

$$F^* - F(y_\tau) \leqslant 3\varepsilon_\tau + 2\sqrt{\frac{\ln T}{m_\tau}}. \tag{4.50}$$

By Lemma 52, $F^* = R(S^*)$ and therefore the right-hand side of the above inequality is an upper bound on the regret incurred by exploring revenue level y_τ (corresponding to the

138

assortment selection $\mathcal{L}_{y_\tau}$) once. As the exploration is carried out for m_τ times, the total regret for all exploration steps at revenue level x_τ can be upper bounded by

$$m_\tau \left[3\varepsilon_\tau + 2\sqrt{\frac{\ln T}{m_\tau}} \right] \leq 3m_\tau \varepsilon_\tau + \sqrt{4m_\tau \ln T} \leq 3n_\tau \varepsilon_\tau + \sqrt{4n_\tau \ln T} \lesssim \varepsilon_\tau^{-1} \log T. \quad (4.51)$$

Here the last inequality holds because $n_\tau \leq 16\varepsilon_\tau^{-2} \ln(T^2)$.

2. *Regret from exploiting a_τ*: by Lemma 61, $a_\tau \leq \theta^*$, and therefore $F(a_\tau) \geq a_\tau$. In addition, $a_\tau \geq \theta^* - 3\varepsilon_\tau$ by the definition of ε_τ. Subsequently,

$$F(a_\tau) \geq a_\tau \geq \theta^* - 3\varepsilon_\tau = F^* - 3\varepsilon_\tau. \quad (4.52)$$

Re-organizing terms on both sides of the above inequality and noting that $F^* = F(S^*)$, we have

$$F(S^*) - F(a_\tau) \leq 3\varepsilon_\tau. \quad (4.53)$$

Therefore, the regret for each exploitation of revenue level a_τ (corresponding to the assortment selection $\mathcal{L}_{a_\tau}$) can be upper bounded by ε_τ. Because the revenue level a_τ is exploited for n_τ times and $n_\tau \leq 16\varepsilon_\tau^{-2} \ln(T^2)$, the total regret of exploitation of a_τ at outer iteration τ can be upper bounded by

$$n_\tau \cdot 3\varepsilon_\tau \lesssim \varepsilon_\tau^{-1} \log T. \quad (4.54)$$

4.5.5 Proof of Lemma 55

Without loss of generality we assume $X_1, \cdots, X_L \in [0, 1]$ almost surely, while the general case of $X_1, \cdots, X_L \in [a, b]$ can be dealt with by a simple re-scaling argument. Denote $k := \lfloor \log_2 L \rfloor$. For each $\ell \in \{1, 2, 4, \cdots, 2^k\}$, by standard Hoeffding's inequality (Lemma 89), we have

$$\Pr\left[\left| \frac{1}{\ell} \sum_{i=1}^{\ell} X_i - \mu \right| \leq \sqrt{\frac{\ln[8/(\delta\ell)]}{2\ell}} \right] \geq 1 - \frac{\delta\ell}{4}.$$

Subsequently, by union bound and the fact that $1 + 2 + 4 + \cdots + 2^k \leq 2^{k+1} \leq 2L$, we have

$$\Pr\left[\forall \ell = 1, 2, 4, \cdots, 2^k, \left| \frac{1}{\ell} \sum_{i=1}^{\ell} X_i - \mu \right| \leq \sqrt{\frac{\ln[8/(\delta\ell)]}{2\ell}} \right] \geq 1 - \frac{\delta L}{2}. \quad (4.55)$$

Next consider any $\ell \in \{1, 2, 4, \cdots, 2^k\}$. By Hoeffding's maximal inequality (Lemma 93), we have

$$\Pr\left[\forall i \leq \min\{\ell, n - \ell\}, \left| X_{\ell+1} + \cdots + X_{\ell+i} - i \cdot \mu \right| \leq \sqrt{\frac{\ell}{2} \ln[8/(\delta\ell)]} \right] \geq 1 - \frac{\delta\ell}{4}.$$

Again using union bound over all $\ell = 1, 2, 4, \cdots, 2^k$ and the fact that $1 + 2 + 4 + \cdots + 2^k \leq 2^{k+1} \leq 2L$, we have

$$\Pr\left[\forall \ell = 1, 2, \cdots, 2^k, i \leq \min\{\ell, n - \ell\}, \left| X_{\ell+1} + \cdots + X_{\ell+i} - i \cdot \mu \right| \leq \sqrt{\frac{\ell}{2} \ln[8/(\delta\ell)]} \right] \geq 1 - \frac{\delta L}{2}.$$

$$(4.56)$$

Combining Eqs. (4.55,4.56), we have with probability $1-\delta L$ uniformly over all $\ell = 1, 2, 4, \cdots, 2^k$ and $i \leqslant \min\{\ell, n - \ell\}$ that

$$\left|X_1 + \cdots + X_\ell + X_{\ell+1} + \cdots + X_{\ell+i} - (\ell + i)\mu\right| \leqslant \sqrt{2\ell \ln[8/(\delta\ell)]}.$$

Dividing both sides of the above inequality by $(\ell + i)$ we complete the proof of Lemma 55.

4.5.6   Proof of Theorem 14

We first define some notations. Let $\tau = 0, 1, \cdots$ be the number of outer iterations in Algorithm 8, $\varepsilon_\tau = (y_\tau - x_\tau)$ be the distance between the two trisection points at outer iteration τ, and $n_\tau = 8\lceil\varepsilon_\tau^{-2}\ln(8T\varepsilon_\tau^2)\rceil$ be the pre-specified number of inner iterations. Recall also that $\theta^* = F(\theta^*) = F^*$ is the optimal revenue value suggested by Lemma 52.

Define the following three disjoint events that partition the entire probabilistic space:

- Event $\mathcal{E}_1(\tau)$: $\theta^* < a_\tau < b_\tau$;

- Event $\mathcal{E}_2(\tau)$: $a_\tau \leqslant \theta^* \leqslant b_\tau$;

- Event $\mathcal{E}_3(\tau)$: $a_\tau < b_\tau < \theta^*$.

Let $\tau_0 \in \mathbb{N}$ be the last outer iteration in Algorithm 8. Let also $\mathcal{T}(\tau) \subseteq [T]$ be the indices of inner iterations in outer iteration τ, satisfying $|\mathcal{T}(\tau)| \leqslant 2n_\tau$ almost surely. For $\omega \in \{1, 2, 3\}$, $\tau \in \mathbb{N}$ and $\alpha, \beta \in \mathbb{R}^+$, define

$$\psi_\tau^\omega(\alpha, \beta) := \mathbb{E}\left[\sum_{\tau'=\tau}^{\tau_0} \sum_{t\in\mathcal{T}(\tau')} R(S^*) - R(S_t)\Big|\mathcal{E}_\omega(\tau), |a_\tau - \theta^*| = \alpha, |F(a_\tau) - a_\tau| = \beta\right]. \quad (4.57)$$

Intuitively, $\psi_\tau^\omega(\alpha, \beta)$ is the expected regret Algorithm 8 incurs for outer iterations $\tau, \tau+1, \cdots, \tau_0$, conditioned on the event $\mathcal{E}_\omega(\tau)$ and other boundary conditions at the left margin a_τ.

The following three lemmas are the central steps in our proof, which establish recurrence relationships among $\psi_\tau^\omega(\alpha, \beta)$, for $\omega \in \{1, 2, 3\}$. The proofs are technically involved and given later. To simplify notations, we write $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ if there exists a *universal* constant $C > 0$ such that $|a_n| \leqslant C|b_n|$ for all $n \in \mathbb{N}$.

Lemma 63 (Regret in Case 1). $\psi_\tau^1(\alpha, \beta) \leqslant \beta T + \sum_{\tau'=\tau+1}^{\tau_0} \sup_{\Delta>\varepsilon_{\tau'}} \Delta T \exp\{-n_\tau\Delta^2\} + O(\varepsilon_{\tau'}^{-1}\log(T\varepsilon_{\tau'}^2))$.

Lemma 64 (Regret in Case 2). $\psi_\tau^2(\alpha, \beta) \leqslant O(\varepsilon_\tau^{-1}\log(T\varepsilon_\tau^2)) + \psi_{\tau+1}^2(\alpha_2', \beta_2') + \psi_{\tau+1}^3(\alpha_3', \beta_3') \cdot O(\log(T\varepsilon_\tau^2)/(T\varepsilon_\tau^2)) + \sup_{\Delta>\varepsilon_\tau}\psi_{\tau+1}^1(\alpha_1', \beta_1'(\Delta))\exp\{-n_\tau\Delta_\tau^2\}$ *for parameters* $\alpha_1', \beta_1'(\Delta), \alpha_2', \beta_2', \alpha_3', \beta_3'$ *that satisfy* $\beta_1'(\Delta) \leqslant \Delta$ *and* $\alpha_3' \leqslant 3\varepsilon_\tau$.

Lemma 65 (Regret in Case 3). $\psi_\tau^3(\alpha, \beta) \leqslant \alpha T$.

We are now ready to complete the proof of Theorem 14 by combining Lemmas 64, 63 and 65.

Proof. We first get a cleaning expression of $\psi_\tau^1(\alpha, \beta)$ using Lemma 63. First note that $\Delta \mapsto \Delta\exp\{-n_\tau\Delta^2\}$ attains its maximum on $\Delta > 0$ at $\Delta = \sqrt{1/2n_\tau}$. Also note that $n_\tau = \lceil 8\varepsilon_\tau^{-2}\ln(8T\varepsilon_\tau^2)\rceil$ and therefore $\sqrt{1/2n_\tau} \leqslant \varepsilon_\tau$. Subsequently,

$$\sum_{\tau'=\tau}^{\tau_0} \sup_{\Delta>\varepsilon_\tau} \Delta T \exp\{-n_\tau\Delta^2\} \leqslant \sum_{\tau'=\tau}^{\tau_0} \varepsilon_\tau T \exp\{-n_\tau\varepsilon_\tau^2\} \leqslant \sum_{\tau'=\tau}^{\tau_0} \varepsilon_\tau T \exp\{-\ln(T\varepsilon_\tau^2)\}$$

$$\leqslant \sum_{\tau'=\tau}^{\tau_0} \varepsilon_\tau^{-1} = O(\varepsilon_{\tau_0}^{-1}), \tag{4.58}$$

where the last asymptotic holds because $\{\varepsilon_\tau\}$ forms a geometric series. Subsequently,

$$\psi_\tau^1(\alpha, \beta) \leqslant \beta T + \sum_{\tau'=\tau}^{\tau_0} O(\varepsilon_{\tau'}^{-1} \log(T\varepsilon_\tau^2)). \tag{4.59}$$

It remains the bound the summation term on the right-hand side of the above inequality. Denote $s_{\tau'} = \varepsilon_{\tau'}^{-1} \ln(T\varepsilon_{\tau'}^2) = \rho^{-\tau'} \ln(T\rho^{2\tau'})$, where $\rho = 2/3$. We then have $s_{\tau'} = \rho^{\tau_0-\tau'}[1 + \ln\rho^{-2(\tau_0-\tau')}]s_{\tau_0} \leqslant 2(\tau_0 - \tau' + 1)\rho^{\tau_0-\tau'} \ln(1/\rho)$ for all $\tau' \leqslant \tau_0$. Subsequently,

$$\sum_{\tau'=\tau}^{\tau_0} s_{\tau'} \leqslant \sum_{\tau'=0}^{\tau_0} 2(\tau_0 - \tau' + 1)\rho^{\tau_0-\tau'} \ln(1/\rho) \cdot s_{\tau_0} \leqslant O(1) \cdot s_{\tau_0}. \tag{4.60}$$

Therefore,

$$\psi_\tau^1(\alpha, \beta) \leqslant \beta T + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)). \tag{4.61}$$

We are now ready to derive the final regret upper bound by analyzing $\psi_0^2(\alpha, \beta)$, because the event $\mathcal{E}_2(0)$ always holds since $0 \leqslant \theta^* \leqslant 1$. Applying Lemma 64 with Lemma 65 and Eq. (4.61), we have for all $\tau \in \{0, 1, \cdots, \tau_0\}$ that

$$\begin{aligned}
\psi_\tau^2(\alpha, \beta) &\leqslant \psi_{\tau+1}^2(\alpha_2', \beta_2') + O(\varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2)) + O(\varepsilon_\tau T) \cdot \frac{\ln(T\varepsilon_\tau^2)}{T\varepsilon_\tau^2} \\
&\quad + \sup_{\Delta > \varepsilon_\tau} \left(\Delta T + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2))\right) \exp\{-n_\tau \Delta^2\} \\
&\leqslant \psi_{\tau+1}^2(\alpha_2', \beta_2') + O(\varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2)) + \sup_{\Delta > \varepsilon_\tau} \Delta T \exp\{-n_\tau \Delta^2\} \\
&\quad + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)) \cdot \exp\{-n_\tau \varepsilon_\tau^2\}. \tag{4.62}
\end{aligned}$$

Using the same analysis as in Eq. (4.58), we know $\sup_{\Delta > \varepsilon_\tau} \Delta T \exp\{-n_\tau \Delta^2\} \leqslant O(\varepsilon_\tau^{-1})$ and $\exp\{-n_\tau \varepsilon_\tau^2\} \leqslant 1/(T\varepsilon_\tau^2)$. Subsequently, summing all terms $\tau = 0, 1, \cdots, \tau_0$ together we have

$$\begin{aligned}
\psi_0^2(\alpha, \beta) &\leqslant \sum_{\tau=0}^{\tau_0} O(\varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2)) + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)) \cdot \frac{1}{T\varepsilon_\tau^2} \\
&\lesssim \varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2) \cdot (1 + 1/(T\varepsilon_{\tau_0}^2)). \tag{4.63}
\end{aligned}$$

Finally, note that $n_{\tau_0} \gtrsim \varepsilon_{\tau_0}^{-2}$ and $n_{\tau_0} \leqslant T$, implying that $\varepsilon_{\tau_0} \gtrsim \sqrt{1/T}$. Plugging the lower bound on ε_{τ_0} into the above inequality we have $\psi_0^2(\alpha, \beta) \lesssim \sqrt{T}$, which completes the proof of Theorem 14. $\qquad\square$

Proof of Lemma 64 First analyze the expected regret incurred at outer iteration τ. by exploiting the left end-point a_τ (corresponding to assortment $\mathcal{L}_{a_\tau}$) for n_τ iterations. Also, because $a_\tau \leqslant \theta^* \leqslant b_\tau$ conditioned on $\mathcal{E}_2(\tau)$, by Lemmas 52 and 54 we have $F(a_\tau) \geqslant a_\tau \geqslant \theta^* - |b_\tau - a_\tau| = F(\theta^*) - |b_\tau - a_\tau| \geqslant R(S^*) - 3\varepsilon_\tau$. Subsequently,

$$\text{Regret by exploiting } \mathcal{L}_{a_\tau}: \quad \leqslant 3\varepsilon_\tau \cdot n_\tau \lesssim \varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2). \tag{4.64}$$

Next we analyze the expected regret incurred at outer iteration τ by exploring the right trisection point y_τ (corresponding to assortment $\mathcal{L}_{y_\tau}$). This is done by a case analysis. If $y_\tau \leqslant \theta^*$, then the regret incurred by exploiting $\mathcal{L}_{y_\tau}$ at outer iteration τ is again upper bounded (up to numerical constants) by $\varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2)$, similar to Eq. (4.64). Otherwise, for the case of $y_\tau > \theta^*$, define $\Delta_\tau := y_\tau - F(y_\tau)$. By Lemma 53, we know $\Delta_\tau \geqslant 0$, and also by Lemma 52, each exploration of $\mathcal{L}_{y_\tau}$ incurs a regret of no more than Δ_τ. Let m_τ be the number of times $\mathcal{L}_{y_\tau}$ is explored at outer iteration τ. By definition of the stopping rule in Algorithm 8, we have

$$
\Pr\left[m_\tau \geqslant \ell\right] \leqslant \Pr\left[\frac{\rho_\ell}{\ell} + \sqrt{\frac{2\ln(8T/\ell)}{\ell}} \geqslant y_\tau\right]
$$
$$
= \Pr\left[\frac{\rho_\ell}{\ell} - F(y_\tau) \geqslant \Delta_\tau - \sqrt{\frac{2\ln(8T/\ell)}{\ell}}\right]. \tag{4.65}
$$

Because ρ_ℓ is a sum of ℓ i.i.d. random variables with mean $F(y_\tau)$ and values in $[0,1]$ almost surely, applying Hoeffding's inequality (Lemma 89) we have

$$
\Pr\left[m_\tau \geqslant \ell\right] \leqslant \exp\left\{-2\left(\sqrt{\ell}\Delta_\tau - \sqrt{2\ln(8T/\ell)}\right)^2\right\}
$$
$$
\lesssim \begin{cases} 1, & \text{if } \Delta_\tau \leqslant \sqrt{8\ln(8T/\ell)/\ell}; \\ \exp\{-\ell\Delta_\tau^2/2\}, & \text{otherwise.} \end{cases}
$$

Subsequently,

$$
\textit{Regret by exploring } \mathcal{L}_{y_\tau}: \quad \leqslant \sum_{\ell=1}^{n_\tau} \ell\Delta_\tau \Pr[m_\tau = \ell] \leqslant \sum_{\ell=1}^{n_\tau} \Delta_\tau \Pr[m_\tau \geqslant \ell]
$$

$$
\lesssim \sum_{\ell=1}^{\ell_0-1} \sqrt{\frac{\ln(T/\ell)}{\ell}} + \sum_{\ell=\ell_0}^{n_\tau} \Delta_\tau \exp\{-\ell\Delta_\tau^2/2\} \tag{4.66}
$$

$$
\lesssim \sqrt{\ell_0\ln(T/\ell_0)} + \sup_{\Delta > \sqrt{8\ln(8T/\ell_0)/\ell_0}} \Delta \cdot \sum_{\ell=\ell_0}^{\infty} \exp\{-\ell\Delta^2/2\} \tag{4.67}
$$

$$
\leqslant \sqrt{\ell_0\ln(T/\ell_0)} + \sup_{\Delta > \sqrt{8\ln(8T/\ell_0)/\ell_0}} \frac{\Delta \exp\{-\ell_0\Delta^2/2\}}{1 - \exp\{-\Delta^2/2\}}
$$

$$
\leqslant \sqrt{\ell_0\ln(T/\ell_0)} + \sup_{\Delta > \sqrt{8\ln(8T/\ell_0)/\ell_0}} \frac{\Delta \exp\{-\Delta^2/2\}}{1 - \exp\{-\Delta^2/2\}}
$$

$$
\leqslant \sqrt{\ell_0\ln(T/\ell_0)} + \sqrt{\frac{8\ln(8T/\ell_0)}{\ell_0}} \cdot \frac{1}{1 - \exp\{-4\ln(8T/\ell_0)\}} \tag{4.68}
$$

$$
\lesssim \sqrt{\ell_0\ln(T/\ell_0)}. \tag{4.69}
$$

Here in Eq. (4.66), ℓ_0 is the smallset positive integer not exceeding n_τ such that $\Delta_\tau > \sqrt{8\ln(8T/\ell_0)/\ell_0}$. (If $\Delta_\tau \leqslant \sqrt{8\ln(8T/\ell_0)/\ell_0}$ holds for all $1 \leqslant \ell_0 \leqslant n_\tau$, then the second term in Eq. (4.66) is 0 and

one can conveniently set $\ell_0 = n_\tau + 1$ in this case.); Eq. (4.67) holds because

$$\sum_{\ell=1}^{\ell_0} \sqrt{\frac{\ln(T/\ell)}{\ell}} \lesssim \sum_{j=1}^{\lceil \log_2 \ell_0 \rceil} 2^j \sqrt{\frac{\ln(T/2^j)}{2^j}} = \sum_{j=1}^{\lceil \log_2 \ell_0 \rceil} \sqrt{2^j \ln(T/2^j)} \lesssim \sqrt{\ell_0 \ln(T/\ell_0)};$$

Eq. (4.68) holds because $\Delta \mapsto \Delta e^{-\Delta^2/2}/(1 - e^{-\Delta^2/2})$ is monotonically decreasing on $\Delta > 0$. Finally, because $\ell_0 \leqslant n_\tau$ and $n_\tau \lesssim \varepsilon_\tau^{-2} \log(T\varepsilon_\tau^2) \geqslant \varepsilon_\tau^{-2}$, we have

$$\textit{Regret by exploring } \mathcal{L}_{y_\tau} \lesssim \sqrt{n_\tau \ln(T/n_\tau)} \lesssim \varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2). \tag{4.70}$$

Finally, we consider regret incurred at later outer iterations $\tau' = \tau + 1, \cdots, \tau_0$. This is done by another case analysis on the relative location of θ^* with respect to $a_{\tau+1}$ and $b_{\tau+1}$:

- $\mathcal{E}_2(\tau + 1)$: $a_{\tau+1} \leqslant \theta^* \leqslant b_{\tau+1}$: the additional regret is upper bounded by $\psi_{\tau+1}^2(\alpha_1', \beta_1')$ for some values of α_1', β_1' that are not important;

- $\mathcal{E}_1(\tau + 1)$: $\theta^* < a_{\tau+1} < b_{\tau+1}$: the additional regret is upper bounded by $\psi_{\tau+1}^1(\alpha_2', \beta_2')$ with $\beta_2' \leqslant \Delta_\tau = y_\tau - F(y_\tau)$ and the value of α_2' not important;

- $\mathcal{E}_3(\tau + 1)$: $a_{\tau+1} < b_{\tau+1} < \theta^*$: the additional regret is upper bounded by $\psi_{\tau+1}^3(\alpha_3', \beta_3')$ with $\alpha_3' \leqslant 3\varepsilon_\tau$ and the value of β_3' not important.

It remains to upper bound the probability the latter two cases above occur. $\mathcal{E}_1(\tau + 1)$ occurs if for all inner iterations $t \in \mathcal{T}(\tau)$, the exploration step fails to detect $F(y_\tau)$ below y_τ, meaning that $\frac{\rho_\ell}{\ell} + \sqrt{\frac{2\ln(8T/\ell)}{\ell}} > y_\tau$ for all $\ell \in \{1, \cdots, n_\tau\}$. Also note that because $\theta^* < a_{\tau+1} = x_\tau = y_\tau - \varepsilon_\tau$, by Lemma 53 we know that $\Delta_\tau = y_\tau - F(y_\tau) \geqslant \varepsilon_\tau$. Using Hoeffding's inequality, we have

$$\Pr[\mathcal{E}_1(\tau + 1)] \leqslant \Pr\left[\forall \ell, \frac{\rho_\ell}{\ell} - F(y_\tau) > \Delta_\tau - \sqrt{\frac{2\ln(8T/\ell)}{\ell}}\right]$$

$$\leqslant \Pr\left[\frac{\rho_{n_\tau}}{n_\tau} - F(y_\tau) > \Delta_\tau - \sqrt{\frac{2\ln(8T/\ell)}{n_\tau}}\right]$$

$$\leqslant \exp\left\{-2\left(\sqrt{n_\tau}\Delta_\tau - \sqrt{2\ln(8T/n_\tau)}\right)^2\right\}$$

$$\leqslant \exp\left\{-n_\tau\Delta_\tau^2\right\}. \tag{4.71}$$

Here Eq. (4.71) holds because $\sqrt{n_\tau}\Delta_\tau \geqslant \sqrt{n_\tau}\varepsilon_\tau \geqslant \sqrt{8\ln(8T\varepsilon_\tau^2)} \geqslant 2\sqrt{2\ln(8T/n_\tau)}$ by the choice of n_τ.

The $\mathcal{E}_3(\tau + 1)$ event occurs if the exploration step in Algorithm 8 falsely detects $y_\tau > F(y_\tau)$ at some stage $\ell \in \{1, \cdots, n_\tau\}$, meaning that $\frac{\rho_\ell}{\ell} + \sqrt{\frac{2\ln(8T/\ell)}{\ell}} < y_\tau$. Note that because $b_{\tau+1} = y_\tau < \theta^*$, by Lemma 54, we know $F(y_\tau) \geqslant y_\tau$. By Lemma 55,

$$\Pr[\mathcal{E}_3(\tau + 1)] = \Pr\left[\exists \ell, \frac{\rho_\ell}{\ell} < y_\tau - \sqrt{\frac{2\ln(8T/\ell)}{\ell}}\right]$$

$$\leqslant \Pr\left[\exists \ell, \left|\frac{\rho_\ell}{\ell} - F(y_\tau)\right| > \sqrt{\frac{2\ln(8T/\ell)}{\ell}}\right]$$

143

$$\lesssim \frac{n_\tau}{T} \lesssim \frac{\ln(T\varepsilon_\tau^2)}{T\varepsilon_\tau^2}. \tag{4.72}$$

Combining all regret parts we complete the proof of Lemma 64.

Proof of Lemma 63 The regret for all outer iterations after τ (conditioned on $\mathcal{E}_1(\tau) : \theta^* < a_\tau < b_\tau$) consists of two parts: the regret from exploiting $\mathcal{L}_{y_{\tau'}}$ for $\tau' \geqslant \tau$, and the regret from exploring $\mathcal{L}_{a_{\tau'}}$.

For any $\tau' \in \{\tau, \tau+1, \cdots, \tau_0\}$, the expected regret from exploiting $\mathcal{L}_{y'_\tau}$ can always be upper bounded by $O(\varepsilon_{\tau'}^{-1} \log(T\varepsilon_{\tau'}^2))$ by the same analysis in the proof of Lemma 64 (more specifically the array of inequalities leading to Eqs. (4.67) and (4.69)), regardless of the values of α and β. This corresponds to the $\sum_{\tau'=\tau}^{\tau_0} O(\varepsilon_{\tau'}^{-1} \log(T\varepsilon_{\tau'}^2))$ term in Lemma 63.

We next upper bound the expected regret incurred by exploring $\mathcal{L}_{a_{\tau'}}$ for all $\tau' = \tau, \tau + 1, \cdots, \tau_0$. Because $a_\tau - F(a_\tau) = \beta$ by the definition of $\psi_\tau^1(\alpha, \beta)$, the expected regret incurred by exploring $\mathcal{L}_{a_{\tau'}}$, $\tau' \in \{\tau, \tau+1, \cdots, \tau_0\}$ is at most βT *assuming* $a_\tau = a_{\tau+1} = \cdots = a_{\tau_0}$. It then remains to bound the additional regret incurred by the movements of $a_{\tau'}$ in subsequent outer iterations.

Let $\mathcal{W} = \{\tau_1', \tau_2', \cdots, \tau_\ell'\}$ be outer iterations at which the update rule $a_{\tau+1} \leftarrow x_\tau$ is applied. We then have the following observations:

1. Each $\tau' \in \mathcal{W}$ would incur an additional regret upper bounded by $\Delta_{\tau'} T$, where $\Delta_{\tau'} = y_{\tau'} - F(y_{\tau'}) \geqslant \varepsilon_{\tau'}$;

2. For each $\tau' \in \{\tau, \tau+1, \cdots, \tau_0\}$, the probability update $a_{\tau'+1} \leftarrow x_{\tau'}$ is applied is at most $\exp\{-n_{\tau'}\Delta_{\tau'}\}$, using the same analysis in the proof of Lemma 64 (more specifically the array of inequalities leading to Eq. (4.71)).

Summarizing the above observations, by the law of total expectation the expected regret from exploring $\mathcal{L}_{a_{\tau'}}$ at subsequent iterations $\tau' \geqslant \tau$ can be upper bounded by $\beta T + \sum_{\tau'=\tau}^{\tau_0} \sup_{\Delta > \varepsilon_\tau} \Delta T \exp\{-n_\tau \Delta^2\}$.

Proof of Lemma 65 Because $a_\tau = \theta^* - \alpha < \theta^*$, by Lemma 54 we have $F(a_\tau) \geqslant a_\tau = \theta^* - \alpha = F(\theta^*) - \alpha$. Subsequently, $F(S^*) - F(a_\tau) \leqslant \alpha$ thanks to Lemma 52. Also note that conditioned on $\mathcal{E}_3(\tau)$, the revenue levels explored or exploited at each time epoch $t \in \mathcal{T}(\tau')$, $\tau \leqslant \tau' \leqslant \tau_0$ are sandwiched between a_τ and θ^*, and therefore $R(S^*) - R(S_t) \leqslant \alpha$. Hence, $\psi_\tau^3(\alpha, \beta) \leqslant \alpha \cdot \mathbb{E} \sum_{\tau'=\tau}^{\tau_0} |\mathcal{T}(\tau')| \leqslant \alpha T$.

4.5.7 Proof of Theorem 15

We first describe the underlying parameter values on which our lower bound proof is built. Fix revenue parameters $\{r_i\}_{i=1}^N$ as $r_1 = 1$, $r_2 = 1/2$ and $r_3 = \cdots = r_N = 0$, which are known a priori. We then consider two constructions of the unknown utility parameters $\{v_i\}_{i=1}^N$:

$$P_0: \quad v_1 = 1 - 1/4\sqrt{T}, \ v_2 = 1, \ v_3 = \cdots = v_N = 0;$$
$$P_1: \quad v_1 = 1 + 1/4\sqrt{T}, \ v_2 = 1, \ v_3 = \cdots = v_N = 0.$$

We note that P_0 and P_1 also give the probability distributions that characterize the customer random purchasing actions; and thus we will use $P_j[A]$ to denote the probability of event A under the utility parameters specified by P_j for $j \in \{0, 1\}$.

The first lemma shows that there does not exist estimators that can identify P_0 from P_1 with high probability with only T observations of random purchasing actions. Its proof involves careful calculation of the Kullback-Leibler (KL) divergence between the two hypothesized distributions and subsequent application of Le Cam's lemma to the testing question between P_0 and P_1.

Lemma 66. *For any estimator $\widehat{\psi} \in \{0,1\}$ whose inputs are T random purchasing actions $i_1, \cdots, i_T$, it holds that $\max_{j \in \{0,1\}} P_j[\widehat{\psi} \neq j] \geqslant 1/3$.*

On the other hand, the following lemma shows that, if the policy π can achieve a small regret under both P_0 and P_1, then one can construct an estimator based on π such that with large probability the estimator can distinguish between P_0 and P_1 from observed customers' purchasing actions.

Lemma 67. *Suppose a policy π satisfies $\mathrm{Regret}(\{S_t\}_{t=1}^T) < \sqrt{T}/384$ for both P_0 and P_1. Then there exists an estimator $\widehat{\psi} \in \{0,1\}$ such that $P_j[\widehat{\psi} \neq j] \leqslant 1/4$ for both $j = 0$ and $j = 1$.*

Lemma 67 is proved by explicitly constructing a classifier (tester) $\widehat{\psi}$ from any sequence of low regret. In particular, for any assortment sequence $\{S_t\}_{t=1}^T$, we construct $\widehat{\psi}$ as $\widehat{\psi} = 0$ if $\frac{1}{T} \sum_{t=1}^T \mathbb{I}[1 \in S_t, 2 \notin S_t] \geqslant 1/2$ and $\widehat{\psi} = 1$ otherwise. Using Markov's inequality and the construction of $\{r_i, v_i\}$, it can be shown that if $\mathrm{Regret}(\{S_t\}_{t=1}^T) > \sqrt{T}/384$ then $\widehat{\psi}$ is a good tester with small testing error. Detailed calculations and the complete proof is deferred to the supplement.

Combining Lemmas 66 and 67 we proved our lower bound result in Theorem 15.

Proof of Lemma 66 We first state a lemma that upper bounds the KL divergence under P_0 and P_1 for arbitrary assortment selections $S \in \mathbb{S}$.

Lemma 68. *For any $S \in \mathbb{S}$ let $P_0(S)$ and $P_1(S)$ be the distribution of the purchasing action under P_0 and P_1, respectively. Then $\mathrm{KL}(P_0(S) \| P_1(S)) \leqslant 1/18T$.*

Proof of Lemma 68. If $1 \notin S$ then $P_0(S) \equiv P_1(S)$ and therefore $\mathrm{KL}(P_0(S) \| P_1(S)) = 0$. In addition, because $v_i = r_i = 0$ for all $i \geqslant 3$, the items apart from 1 and 2 in S do not affect the distribution of the purchasing action under both P_0 and P_1. Therefore, it suffices to compute $\mathrm{KL}(P_0(\{1\}) \| P_1(\{1\}))$ and $\mathrm{KL}(P_0(\{1,2\}) \| P_1(\{1,2\}))$.

Before delving into detailed calculations, we first state a simple proposition bounding the KL divergence between two categorical distributions. It is simple to verify.

Proposition 18. *Let P and Q be two categorical distributions on J items, with parameters $p_1, \cdots, p_J$ and $q_1, \cdots, q_J$ respectively. Denote also $\varepsilon_j := p_j - q_j$. Then $\mathrm{KL}(P \| Q) \leqslant \sum_{j=1}^J \varepsilon_j^2 / q_j$.*

We first consider $\mathrm{KL}(P_0(\{1\}) \| P_1(\{1\}))$. By definition, $P_0(i = 1 | \{1\}) \leqslant 1/2 - 1/24\sqrt{T}$ and $P_1[i = 2 | \{2\}] \leqslant 1/2 + 1/24\sqrt{T}$. Also, $\min_{i=0,1}\{P_1(i | \{1\})\} \geqslant 1/3$. Subsequently,

$$\mathrm{KL}(P_0(\{1\}) \| P_1(\{1\})) \leqslant 2 \times \frac{1/144T}{1/3} \leqslant \frac{1}{24T} \leqslant \frac{1}{18T}. \tag{4.73}$$

We next consider $\mathrm{KL}(P_0(\{1,2\}) \| P_1(\{1,2\}))$. Note that $P_0(i = 0 | \{1,2\}) > P_1(i = 0 | \{1,2\})$, $P_0(i = 1 | \{1,2\}) < P_1(i = 1 | \{1,2\})$ and $P_0(i = 2 | \{1,2\}) > P_1(i = 2 | \{1,2\})$. Also, $P_0(i =$

$1|\{1,2\}) \leqslant 1/3 - 1/48\sqrt{T}, P_1(i = 1|\{1,2\}) \geqslant 1/3 + 1/48\sqrt{T}$ and $\min_{0 \leqslant i \leqslant 2}\{P_1(i|\{1,2\})\} \geqslant 1/4$. Subsequently,

$$\mathrm{KL}(P_0(\{1,2\})\|P_1(\{1,2\})) \leqslant 3 \times \frac{1/576T}{1/4} \leqslant \frac{1}{48T} \leqslant \frac{1}{18T}. \tag{4.74}$$

The lemma is thus proved. $\qquad\square$

We are now ready to prove Lemma 66.

Proof of Lemma 66. Denote $\|P - Q\|_{\mathsf{TV}} := 2\sup_A |P(A) - Q(A)|$ as the total variation norm between P and Q, and let $P_0^{\otimes T}, P_1^{\otimes T}$ denote the distribution of $\{i_t|S_t\}_{t=1}^T$ parameterized by P_0 and P_1. By Pinsker's inequality and the conditional independence of i_t conditioned on S_t, we have

$$\|P_0^{\otimes T} - P_1^{\otimes T}\|_{\mathsf{TV}} \leqslant \sqrt{2\mathrm{KL}(P_0^{\otimes T}\|P_1^{\otimes T})} \leqslant \sup_{S^{(1)},\cdots,S_t} \sqrt{2\prod_{t=1}^{T} \mathrm{KL}(P_0(S_t)\|P_1(S_t))} \tag{4.75}$$

$$\leqslant \sqrt{2T} \cdot \sup_{S} \sqrt{\mathrm{KL}(P_0(S)\|P_1(S))} \leqslant \sqrt{2T} \cdot \sqrt{1/18T} \leqslant 1/3. \tag{4.76}$$

Using Le Cam's inequality we have

$$\inf_{\widehat{\psi}} \max_{j=0,1} P_j\left[\widehat{\psi} \neq j\right] \geqslant \frac{1}{2}\left(1 - \|P_0^{\otimes T} - P_1^{\otimes T}\|_{\mathsf{TV}}\right) \geqslant \frac{1}{3}, \tag{4.77}$$

$\qquad\square$

Proof of Lemma 67 Denote $\wp_0 := 1/T \cdot \sum_{t=1}^T \mathbb{I}[1 \in S_t, 2 \notin S_t]$, $\wp_1 := 1/T \cdot \sum_{t=1}^T \mathbb{I}[1, 2 \in S_t]$, $\wp_2 := 1/T \cdot \sum_{t=1}^T \mathbb{I}[2 \in S_t, 1 \notin S_t]$ and $\overline{\wp} := 1/T \cdot \sum_{t=1}^T \mathbb{I}[1, 2 \notin S_t]$. Because the four events partition the entire probability space, we have $\wp_0 + \wp_1 + \wp_2 + \overline{\wp} = 1$. In addition, it is easy to verify that $S^* = \{1\}$ under P_0 and under P_1. Subsequently,

$$\frac{\mathrm{Regret}_\pi(T)}{T} \leqslant \frac{\wp_0}{12\sqrt{T}} + \frac{\wp_2 + \overline{\wp}}{24} \quad \text{under } P_0;$$

$$\frac{\mathrm{Regret}_\pi(T)}{T} \leqslant \frac{\wp_1}{48\sqrt{T}} + \frac{\wp_2 + \overline{\wp}}{6} \quad \text{under } P_1.$$

Using Markov's inequality and the fact that $\mathrm{Regret}_\pi(T) \leqslant \sqrt{T}/384$ under both P_0 and P_1, we have

$$P_0\left[\frac{\wp_0}{12\sqrt{T}} + \frac{\wp_2 + \overline{\wp}}{24} > \frac{1}{96\sqrt{T}}\right] \leqslant \frac{1}{4} \quad \text{and} \quad P_1\left[\frac{\wp_1}{48\sqrt{T}} + \frac{\wp_2 + \overline{\wp}}{6} > \frac{1}{96\sqrt{T}}\right] \leqslant \frac{1}{4}. \tag{4.78}$$

Subsequently, because $\wp_0 + \wp_1 + \wp_2 + \overline{\wp} = 1$, we know that $\wp_0 > 1/2$ with probability $\geqslant 2/3$ under P_0 and $\wp_0 < 1/2$ with probability $\geqslant 2/3$ under P_1. Define $\widehat{\psi}$ as

$$\widehat{\psi} := \begin{cases} 0 & \text{if } \wp_0 \geqslant 1/2; \\ 1 & \text{if } \wp_0 < 1/2. \end{cases} \tag{4.79}$$

The estimator $\widehat{\psi}$ then satisfies Lemma 67 by the above argument.

4.5.8 Proof of Theorem 16

Throughout the proof we set $r_1 = \cdots = r_N = 1$ and $v_1, \cdots, v_N \in \{1/K, (1 + \epsilon)/K\}$ for some parameter $\epsilon \in (0, 1/2]$ to be specified later. For any subset $S \subseteq [N]$, we use θ_S to indicate the parameterization where $v_i = (1 + \epsilon)/K$ if $i \in S$ and $v_i = 1/K$ if $i \notin S$.

For the ease of presentation, we further define some notations. We use $\mathcal{S}_K$ to denote all subsets of $[N]$ of size K; that is, $S \in \mathcal{S}_K$ implies $|S| = K$. Clearly, $|\mathcal{S}_K| = \binom{N}{K}$. We use P_S and $\mathbb{E}_S$ to denote the law and expectation under the parameterization θ_S.

The first step in our proof is to show that under problem parameter θ_{S_0} for some fixed $S_0 \in \mathcal{S}_K$, any assortment selection $\widetilde{S}_t \in \mathcal{S}_K$ that differs significantly from S_0 would incur a large one-stage regret. This is formalized in Lemma 69, which shows that, if a δ portion of items differ between S_0 and $\widetilde{S}_t$ then the assortment $\widetilde{S}_t$ incurs a one-stage regret of $\Omega(\delta\epsilon)$. This reduces the problem of lower bounding the regret of any policy to lower bounding the (expected) number of times a specific item $i \in [N]$ is offered, denoted as $\widetilde{N}_i$ in our proof.

At the second step we show, through a "neighboring argument" detailed in Eq. (4.83), the question of bounding $\mathbb{E}[\widetilde{N}_i]$ can be reduced to upper bounding the discrepancy between $\mathbb{E}_S[\widetilde{N}_i]$ and $\mathbb{E}_{S'}[\widetilde{N}_i]$ under two "neighboring" parameterizations θ_S and $\theta_{S'}$. Such an upper bound can be established by using the *Pinsker's inequality*, together with an upper bound on the Kullback-Leibler (KL) divergence between P_S and $P_{S'}$, which is stated in Lemma 70.

Finally, by appropriately setting the parameter ϵ which scales with N, T and K (more specifically, ϵ is set to $\epsilon = \min\{0.05\sqrt{N/T}, 0.5\}$), we complete the proof of Theorem 16.

The counting argument

We first prove the following lemma that bounds the regret of any assortment selection $\widetilde{S}_t \in \mathcal{S}_K$:

Lemma 69. *Fix arbitrary $S_0 \in \mathcal{S}_K$ and let v be the parameter associated with θ_{S_0}; that is, $v_i = (1 + \epsilon)/K$ for $i \in S_0$ and $v_i = 1/K$ for $i \in [N]\backslash S_0$, where $\epsilon \in (0, 1/2]$. For any $\widetilde{S}_t \in \mathcal{S}_K$, it holds that*

$$\max_{S \in \mathcal{S}_K} \{R_v(S)\} - R_v(\widetilde{S}_t) \geqslant \frac{\delta\epsilon}{9},$$

where $\delta = 1 - (|\widetilde{S}_t \cap S_0|/K)$.

Proof. By construction of v, it is clear that $\max_{S \in \mathcal{S}_K} \{R_v(S)\} = R_v(S_0) = (1 + \epsilon)/(2 + \epsilon)$. On the other hand, $R_v(\widetilde{S}_t) = (1 + (1 - \delta)\epsilon)/(2 + (1 - \delta)\epsilon)$. Subsequently,

$$\max_{S \in \mathcal{S}_k} \{R_v(S)\} - R_v(\widetilde{S}_t) = \frac{1 + \epsilon}{2 + \epsilon} - \frac{1 + (1 - \delta)\epsilon}{2 + (1 - \delta)\epsilon}$$

$$= \frac{\delta\epsilon}{(2 + \epsilon)(2 + (1 - \delta)\epsilon)} \geqslant \frac{\delta\epsilon}{9},$$

where the last inequality holds because $0 < \epsilon \leqslant 1/2$. $\qquad\square$

For each assortment selection $S_t \subseteq [N]$, $|S_t| \leqslant K$, let $\widetilde{S}_t \supseteq S_t$ be an arbitrary subset of size K that contains S_t; that is, $\widetilde{S}_t \supseteq S_t$, $\widetilde{S}_t \subseteq [N]$ and $|\widetilde{S}_t| = K$. For example, when $|S_t| = K$

one may directly set $\widetilde{S}_t = S_t$. Define $\widetilde{N}_i := \sum_{t=1}^T \mathbb{I}[i \in \widetilde{S}_t]$. Using Lemma 69 and the fact that $\{\widetilde{S}_t\}_{t=1}^T$ suffers less regret than $\{S_t\}_{t=1}^T$, we have

$$\max_{S \in \mathcal{S}_K} \mathbb{E}_S\left[\sum_{t=1}^T R_v(S) - R_v(S_t)\right] \geqslant \max_{S \in \mathcal{S}_K} \mathbb{E}_S\left[\sum_{t=1}^T R_v(S) - R_v(\widetilde{S}_t)\right]$$

$$\geqslant \frac{1}{|\mathcal{S}_K|} \sum_{S \in \mathcal{S}_K} \mathbb{E}_S\left[\sum_{t=1}^T R_v(S) - R_v(\widetilde{S}_t)\right] \tag{4.80}$$

$$\geqslant \frac{1}{|\mathcal{S}_K|} \sum_{S \in \mathcal{S}_K} \sum_{i \notin S} \mathbb{E}_S[\widetilde{N}_i] \cdot \frac{\epsilon}{9K} \tag{4.81}$$

$$= \frac{\epsilon}{9}\left(T - \frac{1}{|\mathcal{S}_K|} \sum_{S \in \mathcal{S}_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S[\widetilde{N}_i]\right). \tag{4.82}$$

Here Eq. (4.80) holds because the maximum regret is always lower bounded by the average regret (averaging over all parameterization θ_S for $S \in \mathcal{S}_K$), Eq. (4.81) follows from Lemma 69, and Eq. (4.82) holds because $\sum_{i=1}^N \mathbb{E}_S[\widetilde{N}_i] = \mathbb{E}_S\left[\sum_{i=1}^N \widetilde{N}_i\right] = TK$ for any $S \subseteq [N]$. The lower bound proof is then reduced to finding the largest ϵ such that the summation term in Eq. (4.82) is upper bounded by, say, cT for some constant $c < 1$.

Pinsker's inequality

The major challenge of bounding the summation term on the right-hand side of Eq. (4.82) is the $\sum_{i \in S} \mathbb{E}_S[\widetilde{N}_i]$ term. Ideally, we expect this term to be small (e.g., around K/N fraction of $\sum_{i=1}^N \mathbb{E}_S[\widetilde{N}_i] = KT$) because $S \in \mathcal{S}_K$ is of size K. However, a bandit assortment selection algorithm, with knowledge of S, could potentially allocate its assortment selections so that $\widetilde{N}_i$ becomes significantly larger for $i \in S$ than $i \notin S$. To overcome such difficulties, we use an analysis similar to the proof of Theorem 3.5 in (Bubeck & Cesa-Bianchi, 2012) to exploit the $\sum_{i=1}^N \mathbb{E}_S[\widetilde{N}_i] = NK$ property and Pinsker's inequality (Tsybakov, 2009) to bound the discrepancy in expectations under different parameterization.

Let $\mathcal{S}_{K-1}^{(i)} = \mathcal{S}_{K-1} \cap \{S \subseteq [N] : i \notin S\}$ be all subsets of size $K - 1$ that do not include i. Re-arranging summation order we have

$$\frac{1}{|\mathcal{S}_K|} \sum_{S \in \mathcal{S}_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S[\widetilde{N}_i] = \frac{1}{K} \sum_{i=1}^N \frac{1}{|\mathcal{S}_K|} \sum_{S \in \mathcal{S}_K, i \in S} \mathbb{E}_S[\widetilde{N}_i]$$

$$= \frac{1}{K} \sum_{i=1}^N \frac{1}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}^{(i)}} \mathbb{E}_{S' \cup \{i\}}[\widetilde{N}_i]. \tag{4.83}$$

Denote $P = P_{S'}$ and $Q = P_{S' \cup \{i\}}$. Also note that $0 \leqslant \widetilde{N}_i \leqslant T$ almost surely under both P and Q. Using Pinsker's inequality we have that

$$\left|\mathbb{E}_P[\widetilde{N}_i] - \mathbb{E}_Q[\widetilde{N}_i]\right| \leqslant \sum_{j=0}^T j \cdot \left|P[\widetilde{N}_i = j] - Q[\widetilde{N}_i = j]\right|$$

148

$$\leqslant T \cdot \sum_{j=0}^{T} \left| P[\widetilde{N}_i = j] - Q[\widetilde{N}_i = j] \right|$$

$$\leqslant T \cdot \|P - Q\|_{\mathrm{TV}} \leqslant T \cdot \sqrt{\frac{1}{2} \mathrm{KL}(P \| Q)}.$$

Here $\|P - Q\|_{\mathrm{TV}} = \sup_A |P(A) - Q(A)|$ and $\mathrm{KL}(P\|Q) = \int (\log \mathrm{d}P/\mathrm{d}Q) \mathrm{d}P$ are the total variation and the Kullback-Leibler (KL) divergence between P and Q, respectively. Subsequently,

$$\frac{1}{|\mathcal{S}_K|} \sum_{S \in \mathcal{S}_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S[\widetilde{N}_i]$$

$$\leqslant \frac{1}{K} \sum_{i=1}^{N} \frac{1}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}^{(i)}} \left(\mathbb{E}_{S'}[\widetilde{N}_i] + T\sqrt{\frac{1}{2}\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})} \right). \quad (4.84)$$

The first term on the right-hand side of Eq. (4.120) is easily bounded:

$$\frac{1}{K} \sum_{i=1}^{N} \frac{1}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}^{(i)}} \mathbb{E}_{S'}[\widetilde{N}_i] = \frac{1}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}} \frac{1}{K} \sum_{i \notin S'} \mathbb{E}_{S'}[\widetilde{N}_i]$$

$$\leqslant \frac{1}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}} \frac{1}{K} \sum_{i=1}^{N} \mathbb{E}_{S'}[\widetilde{N}_i]$$

$$= \frac{|\mathcal{S}_{K-1}|}{K|\mathcal{S}_K|} \cdot TK = \frac{\binom{N}{K-1}}{K\binom{N}{K}} \cdot TK = \frac{TK}{N-K+1} \leqslant \frac{T}{3}. \quad (4.85)$$

Here the last inequality holds because $K \leqslant N/4$ and hence $\frac{TK}{N-K+1} \leqslant \frac{TK}{3K+1} \leqslant \frac{T}{3}$. Combining all inequalities we have that

$$\max_{S \in \mathcal{S}_K} \mathbb{E}_S \left[\sum_{t=1}^{T} R_v(S_v^*) - R_v(S_t) \right]$$

$$\geqslant \frac{\epsilon}{9} \left(\frac{2T}{3} - \frac{T}{|\mathcal{S}_K|} \sum_{S' \in \mathcal{S}_{K-1}} \frac{1}{K} \sum_{i \notin S'} \sqrt{\frac{1}{2}\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})} \right). \quad (4.86)$$

It remains to bound the KL divergence between two "neighboring" parameterization $\theta_{S'}$ and $\theta_{S'\cup\{i\}}$ for all $S' \in \mathcal{S}_{K-1}$ and $i \notin S'$, which we elaborate in the next section.

KL-divergence between assortment selections

Define $N_i := \sum_{t=1}^{T} \mathbb{I}[i \in S_t]$. Note that because $S_t \subseteq \widetilde{S}_t$, we have $N_i \leqslant \widetilde{N}_i$ almost surely and hence $\sum_{i=1}^{N} \mathbb{E}_S[N_i] \leqslant \sum_{i=1}^{N} \mathbb{E}_S[\widetilde{N}_i] = TK$ for all $S \subseteq [N]$.

Lemma 70. *Suppose $\epsilon \in (0, 1/2]$. For any $S' \in \mathcal{S}_{K-1}$ and $i \notin S'$, it holds that $\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}}) \leqslant \mathbb{E}_{S'}[N_i] \cdot 63\epsilon^2/K$.*

Before proving Lemma 70 we first prove an upper bound on KL-divergence between categorical distributions.

Lemma 71. *Suppose P is a categorical distribution with parameters $p_0, \cdots, p_J$, meaning that $P(X = j) = p_j$ for $j = 0, \cdots, J$, and Q is a categorical distribution with parameters $q_0, \cdots, q_J$. Suppose also $p_j = q_j + \varepsilon_j$ for all $j = 0, \cdots, J$. Then*

$$\mathrm{KL}(P\|Q) \leqslant \sum_{j=0}^{J} \frac{\varepsilon_j^2}{q_j}.$$

Proof. We have that

$$\mathrm{KL}(P\|Q) = \sum_{j=0}^{J} (q_j + \varepsilon_j) \log \frac{q_j + \varepsilon_j}{q_j}$$

$$\overset{(a)}{\leqslant} \sum_{j=0}^{J} (q_j + \varepsilon_j) \frac{\varepsilon_j}{q_j} \overset{(b)}{=} \sum_{j=0}^{J} \frac{\varepsilon_j^2}{q_j}.$$

Here (a) holds because $\log(1 + x) \leqslant x$ for all $x > -1$ and (b) holds because $\sum_{j=0}^{J} \varepsilon_j = 0$. $\square$

We are now ready to prove Lemma 70.

Proof. It is clear that for any $S_t \subseteq [N], |S_t| \leqslant K$ such that $i \notin S_t$, we have $\mathrm{KL}(P_{S'}(\cdot|S_t)\|P_{S'\cup\{i\}}(\cdot|S_t)) = 0$. Therefore, we shall focus only on those $S_t \subseteq [N]$ with $i \in S_t$, which happens for $\mathbb{E}_{S'}[N_i]$ epochs in expectation. Define $K' := |S_t| \leqslant K$ and $J := |S_t \cap S'| \leqslant K - 1$. Re-write the probability of $i_t = j$ as $p_j = v_j/(a + J\epsilon/K)$ and $q_j = v_j/(a + (J + 1)\epsilon/K)$ under $P_{S'}$ and $P_{S'\cup\{i\}}$, respectively, where $a = 1 + K'/K \in (1, 2]$. We then have that

$$|p_0 - q_0| = \left| \frac{1}{a + J\epsilon/K} - \frac{1}{a + (J + 1)\epsilon/K} \right| \leqslant \frac{\epsilon}{K};$$

$$|p_j - q_j| \leqslant \frac{1 + \epsilon}{K} \left| \frac{1}{a + J\epsilon/K} - \frac{1}{a + (J + 1)\epsilon/K} \right| \leqslant \frac{2\epsilon}{K^2},$$
$$\text{if } 1 \leqslant j \leqslant N, j \neq i;$$

$$|p_j - q_j| \leqslant \left| \frac{1}{K} \frac{1}{a + J\epsilon/K} - \frac{1 + \epsilon}{K} \frac{1}{a + (J + 1)\epsilon/K} \right|$$
$$\leqslant \frac{\epsilon}{K} \frac{1}{a + (J + 1)\epsilon/K} + \frac{1}{K} \left| \frac{1}{a + J\epsilon/K} - \frac{1}{a + (J + 1)\epsilon/K} \right|$$
$$\leqslant \frac{\epsilon}{K} + \frac{1}{K} \cdot \frac{\epsilon}{K} \leqslant \frac{\epsilon}{K^2} + \frac{\epsilon}{K} \leqslant \frac{4\epsilon}{K}, \quad \text{if } j = i.$$

150

Note that $q_0 \geqslant 1/3$ and $q_j \geqslant 1/(3K)$ for $j \geqslant 1$, because $\epsilon \in (0, 1/2]$, $a \in (1, 2]$ and $J \leqslant K - 1$. Invoking Lemma 71 we have that

$$
\begin{aligned}
\mathrm{KL}(P_{S'}(\cdot|S_t)\|P_{S'\cup\{i\}}(\cdot|S_t)) &\leqslant \frac{3\epsilon^2}{K^2} + 3K \cdot \frac{4J\epsilon^2}{K^4} + 3K \cdot \frac{16\epsilon^2}{K^2} \\
&\leqslant \frac{3\epsilon^2}{K^2} + \frac{12\epsilon^2}{K^2} + \frac{48\epsilon^2}{K} \leqslant \frac{63\epsilon^2}{K}.
\end{aligned}
$$

$\square$

Putting everything together

Using Hölder's inequality, we have that

$$
\begin{aligned}
\frac{T}{|\mathcal{S}_K|} &\sum_{S'\in\mathcal{S}_{K-1}} \frac{1}{K} \sum_{i\notin S'} \sqrt{\frac{1}{2}\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})} \\
&\leqslant \frac{T|\mathcal{S}_{K-1}|}{K|\mathcal{S}_K|} \cdot \max_{S'\in\mathcal{S}_{K-1}} \sum_{i\notin S'} \sqrt{\frac{1}{2}\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})} \\
&= \max_{S'\in\mathcal{S}_{K-1}} \frac{T}{N-K+1} \sum_{i\notin S'} \sqrt{\frac{1}{2}\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})}.
\end{aligned}
$$

By Jensen's inequality and the concavity of the square root, we have

$$
\begin{aligned}
\frac{1}{N-K+1} &\sum_{i\notin S'} \sqrt{\frac{1}{2}\mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})} \\
&\leqslant \sqrt{\frac{1}{2(N-K+1)} \sum_{i\notin S'} \mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}})}.
\end{aligned}
$$

Invoking Lemma 70, we obtain

$$
\begin{aligned}
\frac{1}{N-K+1} \sum_{i\notin S'} \mathrm{KL}(P_{S'}\|P_{S'\cup\{i\}}) &\leqslant \frac{1}{N-K+1} \sum_{i\notin S'} \mathbb{E}_{S'}[N_i] \cdot \frac{63\epsilon^2}{K} \\
&\leqslant \frac{63\epsilon^2}{K(N-K+1)} \sum_{i=1}^{N} \mathbb{E}_{S'}[N_i] \\
&\leqslant \frac{126\epsilon^2}{NK} \cdot TK = \frac{126T\epsilon^2}{N}.
\end{aligned}
$$

Subsequently, setting $\epsilon = \min\{0.05\sqrt{N/T}, 0.5\}$ the term inside the bracket on the right-hand side of Eq. (4.86) can be lower bounded by $T/3$. The overall regret is thus lower bounded by $\epsilon T/27 \geqslant \min\{0.001\sqrt{NT}, T/54\}$. Theorem 16 is thus proved.

4.6 Proofs of results in Sec. 4.2

4.6.1 Proof of Theorem 17

The following lemma is the key step in our proof of Theorem 17, which shows that the estimates $\widehat{\phi}_{i,\theta}, \widehat{u}_{i,\theta}$ concentrate around the true values $\phi_{i,\theta}, u_{i,\theta}$.

Lemma 72. *Suppose* $T(i,\theta) \geqslant 96 \ln(2MTK)$. *With probability* $1 - T^{-1}$ *uniformly over all* $i \in [M]$, $\theta \in \mathcal{K}_i$ *and* $t \in [T]$

$$\left|\widehat{u}_{i,\theta} - u_{i,\theta}\right| \leqslant \min\left\{ U, \sqrt{\frac{96 \max(\widehat{u}_{i,\theta}, \widehat{u}_{i,\theta}^2) \ln(2MTK)}{T(i,\theta)}} + \frac{144 \ln(2MTK)}{T(i,\theta)} \right\}; \qquad (4.87)$$

$$\left|\widehat{\phi}_{i,\theta} - \phi_{i,\theta}\right| \leqslant \min\left\{ 1, \sqrt{\frac{\ln(2MTK)}{T(i,\theta)\widehat{u}_{i,\theta}}} \right\}. \qquad (4.88)$$

The following corollary is an immediate consequence of Lemma 72:

Corollary 7. *With probability* $1 - T^{-1}$, $\overline{u}_{i,\theta} \geqslant u_{i,\theta}$ *and* $\overline{\phi}_{i,\theta} \geqslant \phi_{i,\theta}$ *for all* $i \in [M]$, $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$.

Corollary 7 shows that (with high probability) $\overline{u}_{i,\theta}$ and $\overline{\phi}_{i,\theta}$ are valid upper bounds for $u_{i,\theta}$ and $\phi_{i,\theta}$. Our next corollary shows that $\overline{R}'$ is also an upper bound for R' at maximizers of $\overline{R}'$ and $\overline{R}$. Recall that $\overline{R}'(\theta) = [\sum_{i=1}^{M} \overline{\phi}_{i,\theta_i} \overline{u}_{i,\theta_i}]/[1 + \sum_{i=1}^{M} \overline{u}_{i,\theta_i}]$ and $R'(\theta) = [\sum_{i=1}^{M} \phi_{i,\theta_i} u_{i,\theta_i}]/[1 + \sum_{i=1}^{M} u_{i,\theta_i}]$.

Corollary 8. *With probability* $1 - T^{-1}$, $\overline{R}'(\widehat{\theta}) \geqslant R(\widehat{\theta})$ *and* $\overline{R}'(\theta*) \geqslant R'(\theta*)$, *where* $\widehat{\theta}, \theta* \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ *are maximizers of* $\overline{R}'$ *and* R', *respectively.*

We now return to the proof of Theorem 17. The first step is to use the classical regret decomposition for UCB-type policies ($\mathcal{A}$ denotes the success event in Corollary 8).

$$\text{Regret}(\{\widehat{\theta}^{(t)}\}_{t=1}^{T}) = \mathbb{E} \sum_{t=1}^{T} R'(\theta*) - R'(\theta^{(t)})$$

$$\leqslant \mathbb{E}\left[\sum_{t=1}^{T} R'(\theta*) - R'(\theta^{(t)}) \Big| \mathcal{A} \right] \Pr[\mathcal{A}] + O(T) \cdot \Pr[\mathcal{A}^c]$$

$$\leqslant O(1) + \mathbb{E}\left[\sum_{t=1}^{T} \overline{R}'(\theta*) - \overline{R}'(\theta^{(t)}) + \overline{R}'(\theta^{(t)}) - R'(\theta^{(t)}) \Big| \mathcal{A} \right] \qquad (4.89)$$

$$\leqslant O(1) + \mathbb{E}\left[\sum_{t=1}^{T} \overline{R}'(\theta^{(t)}) - R'(\theta^{(t)}) \Big| \mathcal{A} \right]. \qquad (4.90)$$

$$= O(1) + \mathbb{E}\left[\sum_{\tau} |\mathcal{E}_\tau| \cdot (\overline{R}'(\widehat{\theta}^{(\tau)}) - R'(\widehat{\theta}^{(\tau)})) \Big| \mathcal{A} \right]. \qquad (4.91)$$

152

Here, $\widehat{\theta}^{(\tau)}$ denotes any $\theta^{(t)}$ in the τ-th epoch $\mathcal{E}_\tau$ [4]. We also note that Eq. (4.89) holds because $\Pr[\mathcal{A}^c] \leqslant T^{-1}$ and $\overline{R}'(\theta^*) \geqslant R'(\theta^*)$, and Eq. (4.90) holds because $\overline{R}'(\theta^{(t)}) \geqslant \overline{R}'(\theta^*)$, since $\theta^{(t)}$ is the maximizer of $\overline{R}'$ at time t.

It remains to upper bound the discrepancy between $\overline{R}'(\widehat{\theta}^{(\tau)})$ and $R'(\widehat{\theta}^{(\tau)})$ at every epoch τ. This is accomplished by the following "aggregation lemma", which is proved in the online supplement.

Lemma 73. *With probability* $1 - T^{-1}$, *for all* $t \in [T]$, $i \in [M]$ *and* $\theta = (\theta_1, \cdots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$,

$$\left| \overline{R}'(\theta) - R'(\theta) \right| \leqslant \frac{1}{1 + \sum_{i=1}^M u_{i,\theta_i}} \left[\sum_{i=1}^M \frac{\overline{u}_{i,\theta_i} - u_{i,\theta_i}}{1 + u_{i,\theta_i}} + \sum_{i=1}^M u_{i,\theta_i} (\overline{\phi}_{i,\theta_i} - \phi_{i,\theta_i}) \right]. \tag{4.92}$$

Note that $\mathbb{E}|\mathcal{E}_\tau| = 1 + \sum_{i=1}^M \mathbb{E}[\widehat{n}_{i,\tau}] = 1 + \sum_{i=1}^M u_{i,\theta_i}$. Combining Lemma 73 with Eq. (4.91) we obtain

$$\text{Regret}(\{\widehat{\theta}^{(t)}\}_{t=1}^T) \leqslant O(1) + \sum_\tau \mathbb{E}\left[\sum_{i=1}^M \frac{\overline{u}_{i,\widehat{\theta}_i^{(\tau)}} - u_{i,\widehat{\theta}_i^{(\tau)}}}{1 + u_{i,\widehat{\theta}_i^{(\tau)}}} + \sum_{i=1}^M u_{i,\widehat{\theta}_i^{(\tau)}} (\overline{\phi}_{i,\widehat{\theta}_i^{(\tau)}} - \phi_{i,\widehat{\theta}_i^{(\tau)}}) \Big| \mathcal{A} \right]. \tag{4.93}$$

The following lemmas upper bound (asymptotically) the two terms in Eq. (4.93) separately.

Lemma 74. *Conditioned on event* $\mathcal{A}$, *it holds that*

$$\sum_\tau \sum_{i=1}^M \frac{\overline{u}_{i,\widehat{\theta}_i^{(\tau)}} - u_{i,\widehat{\theta}_i^{(\tau)}}}{1 + u_{i,\widehat{\theta}_i^{(\tau)}}} \lesssim \sqrt{MKT \log(MTK)} + MKU \log^2(MTK). \tag{4.94}$$

Lemma 75. *Conditioned on event* $\mathcal{A}$, *it holds that*

$$\sum_\tau \sum_{i=1}^M u_{i,\widehat{\theta}_i^{(\tau)}} (\overline{\phi}_{i,\widehat{\theta}_i^{(\tau)}} - \phi_{i,\widehat{\theta}_i^{(\tau)}}) \lesssim \sqrt{MKT \log(MTK)} + MKU \log^2(MTK). \tag{4.95}$$

Combining both lemmas and Eq. (4.93), we complete the proof of Theorem 17.

Proof of Lemma 72 We first prove the upper bound on $|\widehat{u}_{i,\theta} - u_{i,\theta}|$ for fixed $i \in [M]$ and $\theta \in \mathcal{K}_i$.

Case 1: $u_{i,\theta} \leqslant 1$. Let $\delta > 0$ be a parameter to be specified later. Applying Lemma 94, we have

$$\Pr\left[|\widehat{u}_{i,\theta} - u_{i,\theta}| > \delta u_{i,\theta}\right] \leqslant \exp\left\{ -\frac{n u_{i,\theta} \delta^2}{2(1+\delta) \times 4} \right\} + \exp\left\{ -\frac{n u_{i,\theta} \delta^2}{6 \times 4}\left(3 - \frac{2\delta u_{i,\theta}}{2}\right) \right\}$$

$$\leqslant \exp\left\{ -\frac{n u_{i,\theta} \delta^2}{16 \max(1,\delta)} \right\} + \exp\left\{ -\frac{n u_{i,\theta} \delta^2}{24}(3 - \delta u_{i,\theta}) \right\}$$

---

[4] Recall that in Algorithm 11, $\theta^{(t)}$ does not change within the same epoch $\mathcal{E}_\tau$. We write $\widehat{\theta}^{(\tau)}$ to highlight that $\widehat{\theta}^{(\tau)}$ is the maximizer of $\overline{R}'$ in the τ-th epoch (see Step 5 of Algorithm 11).

Suppose in addition that $\delta u_{i,\theta} \leqslant 2$. Then

$$\Pr\left[|\widehat{u}_{i,\theta} - u_{i,\theta}| > \delta u_{i,\theta}\right] \leqslant 2\exp\left\{-\frac{nu_{i,\theta}\min\{\delta,\delta^2\}}{24}\right\}$$

Equating the right-hand side of the above inequality with $1/MKT^2$ we have, we have

$$\delta = \max\left\{\sqrt{\frac{48\ln(2MTK)}{u_{i,\theta}T(i,\theta)}}, \frac{48\ln(2MTK)}{u_{i,\theta}T(i,\theta)}\right\}$$

and applying the union bound over all $i \in [M]$, $\theta \in \mathcal{K}_i$ and $t \in [T]$, with probability $1 - T^{-1}$

$$\left|\widehat{u}_{i,\theta} - u_{i,\theta}\right| \leqslant \delta u_{i,\theta} \leqslant \sqrt{\frac{48u_{i,\theta}\ln(2MTK)}{T(i,\theta)}} + \frac{48\ln(2MTK)}{T(i,\theta)}. \tag{4.96}$$

Note that if $T(i,\theta) \geqslant 48\ln(2MTK)$ the condition $\delta u_{i,\theta} \leqslant 2$ is met. Replacing all occurrences of $u_{i,\theta}$ in Eq. (4.96) by $\widehat{u}_{i,\theta}$ and using the fact that $\sqrt{a+b} \leqslant \sqrt{a} + \sqrt{b}$, we have

$$\left|\widehat{u}_{i,\theta} - u_{i,\theta}\right| \leqslant \sqrt{\frac{48\widehat{u}_{i,\theta}\ln(2MTK)}{T(i,\theta)}} + \frac{48\ln(2MTK)}{T(i,\theta)} + \sqrt{\frac{48|\widehat{u}_{i,\theta} - u_{i,\theta}|\ln(2MTK)}{T(i,\theta)}}$$

$$\leqslant \sqrt{\frac{48\widehat{u}_{i,\theta}\ln(2MTK)}{T(i,\theta)}} + \frac{48\ln(2MTK)}{T(i,\theta)}$$

$$+ \frac{48\ln(2MTK)\sqrt{u_{i,\theta}}}{T(i,\theta)} + \frac{48\ln(2MTK)}{T(i,\theta)}\sqrt{\frac{48\ln(2MTK)}{T(i,\theta)}}$$

$$\leqslant \sqrt{\frac{48\widehat{u}_{i,\theta}\ln(2MTK)}{T(i,\theta)}} + \frac{144\ln(2MTK)}{T(i,\theta)}. \tag{4.97}$$

Case 2: $u_{i,\theta} > 1$. Let $\delta \in (0,1]$ be a parameter to be specified later. Applying Lemma 94, we have

$$\Pr\left[|\widehat{u}_{i,\theta} - u_{i,\theta}| > \delta u_{i,\theta}\right] \leqslant \exp\left\{-\frac{T(i,\theta)u_{i,\theta}^2\delta^2}{6 \times 4u_{i,\theta}^2}\left(3 - \frac{2u_{i,\theta}\delta}{2 \times u_{i,\theta}}\right)\right\} + \exp\left\{-\frac{T(i,\theta)u_{i,\theta}^2\delta^2}{2 \times 4u_{i,\theta}^2}\right\}$$

$$\leqslant 2\exp\{-T(i,\theta)\delta^2/12\}.$$

Equating the right-hand side of the above inequality with $1/MKT^2$ we have

$$\delta = \sqrt{\frac{24\ln(2MTK)}{T(i,\theta)}}$$

154

and applying the union bound over all $i \in [M]$, $\theta \in \mathcal{K}_i$ and $t \in [T]$, with probability $1 - T^{-1}$,

$$\left|\widehat{u}_{i,\theta} - u_{i,\theta}\right| \leqslant \delta u_{i,\theta} \leqslant \sqrt{\frac{24 u_{i,\theta}^2 \ln(2MTK)}{T(i,\theta)}}. \tag{4.98}$$

Note that $\delta \leqslant 1$ holds if $T(i,\theta) \geqslant 24 \ln(2MTK)$. In addition, if $T(i,\theta) \geqslant 96 \ln(2MTK)$ we have $|\widehat{u}_{i,\theta} - u_{i,\theta}| \leqslant 0.5 u_{i,\theta}$ and hence $\widehat{u}_{i,\theta} \geqslant 0.5 u_{i,\theta}$. Subsequently, Eq. (4.98) implies

$$\left|\widehat{u}_{i,\theta} - u_{i,\theta}\right| \leqslant \sqrt{\frac{96 \widehat{u}_{i,\theta}^2 \ln(2MTK)}{T(i,\theta)}}. \tag{4.99}$$

Finally, combining Eqs. (4.97,4.99) we proved the upper bound on $|\widehat{u}_{i,\theta} - u_{i,\theta}|$.

We next prove the upper bound on $|\widehat{\phi}_{i,\theta} - \phi_{i,\theta}|$. Recall that for each $\tau \in \mathcal{T}(i,\theta)$, $\widehat{r}_{ik}$ is the sum of $\widehat{n}_{ik}$ i.i.d. random variables with mean $\phi_{i,\theta}$ and within range $[0,1]$ almost surely. Also note that $\sum_{\tau' \in \mathcal{T}(i,\theta)} \widehat{n}_{i,k} = T(i,\theta) \widehat{u}_{i,\theta}$. Applying Hoeffding's inequality (Lemma 89) we have for any $\delta > 0$ that

$$\Pr\left[\left|\widehat{\phi}_{i,\theta} - \phi_{i,\theta}\right| > \delta\right] \leqslant 2 \exp\left\{-2\delta^2 \cdot T(i,\theta) \widehat{u}_{i,\theta}\right\}.$$

Equating the right-hand side of the above inequality with $1/M(K+1)T^2$ and applying the union bound, we have with probability $1 - T^{-1}$ uniformly over $i \in [M]$, $\theta \in \mathcal{K}_i$ and $t \in [T]$ that

$$\left|\widehat{\phi}_{i,\theta} - \phi_{i,\theta}\right| \leqslant \sqrt{\frac{\ln(2MTK)}{T(i,\theta) \widehat{u}_{i,\theta}}}. \tag{4.100}$$

Proof of Corollary 8 We first prove $\overline{R}'(\widehat{\theta}) \geqslant R'(\widehat{\theta})$. By definition, $\sum_{i=1}^{M}(\overline{\phi}_{i,\widehat{\theta}_i} - \overline{R}'(\widehat{\theta})) \overline{u}_{i,\widehat{\theta}_i} = \overline{R}'(\widehat{\theta})$. In addition, because $\overline{R}'(\widehat{\theta})$ is the maximizer of $\overline{R}'$, setting $\lambda = \overline{R}'(\widehat{\theta})$ and by the second property of Lemma 58 we know that $\psi_\lambda(\widehat{\theta}) = \lambda$ and $\psi_\lambda(\theta) \leqslant \lambda$ for all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$, where $\psi_\lambda(\theta) = \sum_{i=1}^{M}(\overline{\phi}_{i,\theta_i} - \lambda) \overline{u}_{i,\theta_i}$.

We claim that $\overline{\phi}_{i,\widehat{\theta}_i} \geqslant \overline{R}'(\theta)$ whenever $\overline{u}_{i,\widehat{\theta}_i} > 0$. Assume the contrary, that $\overline{\phi}_{i,\widehat{\theta}_i} < \overline{R}'(\theta) = \lambda$ and $\overline{u}_{i,\widehat{\theta}_i} > 0$ for some $i \in [M]$. Consider $\widehat{\theta}' = (\widehat{\theta}'_1, \cdots, \widehat{\theta}'_M)$ defined as $\widehat{\theta}'_i = \infty$ and $\widehat{\theta}'_{i'} = \widehat{\theta}_{i'}$ for all $i' \neq i$. Because $\widehat{\theta}'_i = \infty$ we know that $\overline{u}_{i,\widehat{\theta}'_i} = u_{i,\widehat{\theta}'_i} = 0$. Subsequently, $\psi_\lambda(\widehat{\theta}') = \psi_\lambda(\widehat{\theta}) - (\overline{\phi}_{i,\widehat{\theta}_i} - \lambda) \overline{u}_{i,\widehat{\theta}_i} > \psi_\lambda(\widehat{\theta}) = \lambda$. This contradicts the condition that $\psi_\lambda(\theta) \leqslant \lambda$ for all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$.

Define $\psi_\lambda^0(\theta) := \sum_{i=1}^{M}(\phi_{i,\theta_i} - \lambda) u_{i,\theta_i}$, which is similar to the definition of ψ_λ except all occurrences of $\overline{\phi}_{i,\cdot}$ and $\overline{u}_{i,\cdot}$ are replaced by their true values $\phi_{i,\cdot}, u_{i,\cdot}$. Because $\overline{\phi}_{i,\widehat{\theta}_i} \geqslant \overline{R}'(\widehat{\theta})$ for all $\overline{u}_{i,\theta_i} > 0$, and $\overline{\phi}_{i,\cdot}, \overline{u}_{i,\cdot}$ are upper bounds of $\phi_{i,\cdot}, u_{i,\cdot}$, we conclude that $\psi_\lambda^0(\widehat{\theta}) \leqslant \psi_\lambda(\widehat{\theta}) = \lambda$, implying that $\sum_{i=1}^{M}(\phi_{i,\widehat{\theta}_i} - \lambda) u_{i,\widehat{\theta}_i} \leqslant \lambda$. Re-arranging terms we have $R'(\widehat{\theta}) = [\sum_{i=1}^{M} \phi_{i,\widehat{\theta}_i} u_{i,\widehat{\theta}_i}]/[1 + \sum_{i=1}^{M} u_{i,\widehat{\theta}_i}] \leqslant \lambda = \overline{R}'(\widehat{\theta})$.

We next prove $\overline{R}'(\theta^*) \geqslant R'(\theta^*)$. Recall that $R'(\theta^*) = [\sum_{i=1}^{M} \phi_{i,\theta_i^*} u_{i,\theta_i^*}]/[1 + \sum_{i=1}^{M} u_{i,\theta_i^*}]$. Hence, $\psi_\lambda^0(\theta^*) = \lambda$ for $\lambda = R'(\theta^*)$, meaning that $\sum_{i=1}^{M}(\phi_{i,\theta_i^*} - \lambda) u_{i,\theta_i^*} = \lambda$. By similar analysis, we know that $\phi_{i,\theta_i^*} \geqslant \lambda$ for all $u_{i,\theta_i^*} > 0$ too. Because $\overline{\phi}_{i,\cdot}, \overline{u}_{i,\cdot}$ are upper bounds of $\phi_{i,\cdot}, u_{i,\cdot}$ and

$\overline{u}_{i,\theta_i^*} = 0$ if $u_{i,\theta_i^*} = 0$, we have $\psi_\lambda(\theta^*) = \sum_{i=1}^{M}(\overline{\phi}_{i,\theta_i^*} - \lambda)\overline{u}_{i,\theta_i^*} \geqslant \sum_{i=1}^{M}(\phi_{i,\theta_i^*} - \lambda)u_{i,\theta_i^*} = \psi_\lambda^0(\theta^*) = \lambda$. This implies that $\sum_{i=1}^{M}(\overline{\phi}_{i,\theta_i^*} - R'(\theta^*))\overline{u}_{i,\theta_i^*} \geqslant R'(\theta^*)$, and therefore $\overline{R}'(\theta^*) = [\sum_{i=1}^{M}\overline{\phi}_{i,\theta_i^*}\overline{u}_{i,\theta_i^*}]/[1 + \sum_{i=1}^{M}\overline{u}_{i,\theta_i^*}] \geqslant R'(\theta^*)$.

Proof of Lemma 73 To simplify notations, we shall abbreviation $\phi_i = \phi_{i,\theta_i}, u_i = u_{i,\theta_i}$ and $\overline{\phi}_i = \overline{\phi}_{i,\theta_i}, \overline{u}_i = \overline{u}_{i,\theta_i}$. We also abbreviate $R' = R'(\theta)$ and $\overline{R}' = \overline{R}'(\theta)$.

By definition of R' and $\overline{R}'$, we have

$$
\begin{aligned}
\left(1 + \sum_{i=1}^{M} u_i\right)[\overline{R}' - R'] &= \left(1 + \sum_{i=1}^{M} u_i\right)\left[\frac{\sum_{i=1}^{M}\overline{u}_i\overline{\phi}_i}{1 + \sum_{i=1}^{M}\overline{u}_i} - \frac{\sum_{i=1}^{M}u_i\phi_i}{1 + \sum_{i=1}^{M}u_i}\right] \\
&= \sum_{i=1}^{M}\overline{\phi}_i\left(\overline{u}_i\frac{1 + \sum_{i'=1}^{M}u_{i'}}{1 + \sum_{i'=1}^{M}\overline{u}_{i'}} - u_i\right) + \sum_{i=1}^{M}u_i(\overline{\phi}_i - \phi_i) \\
&\leqslant \sum_{i=1}^{M}\left(\overline{u}_i\frac{1 + \sum_{i'=1}^{M}u_{i'}}{1 + \sum_{i'=1}^{M}\overline{u}_{i'}} - u_i\right) + \sum_{i=1}^{M}u_i(\overline{\phi}_i - \phi_i)
\end{aligned}
\tag{4.101}
$$

The first term on the right-hand side of Eq. (4.101) can be further upper bounded by

$$
\sum_{i=1}^{M}\overline{u}_i\frac{1 + \sum_{i'=1}^{M}u_{i'}}{1 + \sum_{i'=1}^{M}\overline{u}_{i'}} - u_i = \frac{\sum_{i=1}^{M}\overline{u}_i(1 + \sum_{i'=1}^{M}u_{i'}) - \sum_{i=1}^{M}u_i(1 + \sum_{i'=1}^{M}u_{i'})}{1 + \sum_{i=1}^{M}\overline{u}_i} \leqslant \sum_{i=1}^{M}\frac{\overline{u}_i - u_i}{1 + u_i}.
\tag{4.102}
$$

Here the last inequalities holds because the $\sum_{i,i'=1}^{M}u_i\overline{u}_{i'}$ term cancels out, and $1 + \sum_{i'=1}^{M}\overline{u}_{i'} \geqslant 1 + \overline{u}_i \geqslant 1 + u_i$.

Proof of Lemma 74 Define $T_\tau(i, \theta)$ as the size of $\mathcal{T}(i, \theta)$ after epoch τ, and $T_0(i, \theta)$ as the final size of $\mathcal{T}(i, \theta)$ when Algorithm 11 terminates (i.e., the total number of epochs in which singleton $\theta \in \mathcal{K}_i$ was offered in nest i). Define also $\widehat{u}_{i,\theta}^{(\tau)}$ to be the estimate of $u_{i,\widehat{\theta}_i^{(\tau)}}$ at epoch τ. Using Lemma 72, the expectation of the first term in Eq. (4.93) can be upper bounded by,

$$
\begin{aligned}
\sum_{\tau}\sum_{i=1}^{M}U \cdot &\mathbb{1}\{T_\tau(i, \widehat{\theta}_i^{(\tau)}) < 96\ln(2MTK)\} + \left[\sqrt{\frac{96\max(\widehat{u}_{i,\theta}^{(\tau)}, [\widehat{u}_{i,\theta}^{(\tau)}]^2)\ln(2MTK)}{T_\tau(i, \widehat{\theta}_i^{(\tau)}) \cdot (1 + u_{i,\theta}^{(\tau)})^2}}\right. \\
&\left. + \frac{144\ln(2MTK)}{T_\tau(i, \widehat{\theta}_i^{(\tau)})}\right] \cdot \mathbb{1}\{T_\tau(i, \widehat{\theta}_i^{(\tau)}) \geqslant 96\ln(2MTK)\} \\
= \sum_{i=1}^{M}\sum_{\theta \in \mathcal{K}_i}\sum_{\ell=0}^{T_0(i,\theta)}U \cdot &\mathbb{1}\{\ell < 96\ln(2MTK)\} + \left[\sqrt{\frac{96\max(\widehat{u}_{i,\theta}^{(\tau)}, [\widehat{u}_{i,\theta}^{(\tau)}]^2)\ln(2MTK)}{\ell(1 + u_{i,\theta})^2}}\right. \\
&\left. + \frac{144\ln(2MTK)}{\ell}\right]\mathbb{1}\{\ell \geqslant 96\ln(2MTK)\}
\end{aligned}
$$

156

$$\lesssim \sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} U \log(MTK) + \sqrt{u_{i,\theta} T_0(i,\theta) \log(MTK)} + \log T_0(i,\theta) \log(MTK) \qquad (4.103)$$

$$\leqslant MKU \log^2(MTK) + \sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} \sqrt{u_{i,\theta} T_0(i,\theta) \log(MTK)}. \qquad (4.104)$$

Here Eq. (4.103) holds by plugging in upper bounds on $|\widehat{u}_{i,\theta}^{(\tau)} - u_{i,\theta}|$ (Lemma 72) and noting that, $\max\{a, a^2\}/(1+a)^2 \lesssim a$, $\sum_{\ell \leqslant T_0(i,\theta)} \ell^{-1/2} \lesssim \sqrt{T_0(i,\theta)}$ and $\sum_{\ell \leqslant T_0(i,\theta)} \ell^{-1} \lesssim \log T_0(i,\theta)$ Eq. (4.104) holds by replacing $\log T_0$ with $\log(MTK)$.

Applying Cauchy-Schwartz inequality and the fact that $\mathbb{E}[\widehat{n}_{i,\tau}] = u_{i,\theta_i}$, $\mathbb{E}|\mathcal{E}_\tau| = 1 + \sum_{i=1}^{M} \mathbb{E}[\widehat{n}_{i,\tau}]$, the summation term in (4.104) can be further bounded by

$$\sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} \sqrt{u_{i,\theta} T_0(i,\theta) \log(MTK)} \leqslant \sqrt{M|\mathcal{K}|} \cdot \sqrt{\sum_{i=1}^{N} \sum_{\theta \in \mathcal{K}_i} u_{i,\theta} T_0(i,\theta) \log(MTK)}$$

$$= \sqrt{M|\mathcal{K}|} \cdot \sqrt{\sum_{i=1}^{N} \sum_{\theta \in \mathcal{K}_i} \sum_{\tau \in \mathcal{T}(i,\theta)} \mathbb{E}[\widehat{n}_{i,\tau}] \log(MTK)}$$

$$\leqslant \sqrt{M|\mathcal{K}|} \cdot \sqrt{\sum_{\tau} \mathbb{E}[|\mathcal{E}_\tau|] \log(MTK)}$$

$$= \sqrt{MKT \log(MTK)}.$$

Subsequently,

$$\sum_{\tau} \sum_{i=1}^{M} \frac{\overline{u}_{i,\widehat{\theta}_i^{(\tau)}} - u_{i,\widehat{\theta}_i^{(\tau)}}}{1 + u_{i,\widehat{\theta}_i^{(\tau)}}} \lesssim \sqrt{MKT \log(MTK)} + MKU \log^2(MTK).$$

Proof of Lemma 75 We first state the following result is a corollary of Lemma 72 which gives a lower bound (with high probability) on $T(i,\theta)\widehat{u}_{i,\theta}$ when $u_{i,\theta}$ is not too small. Its proof is given at the end of this section.

Corollary 9. *With probability* $1-T^{-1}$ *for all* $i \in [M]$, $\theta \in \mathcal{K}_i$ *such that* $u_{i,\theta} \geqslant 768 \ln(2MTK)/T(i,\theta)$ *and* $T(i,\theta) \geqslant 96 \ln(2MTK)$, *we have* $T(i,\theta)\widehat{u}_{i,\theta} \geqslant 0.5 T(i,\theta) u_{i,\theta}$.

Proof. First consider the case of $u_{i,\theta} \geqslant 1$. By Eq. (4.98) in the proof of Lemma 72, if $T(i,\theta) \geqslant 96 \ln(2MTK)$ we have $|\widehat{u}_{i,\theta} - u_{i,\theta}| \leqslant 0.5 u_{i,\theta}$ and therefore $T(i,\theta)\widehat{u}_{i,\theta} \geqslant 0.5 T(i,\theta) u_{i,\theta}$.

In the rest of the proof we consider the case of $768 \ln(2MTK)/T(i,\theta) \leqslant u_{i,\theta} \leqslant 1$. By Eq. (4.96) in the proof of Lemma 72, we have

$$T(i,\theta)\widehat{u}_{i,\theta} \geqslant T(i,\theta)u_{i,\theta} - \sqrt{48 T(i,\theta) u_{i,\theta} \ln(2MTK)} - 48 \ln(2MTK).$$

Under the condition that $u_{i,\theta} \geqslant 768 \ln(2MTK)/T(i,\theta)$, the above inequality yields $T(i,\theta)\widehat{u}_{i,\theta} \geqslant 0.5 T(i,\theta) u_{i,\theta}$. $\qquad \square$

Combining Corollary 9 with Lemma 72 and noting that $|\overline{\phi}_{i,\theta} - \phi_{i,\theta}| \leqslant 1$ always holds, the second term on the right-hand side of Eq. (4.93) can be upper bounded by

$$\sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} \sum_{\ell=0}^{T_0(i,\theta)} U \mathbf{1}\{\ell < 96 \ln(2MTK)\} + \left\lceil \frac{768 \ln(2MTK)}{\ell} \right.$$

$$\left. + u_{i,\theta} \sqrt{\frac{2 \ln(2MTK)}{T(i,\theta) u_{i,\theta}}} \right\rceil \cdot \mathbf{1}\{\ell \geqslant 96 \ln(2MTK)\}$$

$$\lesssim \sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} U \log(MKT) + \sqrt{u_{i,\theta} T_0(i,\theta) \log(MKT)} + \log T_0(i,\theta) \log(MKT).$$

$$(4.105)$$

Using similar derivation as in Eq. (4.94), we have

$$\sum_{\tau} \sum_{i=1}^{M} u_{i,\hat{\theta}_i^{(\tau)}} (\overline{\phi}_{i,\hat{\theta}_i^{(\tau)}} - \phi_{i,\hat{\theta}_i^{(\tau)}}) \lesssim \sqrt{MKT \log(MTK)} + MKU \log^2(MTK).$$

4.6.2 Proof of Lemma 59

Let $\theta^* = (\theta_1^*, \cdots, \theta_M^*) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ be the assortment that maximizes R'. Define $\widetilde{\theta}_i^* := \lfloor \theta_i^*/\delta \rfloor \cdot \delta$ for all $i \in [M]$ and $\widetilde{\theta}^* := (\widetilde{\theta}_1^*, \cdots, \widetilde{\theta}_M^*)$. It is easy to verify that $\widetilde{\theta}^* \in \widetilde{\mathcal{K}}_1^\delta \times \cdots \times \widetilde{\mathcal{K}}_M^\delta$. Therefore, it suffices to prove that $R'(\widetilde{\theta}^*) \geqslant R^* - \delta$ where $R^* = R'(\theta^*)$.

To simplify notations, abbreviate $R_i = R_i(\mathcal{L}_i(\theta_i^*))$, $V_i = V_i(\mathcal{L}_i(\theta_i^*))$, $\widetilde{R}_i = R_i(\mathcal{L}_i(\widetilde{\theta}_i^*))$ and $\widetilde{V}_i = V_i(\mathcal{L}_i(\widetilde{\theta}_i^*))$, where $R_i(\cdot)$ and $V_i(\cdot)$ are defined in Eqs. (4.11,4.13). Denote also that $x_i := \widetilde{V}_i - V_i$. By definition of R_i and $\widetilde{R}_i$, we have $R_i V_i = \sum_{r_{ij} \geqslant \theta_i^*} r_{ij} v_{ij}$ and $\widetilde{R}_i \widetilde{V}_i = \sum_{r_{ij} \geqslant \widetilde{\theta}_i^*} r_{ij} v_{ij}$. Subsequently,

$$\widetilde{R}_i \widetilde{V}_i = R_i V_i + \sum_{\theta_i^* > r_{ij} \geqslant \widetilde{\theta}_i^*} r_{ij} v_{ij} \geqslant R_i V_i + x_i(\theta_i^* - \delta). \qquad (4.106)$$

Here the last inequality holds because $|\theta_i^* - \widetilde{\theta}^*| \leqslant \delta$ and $\sum_{\theta_i^* > r_{ij} \geqslant \widetilde{\theta}_i^*} v_{ij} = \widetilde{V}_i - V_i = x_i$. Subsequently,

$$\widetilde{V}_i^{\gamma_i} \left[\widetilde{R}_i - (R^* - \delta) \right] = (V_i + x_i)^{\gamma_i} \left[\widetilde{R}_i - (R^* - \delta) \right] \qquad (4.107)$$

$$\geqslant (V_i + x_i)^{\gamma_i} \left[\frac{R_i V_i + x_i(\theta_i^* - \delta)}{V_i + x_i} - R^* + \delta \right] \qquad (4.108)$$

$$\geqslant (V_i + x_i)^{\gamma_i} \left[\frac{R_i V_i + x_i, \theta_i^*}{V_i + x_i} - R^* \right]. \qquad (4.109)$$

Here in Eq. (4.108) we apply Eq. (4.106), and Eq. (4.109) holds because $x_i/(V_i + x_i) \leqslant 1$.

Proposition 19. *For $i \in [M]$ define function $h_i(\Delta) := (V_i + \Delta)^{\gamma_i}[(R_i V_i + \Delta \theta_i^*)/(V_i + \Delta) - R^*]$. Then h_i is monotonically non-decreasing in Δ for $\Delta \geqslant 0$.*

158

Proof. Note that $h_i(\Delta) = (V_i + \Delta)^{\gamma_i - 1}(R_i V_i + \Delta \theta_i^*) - (V_i + \Delta)^{\gamma_i} R^*$. Differentiating h_i with respect to Δ we have

$$h_i'(\Delta) = (\gamma_i - 1)(V_i + \Delta)^{\gamma_i - 2}(R_i V_i + \Delta \theta_i^*) + \theta_i^*(V_i + \Delta)^{\gamma_i - 1} - \gamma_i(V_i + \Delta)^{\gamma_i - 1} R^*. \quad (4.110)$$

Using the second property of Lemma 56 that $\theta_i^* \geqslant \gamma_i R^* + (1 - \gamma_i) R_i(S_i^*)$, we have for all $\Delta \geqslant 0$ that

$$\begin{aligned} h_i'(\Delta) &\geqslant (\gamma_i - 1)(V_i + \Delta)^{\gamma_i - 2}(R_i V_i + \Delta \theta_i^*) + [\gamma_i R^* + (1 - \gamma_i) R_i] (V_i + \Delta)^{\gamma_i - 1} \\ &\quad - \gamma_i(V_i + \Delta)^{\gamma_i - 1} R^* \quad (4.111) \\ &= (1 - \gamma_i)(V_i + \Delta)^{\gamma_i - 2} [R_i(V_i + \Delta) - R_i V_i - \Delta \theta_i^*] \quad (4.112) \\ &= (1 - \gamma_i)(V_i + \Delta)^{\gamma_i - 2} \cdot (R_i - \theta_i^*) \Delta \geqslant 0. \quad (4.113) \end{aligned}$$

The proposition is then proved, because $h_i'(\Delta) \geqslant 0$ for all $\Delta \geqslant 0$. $\qquad \square$

Invoking Proposition 19, we have that for all $i \in [M]$,

$$\widetilde{V}_i^{\gamma_i} \left[\widetilde{R}_i - (R^* - \delta) \right] \geqslant (V_i + x_i)^{\gamma_i} \left[\frac{R_i V_i + x_i, \theta_i^*}{V_i + x_i} - R^* \right] \geqslant V_i^{\gamma_i} [R_i - R^*]. \quad (4.114)$$

Summing over $i \in [M]$ on both sides of the above inequality and using the definition that $R^* = (\sum_{i \in [M]} R_i V_i^{\gamma_i})/(1 + \sum_{i \in [M]} V_i^{\gamma_i})$,

$$\sum_{i \in [M]} \widetilde{V}_i^{\gamma_i} \left[\widetilde{R}_i - (R^* - \delta) \right] \geqslant \sum_{i \in [M]} R_i V_i^{\gamma_i} - \left(\sum_{i \in [M]} V_i^{\gamma_i} \right) R^* = R^* \geqslant R^* - \delta. \quad (4.115)$$

Re-organizing terms we have

$$R'(\widetilde{\theta}^*) = \frac{\sum_{i=1}^M \phi_{i,\widetilde{\theta}_i^*} u_{i,\widetilde{\theta}_i^*}}{1 + \sum_{i=1}^M u_{i,\widetilde{\theta}_i^*}} = \frac{\sum_{i \in [M]} R_i(\mathcal{L}_i(\widetilde{\theta}_i^*)) V_i(\mathcal{L}_i(\widetilde{\theta}_i^*))^{\gamma_i}}{1 + \sum_{i \in [M]} V_i(\mathcal{L}_i(\widetilde{\theta}_i^*))^{\gamma_i}} = \frac{\sum_{i \in [M]} \widetilde{R}_i \widetilde{V}_i^{\gamma_i}}{1 + \sum_{i \in [M]} \widetilde{V}_i^{\gamma_i}} \geqslant R^* - \delta,$$

which completes the proof.

4.6.3 Proof of Theorem 18

Construction of adversarial model parameters

Let $\epsilon > 0$ be a small positive parameter depending on M and T, which will be specified later. Each nest $i \in [M]$ in our construction consists of $N = 3$ items and is classified into two categories: "Type A" and "Type B", with parameter configurations detailed in Table 4.3. Note that regardless of which type of nest $i \in [M]$ is, the three items in nest i have revenue parameters $(1 + \epsilon)/M^2$, $(1 - \epsilon)/M^2$ and $1/M^2$. Hence it is impossible to decide the type of a nest without observations of customers' purchasing actions. Given the model parameters in Table 4.3, it is easy to verify that for a Type A nest, the optimal assortment is $\{1, 2\}$, while for a Type B nest, the optimal assortment is $\{1, 2, 3\}$.

The following lemma shows that any assortment S_i that does not equal $\{1, 2\}$ for Type A nests or $\{1, 2, 3\}$ for Type B nests incurs an $\Omega(\epsilon/M)$ regret. It is proved in the supplementary material.

Table 4.3: Adversarial construction of two types of nests. The revenue parameter ρ is set to $\rho = 9\sqrt{2}/(1 + \sqrt{2}) \approx 0.694774$.

	Type A Nest			Type B Nest		
	Item 1	Item 2	Item 3	Item 1	Item 2	Item 3
Revenues r_{ij}	1	0.8	ρ	1	0.8	ρ
Preferences v_{ij}	$(1+\epsilon)/M^2$	$(1-\epsilon)/M^2$	$1/M^2$	$(1-\epsilon)/M^2$	$(1+\epsilon)/M^2$	$1/M^2$

Lemma 76. *Let $U \subseteq [M]$ be the set of Type A nests, and by construction $[M]\backslash U$ are all Type B nests. For any $S = (S_1, \cdots, S_M) \in [N]^M$, define $m_U^\sharp(S) := \sum_{i \in U} 1\{S_i \neq \{1,2\}\} + \sum_{i \notin U} 1\{S_i \neq \{1,2,3\}\}$. Then there exists a numerical constant $C > 0$ such that for all S, $R(S^*) - R(S) \geqslant m_U^\sharp(S) \cdot C\epsilon/M$, where $S^* \in \arg\max_S R(S)$ is the optimal assortment combination under U.*

Reduction to average-case regret

Recall that for any policy π, we want to show a lower bound on the *worst-case* regret

$$\sup_{\{r_{ij}, v_{ij}\}} \sum_{t=1}^{T} R^* - \mathbb{E}^\pi \left[R(S^{(t)}) \right]. \tag{4.116}$$

Let $M_0 = M/4$ be an integer (because M is divisible by 4) and $\mathcal{S}_{M_0}$ be all $\binom{M}{M_0}$ subsets of $[M]$ with size M_0. Recall that in our adversarial construction, $U \subseteq [M]$ denotes the set of all Type A nests and the remaining nests $[M]\backslash U$ are Type B. The following inequalities show a reduction to average-case regret:

$$\sup_{\{r_{ij}, v_{ij}\}} \sum_{t=1}^{T} R^* - \mathbb{E}^\pi \left[R(S^{(t)}) \right] \geqslant \sup_{U \in \mathcal{S}_{M_0}} \sum_{t=1}^{T} R^* - \mathbb{E}_U^\pi \left[R(S^{(t)}) \right] \geqslant \frac{1}{|\mathcal{S}_{M_0}|} \sum_{U \in \mathcal{S}_{M_0}} \sum_{t=1}^{T} R^* - \mathbb{E}_U^\pi \left[R(S^{(t)}) \right].$$
$$\tag{4.117}$$

Here we use the $\mathbb{E}_U^\pi$ notation to emphasize that the distribution of $\{S^{(t)}\}$ (and hence the expectation) depends on both the parameter setting (uniquely determined by the set of Type A nests $U \subseteq [M]$) and the policy π itself.

For any $i \in [M]$ and $S \subseteq [N]$, denote $\mathfrak{n}_S(i) := \sum_{t=1}^{T} 1\{S_i^{(t)} = S\}$ as the random variable of the number of times assortment S is offered in nest i. Let $\mathbb{E}_U^\pi[\mathfrak{n}_S(i)]$ be the expectation of $\mathfrak{n}_S(i)$, with expectation taken under model parameters setting U (recall that U is the set of all Type A nests) and policy π. Invoking Lemma 76 and noting that $\sum_{i=1}^{M} \sum_{S \subseteq [N]} \mathbb{E}_U^\pi[\mathfrak{n}_S(i)] = MT$ for any $U \subseteq [M]$ and policy π, the right-hand side of Eq. (4.117) can be lower bounded by,

$$\frac{1}{|\mathcal{S}_{M_0}|} \sum_{U \in \mathcal{S}_{M_0}} \sum_{t=1}^{T} \mathbb{E}^\pi \left[m_U^\sharp(S^{(t)}) \cdot \frac{C\epsilon}{M} \right]$$

$$= \frac{C\epsilon}{M} \cdot \frac{1}{|\mathcal{S}_{M_0}|} \sum_{U \in \mathcal{S}_{M_0}} \left[\sum_{i \in U} \sum_{S \neq \{1,2\}} \mathbb{E}_U^\pi[\mathfrak{n}_S(i)] + \sum_{i \notin U} \sum_{S \neq \{1,2,3\}} \mathbb{E}_U^\pi[\mathfrak{n}_S(i)] \right]$$

$$= \frac{C\epsilon}{M}\left(MT - \frac{1}{|\mathcal{S}_{M_0}|}\sum_{U\in\mathcal{S}_{M_0}}\left[\sum_{i\in U}\mathbb{E}_U^\pi[\mathfrak{n}_{\{1,2\}}(i)] + \sum_{i\notin U}\mathbb{E}_U^\pi[\mathfrak{n}_{\{1,2,3\}}(i)]\right]\right)$$

$$= \frac{C\epsilon}{M}\left(MT - \frac{1}{|\mathcal{S}_{M_0}|}\sum_{U\in\mathcal{S}_{M_0}}\left[\sum_{i\in U}\mathbb{E}_U^\pi[\mathfrak{n}_{\{1,2\}}(i)] + \sum_{i\in U}\mathbb{E}_{U^c}^\pi[\mathfrak{n}_{\{1,2,3\}}(i)]\right]\right). \quad (4.118)$$

Here in the last identity we replace U in the second summation (corresponding to $\mathfrak{n}_{\{1,2,3\}}(i)$ with $U^c := [M]\backslash U$.

To obtain a tight lower bound on the right-hand side of (4.118), we restructure it by considering subsets $U' \subseteq [M]$ with size $M_0 - 1$, which facilitates the use of a KL divergence bound between the probability law with the Type A nests $U' \subseteq [M]$ and that with the Type A nests $U = U' \cup \{i\}$ (see Lemma 77 in the next subsection). More specifically, let $\mathcal{S}_{M_0-1}$ be the collection of all $\binom{M}{M_0-1}$ subsets of $[M]$ with size $M_0 - 1$, and $\mathcal{S}_{M_0-1}^{(i)} := \{U \in \mathcal{S}_{M_0-1} : i \notin U\}$ be a subset of $\mathcal{S}_{M_0-1}$ excluding the ith nest. By swapping summation orders, Eq. (4.118) can be equivalently written as

$$\frac{C\epsilon}{M}\left(MT - \frac{1}{|\mathcal{S}_{M_0}|}\sum_{i=1}^{M}\sum_{U'\in\mathcal{S}_{M_0-1}^{(i)}}\mathbb{E}_{U'\cup\{i\}}^\pi[\mathfrak{n}_{\{1,2\}}(i)] + \mathbb{E}_{(U'\cup\{i\})^c}^\pi[\mathfrak{n}_{\{1,2,3\}}(i)]\right)$$

$$= C\epsilon \cdot T - \frac{C\epsilon}{|\mathcal{S}_{M_0}|}\sum_{U'\in\mathcal{S}_{M_0-1}}\frac{1}{M}\sum_{i\notin U'}\left(\mathbb{E}_{U'\cup\{i\}}^\pi[\mathfrak{n}_{\{1,2\}}(i)] + \mathbb{E}_{(U'\cup\{i\})^c}^\pi[\mathfrak{n}_{\{1,2,3\}}(i)]\right). \quad (4.119)$$

Pinsker's inequality

We upper bound $\mathbb{E}_{U'\cup\{i\}}[\mathfrak{n}_{\{1,2\}}(i)]$ and $\mathbb{E}_{(U'\cup\{i\})^c}[\mathfrak{n}_{\{1,2,3\}}(i)]$ by comparing them with $\mathbb{E}_{U'}[\mathfrak{n}_{\{1,2\}}(i)]$ and $\mathbb{E}_{(U')^c}[\mathfrak{n}_{\{1,2,3\}}(i)]$. In particular, let P_U^π denote the probabilistic law under U and policy π. Then for any $S \subseteq [N]$,

$$\left|\mathbb{E}_U^\pi[\mathfrak{n}_S(i)] - \mathbb{E}_W^\pi[\mathfrak{n}_S(i)]\right| \leqslant \sum_{j=0}^{T} j \cdot \left|P_U^\pi[\mathfrak{n}_S(i) = j] - P_W^\pi[\mathfrak{n}_S(i) = j]\right|$$

$$\leqslant T \cdot \sum_{j=0}^{T}\left|P_U^\pi[\mathfrak{n}_S(i) = j] - P_W^\pi[\mathfrak{n}_S(i) = j]\right|$$

$$= T\|P_U^\pi - P_W^\pi\|_{\text{TV}} \leqslant T\sqrt{\frac{1}{2}\min\{\text{KL}(P_U^\pi\|P_W^\pi), \text{KL}(P_W^\pi\|P_U^\pi)\}} \quad (4.120)$$

$$\leqslant T\sqrt{\frac{T}{2}\min\{\max_S\text{KL}(P_U(\cdot|S)\|P_W(\cdot|S)), \max_S\text{KL}(P_W(\cdot|S)\|P_U(\cdot|S))\}}. \quad (4.121)$$

Here $\|P - Q\|_{\text{TV}}$ and $\text{KL}(P\|Q)$ denote the total variational distance and Kullback-Leibler divergence between two probability laws P and Q, Eq. (4.120) is known as the Pinsker's inequality (see e.g., Csiszar & Körner (2011); Tsybakov (2009)). Note that in the last term P_U and P_W do

161

not have superscript π, because conditioned on a particular assortment combination S the KL divergence no longer depends on π.

The following lemma shows that if U and W differ by only one nest, then the KL divergence between P_U and P_W is small *for all $S = (S_1, \cdots, S_M)$*.

Lemma 77. *Suppose $|U \triangle W| = 1$, where $U \triangle W = (U \backslash W) \cup (W \backslash U)$ denotes the symmetric difference between subsets $U, W \subseteq [M]$. Then there exists a constant $C' > 0$ such that for any $S = (S_1, \cdots, S_M)$, $\min\{\mathrm{KL}(P_U(\cdot|S) \| P_W(\cdot|S)), \mathrm{KL}(P_W(\cdot|S) \| P_U(\cdot|S))\}) \leqslant C' \epsilon^2 / M$.*

Invoking Lemma 77, the right-hand side of Eq. (4.121) can be further upper bounded by

$$T\sqrt{\frac{T}{2} \cdot \frac{C'\epsilon^2}{M}} \lesssim T\sqrt{T\epsilon^2/M}. \tag{4.122}$$

Invoking Eq. (4.119) and noting that $|\mathcal{S}_{M_0-1}| = \binom{M}{M_0-1} = \frac{M_0}{M-M_0+1}\binom{M}{M_0} \leqslant |\mathcal{S}_{M_0}|/3$ and $\sum_{i=1}^{M} \mathbb{E}_{U'}^{\pi}[\mathfrak{n}_S(i)] \leqslant MT$ for all $U' \subseteq [M]$ and $S \subseteq [N]$, we have

$$\frac{1}{|\mathcal{S}_{M_0}|} \sum_{U \in \mathcal{S}_{M_0}} \sum_{t=1}^{T} R^* - \mathbb{E}_U^{\pi}\left[R(S^{(t)})\right]$$

$$\geqslant C\epsilon \cdot T - \frac{C\epsilon}{|\mathcal{S}_{M_0}|} \sum_{U' \in \mathcal{S}_{M_0-1}} \frac{1}{M} \sum_{i \notin U'} \left(\mathbb{E}_{U' \cup \{i\}}^{\pi}[\mathfrak{n}_{\{1,2\}}(i)] + \mathbb{E}_{(U' \cup \{i\})^c}[\mathfrak{n}_{\{1,2,3\}}(i)]\right)$$

$$\geqslant C\epsilon \cdot T - \frac{C\epsilon}{3} \sup_{U' \in \mathcal{S}_{M_0-1}} \frac{1}{M} \sum_{i=1}^{M} \left(\mathbb{E}_{U'}^{\pi}[\mathfrak{n}_{\{1,2\}}(i)] + \left|(\mathbb{E}_{U' \cup \{i\}}^{\pi} - \mathbb{E}_{U'}^{\pi})[\mathfrak{n}_{\{1,2\}}(i)]\right|\right)$$

$$\qquad + \left(\mathbb{E}_{(U')^c}^{\pi}[\mathfrak{n}_{\{1,2,3\}}(i)] + \left|(\mathbb{E}_{(U' \cup \{i\})^c}^{\pi} - \mathbb{E}_{(U')^c}^{\pi})[\mathfrak{n}_{\{1,2,3\}}(i)]\right|\right)$$

$$\geqslant C\epsilon \cdot T - \frac{C\epsilon}{3} \left(\frac{1}{M} \cdot 2MT + \frac{1}{M} \cdot M \cdot O\left(T\sqrt{\frac{T\epsilon^2}{M}}\right)\right). \tag{4.123}$$

Setting $\epsilon = c_0\sqrt{M/T}$ for some sufficiently small positive constant $c_0 > 0$, we complete the proof of Theorem 18.

Proof of Lemma 76 For any $U \subseteq [M]$, $S \subseteq \{1, 2, 3\}$ and $S = (S_1, \cdots, S_M) \in [3]^M$, define $m_{U,S}^{\sharp}(S) := \sum_{i \in U} 1\{S_i = S\}$ and similarly $m_{U^c,S}^{\sharp}(S) := \sum_{i \notin U} 1\{S_i = S\}$. Denote also $S^* = (S_1^*, \cdots, S_M^*)$ as the optimal assortment combination, in which $S_i = \{1, 2\}$ for all $i \in U$ and $S_i = \{1, 2, 3\}$ for all $i \notin U$. Let also $R_U(\cdot), V_U(\cdot), R_{U^c}(\cdot), V_{U^c}(\cdot)$ be revenue and preference of assortment selections in nests of Type A ($R_U(\cdot)$ and $V_U(\cdot)$) or Type B ($R_{U^c}(\cdot)$ and $V_{U^c}(\cdot)$), respectively. Recall that $|U| = M/4$ and $|U^c| = 3M/4$. We then have

$$R(S^*) - R(S) = \frac{R_U(\{1,2\})V_U(\{1,2\})^{1/2} \cdot M/4 + R_{U^c}(\{1,2,3\})V_{U^c}(\{1,2,3\})^{1/2} \cdot 3M/4}{1 + V_U(\{1,2\})^{1/2} \cdot M/4 + V_{U^c}(\{1,2,3\})^{1/2} \cdot 3M/4}$$

$$\qquad - \frac{\sum_{S \subseteq \{1,2,3\}} m_{U,S}^{\sharp}(S) \cdot R_U(S)V_U(S)^{1/2} + m_{U^c,S}^{\sharp}(S) \cdot R_{U^c}(S)V_{U^c}(S)^{1/2}}{1 + \sum_{S \subseteq \{1,2,3\}} m_{U,S}^{\sharp}(S) \cdot V_U(S)^{1/2} + m_{U^c,S}^{\sharp}(S) \cdot V_{U^c}(S)^{1/2}}. \tag{4.124}$$

We next list the values of $V_U(\cdot), R_U(\cdot), V_{U^c}(\cdot)$ and $R_{U^c}(\cdot)$ under our adversarial construction, shown in Table 4.3.

$$
\begin{aligned}
S = \varnothing: \quad & V_U(S)^{1/2} = 0, \; R_U(S) = 0, \; V_{U^c}(S)^{1/2} = 0, \; R_{U^c}(S) = 0; \\
S = \{1\}: \quad & V_U(S)^{1/2} = \frac{\sqrt{1+\epsilon}}{M}, \; R_U(S) = 1, \; V_{U^c}(S)^{1/2} = \frac{\sqrt{1-\epsilon}}{M}, \; R_{U^c}(S) = 1; \\
S = \{2\}: \quad & V_U(S)^{1/2} = \frac{\sqrt{1-\epsilon}}{M}, \; R_U(S) = 0.8, \; V_{U^c}(S)^{1/2} = \frac{\sqrt{1+\epsilon}}{M}, \; R_{U^c}(S) = 0.8; \\
S = \{3\}: \quad & V_U(S)^{1/2} = \frac{1}{M}, \; R_U(S) = \rho, \; V_{U^c}(S)^{1/2} = \frac{1}{M}, \; R_{U^c}(S) = \rho; \\
S = \{1,2\}: \quad & V_U(S)^{1/2} = \frac{\sqrt{2}}{M}, \; R_U(S) = .9 + .1\epsilon, \; V_{U^c}(S)^{1/2} = \frac{\sqrt{2}}{M}, R_{U^c}(S) = .9 - .1\epsilon \\
S = \{1,3\}: \quad & V_U(S)^{1/2} = \frac{\sqrt{1+\epsilon}}{M}, \; R_U(S) = \frac{1+\rho+\epsilon}{2+\epsilon}, \\
& V_{U^c}(S)^{1/2} = \frac{\sqrt{1-\epsilon}}{M}, \; R_{U^c}(S) = \frac{1+\rho-\epsilon}{2-\epsilon}; \\
S = \{2,3\}: \quad & V_U(S)^{1/2} = \frac{\sqrt{1-\epsilon}}{M}, \; R_U(S) = \frac{.8+\rho-.8\epsilon}{2-\epsilon}, \\
& V_{U^c}(S)^{1/2} = \frac{\sqrt{1+\epsilon}}{M}, \; R_{U^c}(S) = \frac{.8+\rho+.8\epsilon}{2+\epsilon}; \\
S = \{1,2,3\}: \quad & V_U(S)^{1/2} = \frac{\sqrt{3}}{M}, \; R_U(S) = \frac{1.8+\rho+.2\epsilon}{3}, \\
& V_{U^c}(S)^{1/2} = \frac{\sqrt{3}}{M}, \; R_{U^c}(S) = \frac{1.8+\rho-.2\epsilon}{3}.
\end{aligned}
$$

Plugging the values of $V_U(\cdot), R_U(\cdot), V_{U^c}(\cdot), R_{U^c}(\cdot)$ into $R(S^*) - R(S)$, and taking $\epsilon \to 0^+$, by detailed algebraic calculations we proved the lemma.

Proof of Lemma 77 By symmetry we may assume without loss of generality that $W = U \cup \{i_0\}$ for some $i_0 \notin U$. The random variables observable are (i, j) where $i \in [M] \cup \{0\}$ indicates the nest in which a purchase is made (if no purchase is made then $i = 0$) and $j \in [N] = \{1, 2, 3\}$ is the particular item purchased in nest i (if $i = 0$ simply define $j = 0$ with probability 1). The KL divergence $\mathrm{KL}(P_U(\cdot|S) \| P_W(\cdot|S))$ can then be written as

$$
\mathrm{KL}(P_U(\cdot|S) \| P_W(\cdot|S)) = -\mathbb{E}_U\left[\log \frac{P_W(i,j|S)}{P_U(i,j|S)}\right] = -\mathbb{E}_U\left[\log \frac{P_W(i|S)}{P_U(i|S)}\right] - \mathbb{E}_U\left[\log \frac{P_W(j|i,S)}{P_U(j|i,S)}\right].
\tag{4.125}
$$

We next upper bound the first term on the right-hand side of Eq. (4.125). By the nested model, the nest-level purchase action $i \in [M] \cup \{0\}$ follows a categorical distribution of $M + 1$ categories, parameterized by probabilities $p = (p_0, \cdots, p_M)$ under U and $q = (q_0, \cdots, q_M)$ under W. By elementary algebra (see for example Lemma 3 in (Chen & Wang, 2018)), $\mathrm{KL}(p\|q)$

163

can be upper bounded as

$$\mathrm{KL}(p\|q) = -\sum_{i=0}^{M} p_i \log \frac{q_i}{p_i} \leqslant \sum_{i=0}^{M} \frac{|p_i - q_i|^2}{q_i}.$$

Note that U and W only differ in nest i_0. Using the nested model description and $\gamma_i \equiv 0.5$, it is easy to verify that $|p_i - q_i| \lesssim \epsilon/M$ for $i \in \{0, i_0\}$, $|p_i - q_i| \lesssim \epsilon/M^2$ if $i \notin \{0, i_0\}$, $q_0 \gtrsim \Omega(1)$ and $q_i \gtrsim 1/M$ for all $i \geqslant 1$. Subsequently,

$$\mathrm{KL}(p\|q) \lesssim \epsilon^2/M. \tag{4.126}$$

We proceed to upper bound the second term on the right-hand side of Eq. (4.125). Because U and W only differ in nest i_0, this term is non-zero only if $i = i_0$. Conditioned on $i = i_0$, it is easy to verify that $\mathrm{KL}(P_U(\cdot|i_0, S_{i_0})\|P_W(\cdot|i_0, S_{i_0})) \lesssim \epsilon^2$ for all $S_{i_0} \subseteq [N]$. In addition, $\max\{P_U(i_0|S), P_W(i_0|S)\} \lesssim 1/M$. Subsequently,

$$-\mathbb{E}_U\left[\log \frac{P_W(j|i, S)}{P_U(j|i, S)}\right] = P_U(i_0|S) \cdot \mathrm{KL}(P_U(\cdot|i_0, S_{i_0})\|P_W(\cdot|i_0, S_{i_0})) \lesssim \epsilon^2/M. \tag{4.127}$$

Combining Eqs. (4.126,4.127) we complete the proof of Lemma 77.

4.7 Proofs of results in Sec. 4.3

4.7.1 Proof of Theorem 17

The proof is divided into four steps. In the first step, we analyze the pilot estimator θ^* obtained from the pure exploration phase of Algorithm 13, and show as a corollary that the true model θ_0 is feasible to all subsequent local MLE formulations with high probability (see Corollary 10). In the second step, we use an ε-net argument to analyze the estimation error of the local MLE. Afterwards, we show in the third step that an upper bound on the estimation error $\widehat{\theta}_{t-1} - \theta_0$ implies an upper bound on the estimation error of the expected revenue $R_t(S)$, hence showing that $\overline{R}_t(S)$ are valid upper confidence bounds. Finally, we apply the *elliptical potential lemma*, which also plays a key role in linear stochastic bandit and its variants, to complete our proof.

Analysis of pure exploration and the pilot estimator

Our first step is to establish an upper bound on the estimation error $\|\theta^* - \theta_0\|_2$ of the pilot estimator θ^*, built using pure exploration data. It should be noted that in the pure exploration phase ($t \in \{1, \cdots, T_0\}$), the assortments $\{S_t\}_{t=1}^{T_0}$ only consist of one item. Therefore the observation model reduces to a standard *generalized linear model* with the sigmoid function $\sigma(x) = 1/(1 + e^{-x}) = e^x/(1 + e^x)$ as the link function, which is essentially a logistic regression model of observing 1 if the customer makes a purchase.

Because the choice model in the pure exploration phase reduces to a generalized linear model, we can cite existing works to upper bound the error $\|\theta^* - \theta_0\|_2$. In particular, the following lemma is cited from (Li et al., 2017b, Eq. (18)), adapted to our model and parameter settings.

Lemma 78. *With probability $1 - \delta$ it holds that*

$$\|\theta^* - \theta_0\|_2 \leqslant \frac{2}{\kappa}\sqrt{\frac{d + \log(1/\delta)}{\lambda_{\min}(V)}} \quad where \quad \kappa = \frac{1}{2e(1+\rho)} \quad and \quad V = \sum_{t=1}^{T_0} v_{t,i_t} v_{t,i_t}^\top. \qquad (4.128)$$

Proof. Because the noise in a logistic regression model is clearly centered and sub-Gaussian with parameter at most $1/4$, it only remains to check (Li et al., 2017b, Assumption 1), that $\inf_{\|x\|_2 \leqslant 1, \|\theta - \theta_0\|_2 \leqslant 1} \sigma'(x^\top \theta) \geqslant \kappa = 2e(1+\rho)$ where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid link function. Because $\sigma'(x) = \sigma(x)(1 - \sigma(x))$, we have $\sigma'(x^\top \theta) = \wp_\theta(1 - \wp_\theta) \geqslant 0.5\wp_\theta$ where $\wp_\theta = \min\{p_\theta(1), 1 - p_\theta(1)\}$ and $p_\theta(1) = \sigma(x^\top\theta) = 1/(1 + \exp\{-x^\top\theta\})$. By (A2), we know that $\wp_{\theta_0} \geqslant 1/(1 + \rho)$. Subsequently, for any $\|x\|_2 \leqslant 1$ and $\|\theta - \theta_0\|_2 \leqslant 1$, we have

$$\wp_\theta = \frac{1}{1 + \exp\{-x^\top\theta\}} = \frac{1}{1 + \exp\{-x^\top(\theta - \theta_0)\}\exp\{-x^\top\theta_0\}} \geqslant \frac{1}{e}\frac{1}{1 + \exp\{x^\top\theta_0\}} \geqslant \frac{1}{e(1+\rho)}.$$

Lemma 78 is then an immediate consequence of (Li et al., 2017b, Eq. (18)). $\qquad\square$

The following corollary immediately follows Lemma 78, by lower bounding $\lambda_{\min}(V)$ using standard matrix concentration inequalities.

Corollary 10. *There exists a universal constant $C_0 > 0$ such that for arbitrary $\tau \in (0, 1/2]$, if $T_0 \geqslant C_0 \max\{\nu^2 d\log T/\lambda_0^2, \rho^2(d + \log T)/(\tau^2\lambda_0)\}$ then with probability $1 - O(T^{-1})$, $\|\theta^* - \theta_0\|_2 \leqslant \tau$.*

Proof. Denote $\Lambda := \mathbb{E}_\mu xx^\top$ and $\widehat{\Lambda} := V/T_0 = \frac{1}{T_0}\sum_{t=1}^{T_0} x_{t,i_t}x_{t,i_t}^\top$. Clearly $\mathbb{E}\widehat{\Lambda} = \Lambda$. In addition, because $\|v_{tj}\|_2 \leqslant \nu$ almost surely, v_{tj} are sub-Gaussian random variables with parameter ν^2. By standard concentration inequalities (see, e.g., (Vershynin, 2012, Proposition 2.1)), we have with probability $1 - O(T^{-2})$ that $\|\widehat{\Lambda} - \Lambda\|_{\mathrm{op}} \lesssim \nu\sqrt{\frac{d\log T}{T_0}}$. Hence, if $T_0 \geqslant C_0\nu^2 d\log T/\lambda_0^2$ for some sufficiently large universal constant C_0, we have $\|\widehat{\Lambda} - \Lambda\|_{\mathrm{op}} \leqslant 0.5\lambda_0 = \lambda_{\min}(\Lambda)$ and therefore $\lambda_{\min}(V) = T_0\lambda_{\min}(\widehat{\Lambda}) \geqslant 0.5T_0\lambda_0$. The corollary then immediately follows Lemma 78. $\qquad\square$

The purpose of Corollary 10 is to establish a connection between the number of pure exploration iterations T_0 and the critical radius τ used in the local MLE formulation. It shows a lower bound on T_0 in order for the estimation error $\|\theta^* - \theta_0\|_2$ to be upper bounded by τ with high probability, which certifies that the true model θ_0 is also a *feasible* local estimator in our MLE-UCB policy. This is an important property for later analysis of local MLE solutions $\widehat{\theta}_{t-1}$.

Analysis of the local MLE

The following lemma upper bounds a Mahalanobis distance between $\widehat{\theta}_t$ and θ_0. For convenience, we adopt the notation that $r_{t0} = 0$ and $v_{t0} = 0$ for all t throughout this section. We also define

$$\begin{aligned} I_t(\theta) &:= \sum_{t'=1}^{t} M_{t'}(\theta), & (4.129)\\ M_{t'}(\theta) &:= -\mathbb{E}_{\theta_0, t'}[\nabla_\theta^2 \log p_{\theta, t'}(j|S_{t'})] \end{aligned}$$

$$= \mathbb{E}_{\theta_0,t'}[v_{t'j}v_{t'j}^\top] - \{\mathbb{E}_{\theta_0,t'}v_{t'j}\}\{\mathbb{E}_{\theta,t'}v_{t'j}\}^\top - \{\mathbb{E}_{\theta,t'}v_{t'j}\}\{\mathbb{E}_{\theta_0,t'}v_{t'j}\}^\top + \{\mathbb{E}_{\theta,t'}v_{t'j}\}\{\mathbb{E}_{\theta,t'}v_{t'j}\}^\top$$

where $\mathbb{E}_{\theta,t'}$ denotes the expectation evaluated under the law $j \sim p_{\theta,t'}(\cdot|S_{t'})$; that is, $p_{\theta,t'}(j|S_{t'}) = \exp\{v_{t'j}^\top\theta\}/(1 + \sum_{k \in S_{t'}} \exp\{v_{t'j}^\top\theta\})$ for $j \in S_{t'}$ and $p_{\theta,t'}(j|S_{t'}) = 0$ for $j \notin S_{t'}$.

Lemma 79. *Suppose $\tau \leqslant 1/\sqrt{8\rho\nu^2 K^2}$. Then there exists a universal constant $C > 0$ such that with probability $1 - O(T^{-1})$ the following holds uniformly over all $t = T_0, \cdots, T - 1$:*

$$(\widehat{\theta}_t - \theta_0)^\top I_t(\theta_0)(\widehat{\theta}_t - \theta_0) \leqslant C \cdot d\log(\rho\nu TK). \tag{4.130}$$

Remark 29. For $\theta = \theta_0$, the expression of $M_{t'}(\theta)$ can be simplified as $M_{t'}(\theta_0) = \mathbb{E}_{\theta_0,t'}[v_{t'j}v_{t'j}^\top] - \{\mathbb{E}_{\theta_0,t'}v_{t'j}\}\{\mathbb{E}_{\theta_0,t'}v_{t'j}\}^\top$.

We next state the proof of Lemma 79. For any $\theta \in \mathbb{R}^d$ define

$$f_{t'}(\theta) := \mathbb{E}_{\theta_0,t'}\left[\log \frac{p_{\theta,t'}(j|S_{t'})}{p_{\theta_0,t'}(j|S_{t'})}\right] = \sum_{j \in S_{t'} \cup \{0\}} p_{\theta_0,t'}(j|S_{t'}) \log \frac{p_{\theta,t'}(j|S_{t'})}{p_{\theta_0,t'}(j|S_{t'})}.$$

By simple algebra calculations, the first and second order derivatives of $f_{t'}$ with respect to θ can be computed as

$$\nabla_\theta f_{t'}(\theta) = \mathbb{E}_{\theta_0,t'}[v_{t'j}] - \mathbb{E}_{\theta,t'}[v_{t'j}]; \tag{4.131}$$

$$\nabla_\theta^2 f_{t'}(\theta) = -\mathbb{E}_{\theta_0,t'}[v_{t'j}v_{t'j}^\top] + \{\mathbb{E}_{\theta_0,t'}v_{t'j}\}\{\mathbb{E}_{\theta,t'}v_{t'j}\}^\top$$
$$+ \{\mathbb{E}_{\theta,t'}v_{t'j}\}\{\mathbb{E}_{\theta_0,t'}v_{t'j}\}^\top - \{\mathbb{E}_{\theta,t'}v_{t'j}\}\{\mathbb{E}_{\theta,t'}v_{t'j}\}^\top. \tag{4.132}$$

In the rest of the section we drop the subscript in $\nabla_\theta, \nabla_\theta^2$, and the ∇, ∇^2 notations should always be understood as with respect to θ.

Define $F_t(\theta) := \sum_{t'=1}^t f_{t'}(\theta)$. It is easy to verify that $-F_t(\theta)$ is the Kullback-Leibler divergence between the conditional distribution of $(i_1, \cdots, i_t)$ parameterized by θ and θ_0, respectively. Therefore, $F_t(\theta)$ is always non-positive. Note also that $F_t(\theta_0) = 0$, $\nabla F_t(\theta_0) = 0$, $\nabla^2 f_{t'}(\theta) = -M_{t'}(\theta)$ and $\nabla^2 F_t(\theta) \equiv -I_t(\theta)$. By Taylor expansion with Lagrangian remainder, there exists $\overline{\theta}_t = \alpha\theta_0 + (1 - \alpha)\widehat{\theta}_t$ for some $\alpha \in (0, 1)$ such that

$$F_t(\widehat{\theta}_t) = -\frac{1}{2}(\widehat{\theta}_t - \theta_0)^\top I_t(\overline{\theta}_t)(\widehat{\theta}_t - \theta_0). \tag{4.133}$$

Our next lemma shows that, if $\overline{\theta}_t$ is close to θ_0 (guaranteed by the constraint that $\|\widehat{\theta}_t - \theta^*\|_2 \leqslant \tau$), then $I_t(\overline{\theta}_t)$ can be *spectrally* lower bounded by $I_t(\theta_0)$. It is proved in the supplementary material.

Lemma 80. *Suppose $\tau \leqslant 1/\sqrt{8\rho\nu^2 K^2}$. Then $I_t(\overline{\theta}_t) \succeq \frac{1}{2}I_t(\theta_0)$ for all t.*

Proof. Because $\widehat{\theta}_t$ is a feasible solution of the local MLE, we know $\|\widehat{\theta}_t - \theta^*\|_2 \leqslant \tau$. Also by Corollary 10 we know that $\|\theta^* - \theta_0\|_2 \leqslant \tau$ with high probability. By triangle inequality and the definition of $\overline{\theta}_t$ we have that $\|\overline{\theta}_t - \theta_0\|_2 \leqslant 2\tau$.

To prove $I_t(\overline{\theta}_t) \succeq \frac{1}{2} I_t(\theta_0)$ we only need to show that $M_{t'}(\overline{\theta}_t) - M_{t'}(\theta_0) \preceq \frac{1}{2} M_{t'}(\theta_0)$ for all $1 \leqslant t' \leqslant t$. This reduces to proving

$$\{\mathbb{E}_{\overline{\theta}_t, t'} v_{t'j} - \mathbb{E}_{\theta_0, t'} v_{t'j}\}\{\mathbb{E}_{\overline{\theta}_t, t'} v_{t'j} - \mathbb{E}_{\theta_0, t'} v_{t'j}\}^\top \preceq \frac{1}{2} \mathbb{E}_{\theta_0, t'} \left[(v_{t'j} - \mathbb{E}_{\theta_0, t'} v_{t'j})(v_{t'j} - \mathbb{E}_{\theta_0, t'} v_{t'j})^\top \right]. \tag{4.134}$$

Fix arbitrary $S_{t'} \subseteq [N]$, $|S_{t'}| = J \leqslant K$ and for convenience denote $x_1, \cdots, x_J \in \mathbb{R}^d$ as the feature vectors of items in $S_{t'}$ (i.e., $\{v_{t'j}\}_{j \in S_{t'}}$). Let also $p_{\theta_0}(j)$ and $p_{\overline{\theta}_t}(j)$ be the probability of choosing action $j \in [J]$ corresponding to x_j parameterized by θ_0 or $\overline{\theta}_t$. Define $\overline{x} := \sum_{j=1}^J p_{\theta_0}(j) x_j$, $w_j := x_j - \overline{x}$ and $\delta_j := p_{\overline{\theta}_t}(j) - p_{\theta_0}(j)$. Recall also that $x_0 = 0$ and $w_0 = -\overline{x}$. Eq. (4.134) is then equivalent to

$$\left\{ \sum_{j=0}^J \delta_j w_j \right\} \left\{ \sum_{j=0}^J \delta_j w_j \right\}^\top \preceq \frac{1}{2} \sum_{j=0}^J p_{\theta_0}(j) w_j w_j^\top. \tag{4.135}$$

Let $L = \mathrm{span}\{w_j\}_{j=0}^J$ and $H \in \mathbb{R}^{L \times d}$ be a whitening matrix such that $H(\sum_j p_{\theta_0}(j) w_j w_j^\top) H^\top = I_{L \times L}$, where $I_{L \times L}$ is the identity matrix of size L. Denote $\widetilde{w}_j := H w_j$. We then have $\sum_{j=0}^J p_{\theta_0}(j) \widetilde{w}_j \widetilde{w}_j^\top = I_{L \times L}$. Eq. (4.135) is then equivalent to

$$\left\| \sum_{j=0}^J \delta_j \widetilde{w}_j \right\|_2^2 \leqslant \frac{1}{2}. \tag{4.136}$$

On the other hand, by (A2) we know that $p_{\theta_0}(j) \geqslant 1/\rho K$ for all j and therefore $\|\widetilde{w}_j\|_2 \leqslant \sqrt{\rho K}$ for all j. Subsequently, we have

$$\left\| \sum_{j=0}^J \delta_j \widetilde{w}_j \right\|_2^2 \leqslant \left(\max_j |\delta_j| \cdot \sum_{j=0}^J \|\widetilde{w}_j\|_2 \right)^2 \leqslant \max_j |\delta_j|^2 \cdot \rho K^2. \tag{4.137}$$

Recall that $\delta_i = p_{\overline{\theta}_t}(i) - p_{\theta_0}(i)$ where $p_\theta(i) = \exp\{x_i^\top \theta\}/(1 + \sum_{j \in S_{t'}} \exp\{x_i^\top \theta\})$. Simple algebra yields that $\nabla_\theta p_\theta(i) = p_\theta(i)[x_i - \mathbb{E}_\theta x_j]$, where $\mathbb{E}_\theta x_j = \sum_{j \in S_{t'}} p_\theta(j) x_j$. Using the mean-value theorem, there exists $\widetilde{\theta}_t = \widetilde{\alpha} \overline{\theta}_t + (1 - \widetilde{\alpha}) \theta_0$ for some $\widetilde{\alpha} \in (0, 1)$ such that

$$\delta_i = \langle \nabla_\theta p_{\widetilde{\theta}_t}(i), \widehat{\theta}_t - \theta_0 \rangle = p_{\widetilde{\theta}_t}(i) \langle x_i - \mathbb{E}_{\widetilde{\theta}_t} x_j, \overline{\theta}_t - \theta_0 \rangle. \tag{4.138}$$

Because $\|x_{ti}\|_2 \leqslant \nu$ almost surely for all $t \in [T]$ and $i \in [N]$, we have

$$\max_j |\delta_j|^2 \cdot \rho K^2 \leqslant 4 \cdot \max_i \|x_i\|_2^2 \cdot \|\overline{\theta}_t - \theta_0\|_2^2 \cdot \rho K^2 \leqslant 4\rho \nu^2 K^2 \cdot \tau^2. \tag{4.139}$$

The lemma is then proved by plugging in the condition on τ. $\qquad\square$

As a corollary of Lemma 80, we have

$$F_t(\widehat{\theta}_t) \leqslant -\frac{1}{4} (\widehat{\theta}_t - \theta_0)^\top I_t(\theta_0)(\widehat{\theta}_t - \theta_0). \tag{4.140}$$

On the other hand, consider the "empirical" version $\widehat{F}_t(\theta) := \sum_{t'=1}^{t} \widehat{f}_{t'}(\theta)$, where

$$\widehat{f}_{t'}(\theta) := \log \frac{p_{\theta,t'}(i_{t'}|S_{t'})}{p_{\theta_0,t'}(i_{t'}|S_{t'})}. \tag{4.141}$$

It is easy to verify that $\widehat{F}_t(\theta_0) = 0$ remains true; in addition, for any *fixed* $\theta \in \mathbb{R}^d$, $\{\widehat{F}_t(\theta)\}_t$ forms a *martingale* [5] and satisfies $\mathbb{E}\widehat{F}_t(\theta) = F_t(\theta)$ for all t. This leads to our following lemma, which upper bounds the *uniform* convergence of $\widehat{F}_t(\theta)$ towards $F_t(\theta)$ for all $\|\theta - \theta_0\| \leqslant 2\tau$.

Lemma 81. *Suppose $\tau \leqslant 1/\sqrt{8\rho^2\nu^2K^2}$. Then there exists a universal constant $C > 0$ such that with probability $1 - O(T^{-1})$ the following holds uniformly for all $t \in \{T_0 + 1, \cdots, T\}$ and $\|\theta - \theta_0\|_2 \leqslant 2\tau$:*

$$\left|\widehat{F}_t(\theta) - F_t(\theta)\right| \leqslant C\left[d\log(\rho\nu TK) + \sqrt{|F_t(\theta)|d\log(\rho\nu TK)}\right]. \tag{4.142}$$

Proof. We first consider a *fixed* $\theta \in \mathbb{R}^d$, $\|\theta - \theta_0\|_2 \leqslant 2\tau$. Define

$$\mathcal{M} := \max_{t' \leqslant t} |\widehat{f}_{t'}(\theta)| \quad \text{and} \quad \mathcal{V}^2 := \sum_{t'=1}^{t} \mathbb{E}_{j \sim \theta_0,t'} \left|\log \frac{p_{\theta,t'}(j|S_{t'})}{p_{\theta_0,t'}(j|S_{t'})}\right|^2. \tag{4.143}$$

Using an Azuma-Bernstein type inequality (see, for example, (Fan et al., 2015, Theorem A), (Freedman, 1975, Theorem (1.6))), we have

$$\left|\widehat{F}_t(\theta) - F_t(\theta)\right| \lesssim \mathcal{M}\log(1/\delta) + \sqrt{\mathcal{V}^2\log(1/\delta)} \quad \text{with probability } 1 - \delta. \tag{4.144}$$

The following lemma upper bounds $\mathcal{M}$ and $\mathcal{V}^2$ using $F_t(\theta)$ and the fact that θ is close to θ_0. It will be proved right after this proof.

Lemma 82. *If $\tau \leqslant 1/\sqrt{8\rho^2\nu^2K^2}$ then $\mathcal{M} \leqslant 1$ and $\mathcal{V}^2 \leqslant 8|F_t(\theta)|$.*

Corollary 11. *Suppose τ satisfies the condition in Lemma 82. Then for any $\|\theta - \theta_0\|_2 \leqslant 2\tau$,*

$$\left|\widehat{F}_t(\theta) - F_t(\theta)\right| \lesssim \log(1/\delta) + \sqrt{|F_t(\theta)|\log(1/\delta)} \quad \text{with probability } 1 - \delta. \tag{4.145}$$

Our next step is to construct an ϵ-net over $\{\theta \in \mathbb{R}^d : \|\theta - \theta_0\|_2 \leqslant 2\tau\}$ and apply union bound on the constructed ϵ-net. This together with a deterministic perturbation argument delivers *uniform* concentration of $\widehat{F}_t(\theta)$ towards $F_t(\theta)$.

For any $\epsilon > 0$, let $\mathcal{H}(\epsilon)$ be a finite covering of $\{\theta \in \mathbb{R}^d : \|\theta - \theta_0\|_2 \leqslant 2\tau\}$ in $\|\cdot\|_2$ up to precision ϵ. That is, $\sup_{\|\theta-\theta_0\|_2\leqslant 2\tau} \min_{\theta' \in \mathcal{H}(\epsilon)} \|\theta - \theta'\|_2 \leqslant \epsilon$. By standard covering number arguments (e.g., (van de Geer, 2000)), such a finite covering set $\mathcal{H}(\epsilon)$ exists whose size can be upper bounded by $\log|\mathcal{H}(\epsilon)| \lesssim d\log(\tau/\epsilon)$. Subsequently, by Corollary 11 and the union bound, we have with probability $1 - O(T^{-1})$ that

$$\left|\widehat{F}_t(\theta) - F_t(\theta)\right| \lesssim d\log(T/\epsilon) + \sqrt{|F_t(\theta)|d\log(T/\epsilon)} \quad \forall T_0 < t \leqslant T, \theta \in \mathcal{H}(\epsilon). \tag{4.146}$$

---

[5]$\{X_k\}_k$ forms a martingale if $\mathbb{E}[X_{k+1}|X_1, \cdots, X_k] = X_k$ for all k.

On the other hand, with probability $1 - O(T^{-1})$ such that Eq. (4.138) holds, we have for arbitrary $\|\theta - \theta'\|_2 \leqslant \epsilon$ that

$$
\begin{aligned}
\left| \widehat{F}_t(\theta) - \widehat{F}_t(\theta') \right| &\leqslant t \cdot \sup_{t' \leqslant t, j \in S_{t'} \cup \{0\}} \left| \log \frac{p_{\theta,t'}(j|S_{t'})}{p_{\theta',t'}(j|S_{t'})} \right| \\
&\leqslant t \cdot \sup_{t' \leqslant t, j \in S_{t'} \cup \{0\}} \frac{|p_{\theta,t'}(j|S_{t'}) - p_{\theta',t'}(j|S_{t'})|}{p_{\theta',t'}(j|S_{t'})} && (4.147) \\
&\leqslant 2\rho T K \cdot \sup_{t' \leqslant t, j \in S_{t'} \cup \{0\}} \left| p_{\theta,t'}(j|S_{t'}) - p_{\theta',t'}(j|S_{t'}) \right| && (4.148) \\
&\leqslant 2\rho T K \cdot \sup_{t' \leqslant t, j \in [N]} 4\|v_{t'j}\|_2^2 \cdot \|\theta - \theta'\|_2 \\
&\lesssim \rho T K \cdot \nu^2 \cdot \epsilon. && (4.149)
\end{aligned}
$$

Here Eq. (4.147) holds because $\log(1 + x) \leqslant x$; Eq. (4.148) holds because $p_{\theta',t'}(j|S_{t'}) \geqslant p_{\theta_0,t'}(j|s_{t'}) - |p_{\theta',t'}(j|S_{t'}) - p_{\theta_0,t'}(j|S_{t'})| \geqslant 1/2\rho K$ thanks to (A2) and Eq. (4.139).

Combining Eqs. (4.146,4.149) and setting $\epsilon \asymp 1/(\rho\nu^2 T K)$ we have with probability $1 - O(T^{-1})$ that

$$
\left| \widehat{F}_t(\theta) - F_t(\theta) \right| \lesssim d \log(\rho\nu T K) + \sqrt{|F_t(\theta)| d \log(\rho\nu T K)} \quad \forall T_0 < t \leqslant T, \|\theta - \theta_0\|_2 \leqslant 2\tau, \tag{4.150}
$$

which is to be demonstrated in Lemma 81.

In the rest of the proof we prove Lemma 82. We first derive an upper bound for M. By (A2), we know that $p_{\theta_0,t'}(j|S_{t'}) \geqslant 1/\rho K$ for all j. Also, Eqs. (4.138,4.139) shows that $|p_{\theta,t'}(j|S_{t'}) - p_{\theta_0,t'}(j|S_{t'})| \leqslant 4\nu^2 \cdot \tau^2$. If $\tau^2 \leqslant 1/\sqrt{8\rho\nu^2 K}$ we have $|p_{\theta,t'}(j|S_{t'}) - p_{\theta_0,t'}(j|S_{t'})| \leqslant 0.5 p_{\theta_0,t'}(j|S_{t'})$ and therefore $|\widehat{f}_{t'}(\theta)| \leqslant \log^2 2 \leqslant 1$.

We next give upper bounds on $\mathcal{V}^2$. Fix arbitrary t', and for notational simplicity let $p_j = p_{\theta_0,t'}(j|S_{t'})$ and $q_j = p_{\theta,t'}(j|S_{t'})$. Because $\log(1 + x) \leqslant x$ for all $x \in (-1, \infty)$, we have

$$
\mathbb{E}_{j \sim \theta_0, t'} \left| \log \frac{p_{\theta,t'}(j|S_{t'})}{p_{\theta_0,t'}(j|S_{t'})} \right|^2 = \sum_{j \in S_{t'} \cup \{0\}} p_j \log^2 \left(1 + \frac{q_j - p_j}{p_j} \right) \leqslant \sum_{j \in S_{t'} \cup \{0\}} \frac{(q_j - p_j)^2}{p_j}. \tag{4.151}
$$

On the other hand, by Taylor expansion we know that for any $x \in (-1, \infty)$, there exists $\overline{x} \in (0, x)$ such that $\log(1 + x) = x - x^2/2(1 + \overline{x})^2$. Subsequently,

$$
\begin{aligned}
-f_{t'}(\theta) &= -\mathbb{E}_{j \sim \theta_0, t'} \left[\log \frac{p_{\theta,t'}(j|S_{t'})}{p_{\theta_0,t'}(j|S_{t'})} \right] = - \sum_{j \in S_{t'} \cup \{0\}} p_j \log \left(1 + \frac{q_j - p_j}{p_j} \right) && (4.152) \\
&= - \sum_{j \in S_{t'} \cup \{0\}} p_j \left(\frac{q_j - p_j}{p_j} - \frac{1}{2(1 + \overline{\delta}_j)^2} \frac{|q_j - p_j|^2}{p_j^2} \right) && (4.153) \\
&\geqslant \frac{1}{2(1 + \max_j |p_j - q_j|/p_j)^2} \cdot \sum_{j \in S_{t'} \cup \{0\}} \frac{(q_j - p_j)^2}{p_j}. && (4.154)
\end{aligned}
$$

Here $\overline{\delta}_j \in (0, (q_j - p_j)/p_j)$ and the last inequality holds because $\sum_j p_j = \sum_j q_j = 1$.

By Eqs. (4.138) and (4.139), we have that $|q_j - p_j|^2 \leqslant 4\nu^2 \cdot \tau^2$. In addition, (A2) implies that $p_j \geqslant 1/\rho K$ for all j. Therefore, if $\tau \leqslant 1/\sqrt{4\rho^2\nu^2 K^2}$ we have $|p_j - q_j|/p_j \leqslant 1$ for all j and hence

$$\mathbb{E}_{j \sim \theta_0, t'} \left| \log \frac{p_{\theta, t'}(j|S_{t'})}{p_{\theta_0, t'}(j|S_{t'})} \right|^2 \leqslant \sum_{j \in S_{t'} \cup \{0\}} \frac{(q_j - p_j)^2}{p_j} \leqslant 8|f_{t'}(\theta)|. \tag{4.155}$$

Summing over all $t' = 1, \cdots, t$ and noting that $f_{t'}(\theta)$ is always non-positive, we complete the proof of Lemma 82. □

We are now ready to prove Lemma 79. By Eq. (4.142) and the fact that $\widehat{F}_t(\widehat{\theta}_t) \leqslant 0 \leqslant F_t(\widehat{\theta}_t)$, we have

$$|F_t(\widehat{\theta}_t)| \leqslant |\widehat{F}_t(\widehat{\theta}_t) - F_t(\widehat{\theta}_t)| \lesssim d \log(\rho\nu TK) + \sqrt{|F_t(\widehat{\theta}_t)| d \log(\rho\nu TK)}. \tag{4.156}$$

Subsequently,

$$|F_t(\widehat{\theta}_t)| \lesssim d \log(\rho\nu NT). \tag{4.157}$$

In addition, because $F_t(\widehat{\theta}_t) \leqslant 0$, by Eq. (4.140) we have

$$-\frac{1}{2}(\widehat{\theta}_t - \theta_0)^\top I_t(\theta_0)(\widehat{\theta}_t - \theta_0) \geqslant F_t(\widehat{\theta}_t) \geqslant d \log(\rho\nu TK). \tag{4.158}$$

Lemma 79 is thus proved.

Analysis of upper confidence bounds

The following technical lemma shows that the upper confidence bounds constructed in Algorithm 13 are valid with high probability. Additionally, we establish an upper bound on the discrepancy between $\overline{R}_t(S)$ and the true value $R_t(S)$ defined in Eq. (4.27).

Lemma 83. *Suppose τ satisfies the condition in Lemma 79. With probability $1 - O(T^{-1})$ the following holds uniformly for all $t > T_0$ and $S \subseteq [N]$, $|S| \leqslant K$ such that*

1. *$\overline{R}_t(S) \geqslant R_t(S)$;*

2. *$|\overline{R}_t(S) - R_t(S)| \lesssim \min\{1, \omega\sqrt{\|I_{t-1}^{-1/2}(\theta_0)M_t(\theta_0|S)I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}}\}$.*

At a higher level, the proof of Lemma 83 can be regarded as a "finite-sample" version of the classical *Delta's method*, which upper bounds estimation error of some functional φ of parameters, i.e., $|\varphi(\widehat{\theta}_{t-1}) - \varphi(\theta_0)|$ using the estimation error of the parameters themselves $\widehat{\theta}_{t-1} - \theta_0$.

We now state the proof of Lemma 83. Without explicit clarification, all statements are conditioned on the success event in Lemma 79, which occurs with probability $1 - O(T^{-1})$ if τ is sufficiently large and satisfies the condition in Lemma 79.

We present below a key technical lemma in the proof of Lemma 83, which is an upper bound on the absolute value difference between $R_t(S) := \mathbb{E}_{\theta_0, t}[r_{tj}|S]$ and $\widehat{R}_t(S) := \mathbb{E}_{\widehat{\theta}_{t-1}, t}[r_{tj}|S]$ using $I_{t-1}(\theta_0)$ and $M_t(\theta_0|S)$, where $I_{t-1}(\theta) = \sum_{t'=1}^{t-1} M_{t'}(\theta)$ and $M_{t'}(\theta) = \mathbb{E}_{\theta_0, t'}[v_{t'j}v_{t'j}^\top] - \{\mathbb{E}_{\theta_0, t'}v_{t'j}\}\{\mathbb{E}_{\theta, t'}v_{t'j}\}^\top - \{\mathbb{E}_{\theta, t'}v_{t'j}\}\{\mathbb{E}_{\theta_0, t'}v_{t'j}\}^\top + \{\mathbb{E}_{\theta, t'}v_{t'j}\}\{\mathbb{E}_{\theta, t'}v_{t'j}\}^\top$. This key lemma can be regarded as a finite sample version of the celebrated *Delta's method* (e.g., (Van der Vaart, 1998)) used widely in classical statistics to estimate and/or infer a functional of unknown quantities.

Lemma 84. *For all $t > T_0$ and $S \subseteq [N]$, $|S| \leq K$, it holds that $|\widehat{R}_t(S) - R_t(S)| \lesssim \sqrt{d \log(\rho \nu TK)} \cdot \sqrt{\|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S) I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}}$, where in $\lesssim$ notation we only hide numerical constants.*

Below we state our proof of Lemma 84. Fix $S \subseteq [N]$. We use $\mathfrak{R}_t(\theta) = \mathbb{E}_{\theta,t}[r_{tj}] = [\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta\}]/[1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta\}]$ to denote the expected revenue of assortment S at time t, evaluated using a specific model $\theta \in \mathbb{R}$. Then

$$
\nabla_\theta \mathfrak{R}_t(\theta) = \frac{\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta\}(1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta\})^2 - (\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta\})(\sum_{j \in S} \exp\{v_{tj}^\top \theta\})}{(1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta\})^2}
$$

$$
= \mathbb{E}_{\theta,t}[r_{tj} v_{tj}] - \{\mathbb{E}_{\theta,t} r_{tj}\}\{\mathbb{E}_{\theta,t} v_{tj}\}. \tag{4.159}
$$

By the mean value theorem, there exists $\widetilde{\theta}_{t-1} = \theta_0 + \xi(\widehat{\theta}_{t-1} - \theta_0)$ for some $\xi \in (0,1)$ such that

$$
\left|\widehat{R}_t(S) - R_t(S)\right| = \left|\mathfrak{R}_t(\widehat{\theta}_{t-1}) - \mathfrak{R}_t(\theta_0)\right| = \left|\langle \nabla \mathfrak{R}_t(\widetilde{\theta}_{t-1}), \widehat{\theta}_{t-1} - \theta_0 \rangle\right|
$$

$$
= \sqrt{(\widehat{\theta}_{t-1} - \theta_0)^\top [\nabla \mathfrak{R}_t(\widetilde{\theta}_{t-1}) \nabla \mathfrak{R}_t(\widetilde{\theta}_{t-1})^\top)](\widehat{\theta}_{t-1} - \theta_0)}. \tag{4.160}
$$

Recall that $\nabla \mathfrak{R}_t(\widetilde{\theta}_{t-1}) = \mathbb{E}_{\widetilde{\theta}_{t-1},t}[r_{tj} v_{tj}] - \{\mathbb{E}_{\widetilde{\theta}_{t-1},t} r_{tj}\}\{\mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj}\} = \mathbb{E}_{\widetilde{\theta}_{t-1},t}[(r_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} r_{tj})(v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})]$. Subsequently, by Jenson's inequality and the fact that $r_{tj} \in [0,1]$ almost surely,

$$
\nabla \mathfrak{R}_t(\widetilde{\theta}_{t-1}) \nabla \mathfrak{R}_t(\widetilde{\theta}_{t-1})^\top \preceq \mathbb{E}_{\widetilde{\theta}_{t-1},t}\left[(r_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} r_{tj})^2 (v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})(v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})^\top\right]
$$

$$
\preceq \mathbb{E}_{\widetilde{\theta}_{t-1},t}\left[(v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})(v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})^\top\right] = \widehat{M}_t(\widetilde{\theta}_{t-1}|S). \tag{4.161}
$$

Define $\widehat{M}_t(\theta|S) := \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t} v_{tj})(v_{tj} - \mathbb{E}_{\theta,t} v_{tj})^\top]$, where $S \subseteq [N]$ is the assortment supplied at iteration t. Combining Eqs. (4.160,4.161) with Lemma 79, we have

$$
\left|\widehat{R}_t(S) - R_t(S)\right| \lesssim \sqrt{d \log(\rho \nu NT)} \cdot \sqrt{\|I_{t-1}(\theta_0)^{-1/2} \widehat{M}_t(\widetilde{\theta}_{t-1}|S) I_{t-1}(\theta_0)^{-1/2}\|_{\mathrm{op}}}. \tag{4.162}
$$

It remains to show that $\widehat{M}_t(\widetilde{\theta}_{t-1}|S)$ and $M_t(\theta_0|S)$ are close, for which we first recall the definitions of both quantities:

$$
\widehat{M}_t(\widetilde{\theta}_{t-1}|S) = \mathbb{E}_{\widetilde{\theta}_{t-1},t}\left[(v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})(v_{tj} - \mathbb{E}_{\widetilde{\theta}_{t-1},t} v_{tj})^\top\right];
$$

$$
M_t(\theta_0|S) = \mathbb{E}_{\theta_0,t}[v_{tj} v_{tj}^\top] - \{\mathbb{E}_{\theta_0,t} v_{tj}\}\{\mathbb{E}_{\theta_0,t} v_{tj}\}^\top = \widehat{M}_t(\theta_0|S).
$$

The next lemma shows that under suitable conditions $\widehat{M}_t(\widetilde{\theta}_{t-1}|S)$ is close to $\widehat{M}_t(\theta_0|S) = M_t(\theta_0|S)$, implying that $\frac{1}{4} M_t(\theta_0|S) \preceq \widehat{M}_t(\widetilde{\theta}_{t-1}|S) \preceq 4 M_t(\theta_0|S)$. It is proved in the supplementary material.

Lemma 85. *Suppose $\tau \leq 1/\sqrt{8\rho^2 \nu^2 K^2}$. Then $\frac{1}{4} M_t(\theta_0|S) \preceq \widehat{M}_t(\widetilde{\theta}_{t-1}|S) \preceq 4 M_t(\theta_0|S)$ for all t, S and θ.*

Proof. Define $\overline{M}_t(\theta|S) := \mathbb{E}_{\theta_0,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top]$, where only the outermost expectation is replaced by taking with respect to the probability law under θ_0. Denote also $\widetilde{w}_j := v_{tj} - \mathbb{E}_{\theta,t}v_{tj}$. Then $\overline{M}_t(\theta|S) = \sum_j p_{\theta_0,t}(j)\widetilde{w}_j\widetilde{w}_j^\top$ and $\overline{M}_t(\theta|S) - \widehat{M}_t(\theta|S) = \sum_j \delta_j \widetilde{w}_j\widetilde{w}_j^\top$, where $\delta_j = p_{\theta_0,t}(j) - p_{\theta,t}(j)$. By Eq. (4.138) and the fact that $\|v_{ti}\|_2 \leqslant \nu$, $\|\theta - \theta_0\|_2 \leqslant \tau$, we have

$$\max_j |\delta_j| \leqslant \sqrt{4\nu^2 \cdot \tau}. \tag{4.163}$$

On the other hand, by (A2) we know that $\min_j p_{\theta_0,t}(j) \geqslant 1/\rho K$ and therefore

$$\overline{M}_t(\theta|S) = \sum_j p_{\theta_0,t}\widetilde{w}_j\widetilde{w}_j^\top \succeq \frac{1}{\rho K}\sum_j \widetilde{w}_j\widetilde{w}_j^\top. \tag{4.164}$$

Combining Eqs. (4.163,4.164) and the fact that $\overline{M}_t(\theta|S) - \widehat{M}_t(\theta|S) = \sum_j \delta_j\widetilde{w}_j\widetilde{w}_j^\top$, we have $\overline{M}_t(\theta|S) - \widehat{M}_t(\theta|S) \preceq \overline{M}_t(\theta|S)/2$ and $\widehat{M}_t(\theta|S) - \overline{M}_t(\theta|S) \preceq \overline{M}_t(\theta|S)/2$, provided that $\tau \leqslant 1/\sqrt{8\rho^2\nu^2 K^2}$. This also implies $\frac{1}{2}\overline{M}_t(\theta|S) \preceq \widehat{M}_t(\theta|S) \preceq 2\overline{M}_t(\theta|S)$.

We next prove that $\frac{1}{2}M_t(\theta_0|S) \preceq \overline{M}_t(\theta|S) \preceq 2M_t(\theta_0|S)$ which, together with $\frac{1}{2}\overline{M}_t(\theta|S) \preceq \widehat{M}_t(\theta|S) \preceq 2\overline{M}_t(\theta|S)$ established in the previous section, implies Lemma 85. Recall the definitions that

$$M_t(\theta_0|S) = \mathbb{E}_{\theta_0,t}\left[(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})^\top\right];$$
$$\overline{M}_t(\theta|S) = \mathbb{E}_{\theta_0,t}\left[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top\right].$$

Adding and subtracting $\mathbb{E}_{\theta,t}v_{tj}, \mathbb{E}_{\theta_0,t}v_{tj}$ terms, we have

$$\overline{M}_t(\theta|S) - M_t(\theta_0|S)$$
$$= \mathbb{E}_{\theta_0,t}\left[(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj} + \mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj} + \mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top\right]$$
$$\quad - \mathbb{E}_{\theta_0,t}\left[(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})^\top\right]$$
$$= \mathbb{E}_{\theta_0,t}\left[(\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})^\top\right] + \mathbb{E}_{\theta_0,t}\left[(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})(\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top\right]$$
$$\quad + (\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top$$
$$= (\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top.$$

By Eq. (4.134) in the proof of Lemma 80, we have that

$$(\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(\mathbb{E}_{\theta_0,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^\top \precsim \frac{1}{2}\mathbb{E}_{\theta_0,t}[(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta_0,t}v_{tj})^\top] = \frac{1}{2}M_t(\theta_0|S)$$

provided that $\tau \leqslant 1/\sqrt{8\rho^2\nu^2 K^2}$, thus implying $\frac{1}{2}M_t(\theta_0|S) \preceq \overline{M}_t(\theta|S) \preceq 2M_t(\theta_0|S)$. $\qquad\square$

As a consequence of Lemma 85, the right-hand side of Eq. (4.162) can be upper bounded by

$$\sqrt{d\log(\rho\nu TK)} \cdot \sqrt{4\|I_{t-1}(\theta_0)^{-1/2}M_t(\theta_0|S)I_{t-1}(\theta_0)^{-1/2}\|_{\mathrm{op}}}.$$

Lemma 84 is thus proved. We are now ready to prove Lemma 83. By Lemma 84, we know that with high probability

$$\left|\widehat{R}_t(S) - R_t(S)\right| \precsim \sqrt{d\log(\rho\nu TK)} \cdot \sqrt{\|I_{t-1}(\theta_0)^{-1/2}M_t(\theta_0|S)I_{t-1}(\theta_0)^{-1/2}\|_{\mathrm{op}}} \tag{4.165}$$

In addition, by Lemma 85 and the fact that $\|\widehat{\theta}_{t-1} - \theta_0\|_2 \leqslant \tau$ thanks to the local MLE formulation, we have $\frac{1}{4} M_t(\theta_0|S) \leqslant \widehat{M}_t(\widehat{\theta}_{t-1}|S) \leqslant 4 M_t(\theta_0|S)$ and subsequently $\frac{1}{4} I_{t-1}(\theta_0) \leqslant \widehat{I}_{t-1}(\widehat{\theta}_{t-1}) \leqslant 4 I_{t-1}(\theta_0)$ because $I_{t-1}(\cdot)$ and $\widehat{I}_{t-1}(\cdot)$ are summations of $M_{t'}(\cdot)$ and $\widehat{M}_{t'}(\cdot)$ terms. Setting $\omega \gtrsim \sqrt{d \log(\rho \nu T K)}$ we proved that $\overline{R}_t(S) \geqslant R_t(S)$. The second property of Lemma 83 can be proved similarly, by invoking the spectral similarities between $I_{t-1}(\cdot)$, $M_{t'}(\cdot)$ and $\widehat{I}_{t-1}(\cdot)$, $\widehat{M}_{t'}(\cdot)$.

The elliptical potential lemma

Let S_t^* be the assortment that maximizes the expected revenue $R_t(\cdot)$ (defined in Eq. (4.27)) at time period t, and S_t be the assortment selected by Algorithm 13. Because $R_t(S) \leqslant \overline{R}_t(S)$ for all S (see Lemma 83), we have the following upper bound for each term in the regret (see Eq. (4.28)):

$$R_t(S_t^*) - R_t(S_t) \leqslant (\overline{R}_t(S_t^*) - \overline{R}_t(S_t)) + (\overline{R}_t(S_t) - R_t(S_t)) \leqslant \overline{R}_t(S_t) - R_t(S_t), \quad (4.166)$$

where the last inequality holds because $\overline{R}_t(S_t^*) - \overline{R}_t(S_t) \leqslant 0$ (note that S_t maximizes $\overline{R}_t(\cdot)$).

Subsequently, invoking Lemma 83 and the Cauchy-Schwarz inequality, we have

$$\sum_{t=T_0+1}^{T} R_t(S_t^*) - R_t(S_t) \lesssim \sqrt{d \log(\rho \nu T K)} \cdot \sum_{t=T_0+1}^{T} \sqrt{\min\{1, \|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}\}}$$

$$\lesssim \sqrt{dT \log(\rho \nu T K) \cdot \sum_{t=T_0+1}^{T} \min\{1, \|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}^2\}}. \quad (4.167)$$

The following lemma is a key result that upper bounds $\sum_{t=T_0+1}^{T} \min\{1, \|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}^2\}$. It is usually referred to as the *elliptical potential lemma* and has found many applications in contextual bandit-type problems (see, e.g., Dani et al. (2008); Filippi et al. (2010); Li et al. (2017b); Rusmevichientong et al. (2010)).

Lemma 86. *It holds that*

$$\sum_{t=T_0+1}^{T} \min\{1, \|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}^2\} \leqslant 4 \log \frac{\det I_T(\theta_0)}{\det I_{T_0}(\theta_0)} \lesssim d \log(\lambda_0^{-1} \rho \nu).$$

Proof. Denote $A_t := I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)$ as d-dimensional positive semi-definite matrices with eigenvalues sorted as $\sigma_1(A_t) \geqslant \cdots \geqslant \sigma_d(A_t) \geqslant 0$. By simple algebra,

$$\sum_{t=T_0+1}^{T} \min\{1, \|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}^2\} = \sum_{t=T_0+1}^{T} \min\{1, \sigma_1(A_t)^2\}$$

$$\leqslant \sum_{t=T_0+1}^{T} 2 \log(1 + \sigma_1(A_t)^2) \leqslant \sum_{t=T_0+1}^{T} 4 \log(1 + \sigma_1(A_t)). \quad (4.168)$$

173

On the other hand, note that $I_t(\theta_0) = I_{t-1}(\theta_0) + M_t(\theta_0|S_t) = I_{t-1}(\theta_0)^{1/2}[I_{d\times d} + A_t]I_{t-1}(\theta_0)^{1/2}$. Hence,

$$\log \det I_t(\theta_0) = \log \det I_{t-1}(\theta_0) + \sum_{j=1}^{d} \log(1 + \sigma_j(A_t)). \tag{4.169}$$

Comparing Eqs. (4.168) and (4.169), we have

$$\sum_{t=T_0+1}^{T} \min\{1, \|I_{t-1}^{-1/2}(\theta_0)M_t(\theta_0|S_t)I_{t-1}^{-1/2}(\theta_0)\|_{\mathrm{op}}^2\} \leqslant 4\log \frac{\det I_T(\theta_0)}{\det I_{T_0}(\theta_0)}, \tag{4.170}$$

which proves the first inequality in Lemma 86.

We next prove the second inequality in Lemma 86. Because assortments have size 1 throughout the pure exploration phase ($t \leqslant T_0$), we have

$$I_{T_0}(\theta_0) = \sum_{t=1}^{T_0} p_{\theta_0,t}(j_t)(1 - p_{\theta_0,t}(j_t))^2 v_{t,j_t} v_{t,j_t}^\top \geqslant \frac{1}{(1+\rho)^3} \cdot \sum_{t=1}^{T_0} v_{t,j_t} v_{t,j_t}^\top, \tag{4.171}$$

where the last inequality holds thanks to assumption (A2), which implies $p_{\theta_0,t}(j_t) \in [1/(1+\rho), \rho/(1+\rho)]$. In addition, by the proof of Corollary 10, with high probability $\lambda_{\min}(\sum_{t=1}^{T_0} v_{t,j_t} v_{t,j_t}^\top) \geqslant 0.5 T_0 \lambda_0$, where $\lambda_0 > 0$ is a parameter specified in assumption (A1). Therefore,

$$\det I_{T_0}(\theta_0) \gtrsim [T_0 \lambda_0/\rho^3]^d. \tag{4.172}$$

On the other hand, because $\max_{t,j} \|v_{tj}\|_2 \leqslant \nu$ we have $I_T(\theta_0) \lesssim T \cdot \nu^2$ and subsequently

$$\det I_T(\theta_0) \lesssim [\nu^2 T]^d. \tag{4.173}$$

Combining Eqs. (4.172) and (4.173) we proved the second inequality in Lemma 86. $\qquad\square$

We are now ready to give the final upper bound on $\mathrm{Regret}(\{S_t\}_{t=1}^T)$ defined in Eq. (4.28). Note that the total regret incurred by the pure exploration phase is upper bounded by T_0, because the revenue parameters r_{tj} are normalized so that they are upper bounded by 1. In addition, as the failure event of $\overline{R}_t(S) \leqslant R_t(S)$ for some S occurs with probability $1 - O(T^{-1})$, the total regret accumulated under the failure event is $O(T^{-1}) \cdot T = O(1)$. Further invoking Eq. (4.167) and Lemma 86, we have

$$\begin{aligned} \mathrm{Regret}(\{S_t\}_{t=1}^T) &\leqslant T_0 + O(1) + \mathbb{E} \sum_{t=T_0+1}^{T} R_t(S_t^*) - R_t(S_t) \\ &\lesssim O(1) + \frac{\nu^2 d \log T}{\lambda_0^2} + \frac{\rho^2(d + \log T)}{\tau^2 \lambda_0} + d\sqrt{T} \cdot \log(\lambda_0^{-1}\rho\nu TK) \\ &\lesssim O(1) + d^2 \lambda_0^{-2} \rho^4 \nu^2 K^2 \log T + d\sqrt{T} \cdot \log(\lambda_0^{-1}\rho\nu TK). \end{aligned} \tag{4.174}$$

4.7.2 Proof of Theorem 20

At a higher level, the proof of Theorem 20 can be divided into three steps (separated into three different sub-sections below). In the first step, we construct an *adversarial parameter* set and reduce the task of lower bounding the *worst-case* regret of any policy to lower bounding the *Bayes risk* of the constructed parameter set. In the second step, we use a "counting argument" similar to the one developed in the work of Chen & Wang (2018) to provide an explicit lower bound on the Bayes risk of the constructed adversarial parameter set, and finally we apply *Pinsker's inequality* (see, e.g., (Tsybakov, 2009)) to derive a complete lower bound.

Adversarial construction and the Bayes risk

Let $\epsilon \in (0, 1/d\sqrt{d})$ be a small positive parameter to be specified later. For every subset $W \subseteq [d]$, define the corresponding parameter $\theta_W \in \mathbb{R}^d$ as $[\theta_W]_i = \epsilon$ for all $i \in W$, and $[\theta_W]_i = 0$ for all $i \notin W$. The parameter set we consider is

$$\theta \in \Theta := \{\theta_W : W \in \mathcal{W}_{d/4}\} := \{\theta_W : W \subseteq [d], |W| = d/4\}. \tag{4.175}$$

Note that $d/4$ is a positive integer because d is divisible by 4, as assumed in Theorem 20. Also, to simplify notation, we use $\mathcal{W}_k$ to denote the class of all subsets of $[d]$ whose size is k.

The feature vectors $\{v_{ti}\}$ are constructed to be invariant across time iterations t. For each t and $U \in \mathcal{W}_{d/4}$, K identical feature vectors v_U are constructed as (recall that K is the maximum allowed assortment capacity)

$$[v_U]_i = 1/\sqrt{d} \quad \text{for } i \in U; \qquad [v_U]_i = 0 \quad \text{for } i \notin U. \tag{4.176}$$

It is easy to check that with the condition $\epsilon \in (0, 1/\sqrt{d})$, $\|\theta_W\|_2 \leqslant 1$ and $\|v_U\|_2 \leqslant 1$ for all $W, U \in \mathcal{W}_{d/4}$. Hence the worst-case regret of any policy π can be lower bounded by the worst-case regret of parameters belonging to Θ, which can be further lower bounded by the "average" regret over a uniform prior over Θ:

$$\sup_{v,\theta} \mathbb{E}_{v,\theta}^{\pi} \sum_{t=1}^{T} R(S_\theta^*) - R(S_t) \geqslant \max_{\theta_W \in \Theta} \mathbb{E}_{v,\theta_W}^{\pi} \sum_{t=1}^{T} R(S_{\theta_W}^*) - R(S_t)$$

$$\geqslant \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_{v,\theta_W}^{\pi} \sum_{t=1}^{T} R(S_{\theta_W}^*) - R(S_t). \tag{4.177}$$

Here S_θ^* is the optimal assortment of size at most K that maximizes (expected) revenue under parameterization θ. By construction, it is easy to verify that $S_{\theta_W}^*$ consists of all K items corresponding to feature v_W. We also employ constant revenue parameters $r_{ti} \equiv 1$ for all $t \in [T]$, $i \in [N]$.

The counting argument

In this section we drive an explicit lower bound on the Bayes risk in Eq. (4.177). For any sequences $\{S_t\}_{t=1}^{T}$ produced by the policy π, we first describe an alternative sequence $\{\widetilde{S}_t\}_{t=1}^{T}$ that provably enjoys less regret under parameterization θ_W, while simplifying our analysis.

Let $v_{U_1}, \cdots, v_{U_L}$ be the distinct feature vectors contained in assortment S_t (if $S_t = \varnothing$ then one may choose an arbitrary feature v_U) with $U_1, \cdots, U_L \in \mathcal{W}_{d/4}$. Let U^* be the subset among $U_1, \cdots, U_L$ that maximizes $\langle v_{U*}, \theta_W \rangle$, where θ_W is the underlying parameter. Let $\widetilde{S}_t$ be the assortment consisting of all K items corresponding to feature v_U^*. We then have the following observation:

Proposition 20. $R(S_t) \leqslant R(\widetilde{S}_t)$ *under* θ_W.

Proof. Because $r_{tj} \equiv 1$ in our construction, we have $R(S_t) = (\sum_{j \in S_t} u_j)/(1 + \sum_{j \in S_t} u_j)$ where $u_j = \exp\{v_j^\top \theta_W\}$ under θ_W. Clearly $R(S)$ is a monotonically non-decreasing function in u_j. By replacing all $v_j \in S_t$ with $v_{U*} \in \widetilde{S}_t$, the u_j values do not decrease and therefore the Proposition holds true. $\qquad\square$

To simplify notation we also use $\widetilde{U}_t$ to denote the unique $U^* \in \mathcal{W}_{d/4}$ in $\widetilde{S}_t$. We also use $\mathbb{E}_W$ and $\mathbb{P}_W$ to denote the law parameterized by θ_W and policy π. The following lemma gives a lower bound on $R(\widetilde{S}_t) - R(S_{\theta_W}^*)$ by comparing it with W.

Lemma 87. *Suppose* $\epsilon \in (0, 1/d\sqrt{d})$ *and define* $\delta := d/4 - |\widetilde{U}_t \cap W|$. *Then*

$$R(S_{\theta_W}^*) - R(\widetilde{S}_t) \geqslant \frac{\delta\epsilon}{4K\sqrt{d}}.$$

Proof. Let $v = v_W$ and $\widehat{v} = v_{\widetilde{U}_t}$ be the corresponding feature vectors. Then

$$
\begin{aligned}
R(S_{\theta_W}^*) - R(\widetilde{S}_t) &= \frac{K \exp\{v^\top \theta_W\}}{1 + K \exp\{v^\top \theta_W\}} - \frac{K \exp\{\widehat{v}^\top \theta_W\}}{1 + K \exp\{\widehat{v}^\top \theta_W\}} \\
&= \frac{K[\exp\{v^\top \theta_W\} - \exp\{\widehat{v}^\top \theta_W\}]}{(1 + K \exp\{v^\top \theta_W\})(1 + K \exp\{\widehat{v}^\top \theta_W\})} \\
&\geqslant \frac{\exp\{v^\top \theta_W\} - \exp\{\widehat{v}^\top \theta_W\}}{2Ke}.
\end{aligned}
$$

Here the last inequality holds because $\max(\exp\{v^\top \theta_W\}, \exp\{\widehat{v}^\top \theta_W\}) \leqslant e$. In addition, by Taylor expansion we know that $1 + x \leqslant e^x \leqslant 1 + x + x^2/2$ for all $x \in [0, 1]$. Subsequently,

$$R(S_{\theta_W}^*) - R(\widetilde{S}_t) \geqslant \frac{(v - \widehat{v})^\top \theta_W - (\widehat{v}^\top \theta_W)^2/2}{2Ke} \geqslant \frac{\delta\epsilon/\sqrt{d} - (\sqrt{d}\epsilon)^2/2}{2Ke}.$$

Finally, noting that $d\epsilon^2/2 \leqslant \delta\epsilon/2\sqrt{d}$ provided that $\epsilon \in (0, 1/d\sqrt{d})$, we finish the proof of Lemma 87. $\qquad\square$

Define random variables $\widetilde{N}_i := \sum_{t=1}^T 1\{i \in \widetilde{U}_t\}$. Lemma 87 immediately implies

$$\mathbb{E}_W \sum_{t=1}^T R(S_{\theta_W}^*) - R(\widetilde{S}_t) \geqslant \frac{\epsilon}{4K\sqrt{d}}\left(\frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\widetilde{N}_i]\right), \quad \forall W \in \mathcal{W}_{d/4}. \qquad (4.178)$$

176

Denote $\mathcal{W}_{d/4}^{(i)} := \{W \in \mathcal{W}_{d/4} : i \in W\}$ and $\mathcal{W}_{d/4-1} := \{W \subseteq [d] : |W| = d/4 - 1\}$. Averaging both sides of Eq. (4.178) with respect to all $W \in \mathcal{W}_{d/4}$ and swapping the summation order, we have

$$\frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S_{\theta_W}^*) - R(S_t) \geq \frac{\epsilon}{4K\sqrt{d}} \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \left(\frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i] \right)$$

$$= \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{4} - \frac{1}{|\mathcal{W}_{d/4}|} \sum_{i=1}^{d} \sum_{W \in \mathcal{W}_{d/4}^{(i)}} \mathbb{E}_W[\tilde{N}_i] \right)$$

$$= \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{4} - \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4-1}} \sum_{i \notin W} \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] \right)$$

$$\geq \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{4} - \frac{|\mathcal{W}_{d/4-1}|}{|\mathcal{W}_{d/4}|} \max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \notin W} \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] \right)$$

$$= \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{4} - \frac{|\mathcal{W}_{d/4-1}|}{|\mathcal{W}_{d/4}|} \max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \notin W} \mathbb{E}_W[\tilde{N}_i] + \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i] \right).$$

Note that for any fixed W, $\sum_{i \notin W} \mathbb{E}_W[\tilde{N}_i] \leq \sum_{i=1}^{d} \mathbb{E}_W[\tilde{N}_i] \leq dT/4$. Also, $|\mathcal{W}_{d/4-1}|/|\mathcal{W}_{d/4}| = \binom{d}{d/4-1}/\binom{d}{d/4} = \frac{d/4}{3d/4+1} \leq 1/3$. Subsequently,

$$\frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S_{\theta_W}^*) - R(S_t) \geq \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{6} - \max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \notin W} |\mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i]| \right).$$
$$(4.179)$$

Pinsker's inequality

In this section we concentrate on upper bounding $|\mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i]|$ for any $W \in \mathcal{W}_{d/4-1}$. Let $P = \mathbb{P}_W$ and $Q = \mathbb{P}_{W \cup \{i\}}$ denote the laws under θ_W and $\theta_{W \cup \{i\}}$, respectively. Then

$$\left| \mathbb{E}_P[\tilde{N}_i] - \mathbb{E}_Q[\tilde{N}_i] \right| \leq \sum_{j=0}^{T} j \cdot \left| P[\tilde{N}_i = j] - Q[\tilde{N}_i = j] \right|$$

$$\leq T \cdot \sum_{j=0}^{T} \left| P[\tilde{N}_i = j] - Q[\tilde{N}_i = j] \right|$$

$$\leq T \cdot \|P - Q\|_{\mathrm{TV}} \leq T \cdot \sqrt{\frac{1}{2} \mathrm{KL}(P \| Q)},$$

where $\|P - Q\|_{\mathrm{TV}} = \sup_A |P(A) - Q(A)|$ is the total variation distance between P, Q, $\mathrm{KL}(P \| Q) = \int (\log \mathrm{d}P/\mathrm{d}Q) \mathrm{d}P$ is the Kullback-Leibler (KL) divergence between P, Q, and the inequality $\|P - Q\|_{\mathrm{TV}} \leq \sqrt{\frac{1}{2} \mathrm{KL}(P \| Q)}$ is the celebrated Pinsker's inequality.

For every $i \in [d]$ define random variables $N_i := \sum_{t=1}^{T} \frac{1}{K} \sum_{v_U \in S_t} 1\{i \in U\}$. The next lemma upper bound the KL divergence:

Lemma 88. *For any $W \in \mathcal{W}_{d/4-1}$ and $i \in [d]$, $\mathrm{KL}(P_W \| P_{W \cup \{i\}}) \leqslant C_{\mathrm{KL}} \cdot \mathbb{E}_W[N_i] \cdot \epsilon^2/d$ for some universal constant $C_{\mathrm{KL}} > 0$.*

Proof. Fix a time t with policy's assortment choice S_t, and define $n_i(S_t) := \sum_{v_U \in S_t} 1\{i \in U\}/K$. Let $\{p_j\}_{j \in S_t \cup \{0\}}$ and $\{q_j\}_{j \in S_t \cup \{0\}}$ be the probabilities of purchasing item j under parameterization θ_W and $\theta_{W \cup \{i\}}$, respectively. Then

$$\mathrm{KL}(P_W(\cdot|S_t) \| P_{W \cup \{i\}}(\cdot|S_t)) = \sum_{j \in S_t \cup \{0\}} p_j \log \frac{q_j}{p_j} \leqslant \sum_j p_j \frac{p_j - q_j}{q_j} \leqslant \sum_j \frac{|p_j - q_j|^2}{q_j}, \quad (4.180)$$

where the only inequality holds because $\log(1 + x) \leqslant x$ for all $x > -1$. Because $q_j \geqslant e^{-1}/(1 + Ke) \geqslant 1/(2Ke^2)$ for all $j \in S_t \cup \{0\}$, Eq. (4.180) is reduced to

$$\mathrm{KL}(P_W(\cdot|S_t) \| P_{W \cup \{i\}}(\cdot|S_t)) \leqslant 2e^2 K \cdot \sum_{j \in S_t \cup \{0\}} |p_j - q_j|^2. \quad (4.181)$$

We next upper bound $|p_j - q_j|$ separately. First consider $j = 0$. We have

$$
\begin{aligned}
|p_j - q_j| &= \left| \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_W\}} - \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_{W \cup \{i\}}\}} \right| \\
&\leqslant \frac{1}{(1 + K/e)^2} \cdot 2 \sum_{j \in S_t} \left| v_j^\top (\theta_W - \theta_{W \cup \{i\}}) \right| \\
&\leqslant \frac{2K n_i(S_t)\epsilon/\sqrt{d}}{(1 + K/e)^2} \leqslant \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}}.
\end{aligned}
$$

Here the first inequality holds because $e^x \leqslant 1 + 2x$ for all $x \in [0, 1]$.
For $j > 0$ corresponding to $v_j = v_U$ where $i \notin U$, we have

$$
\begin{aligned}
|p_j - q_j| &= \left| \frac{\exp\{v_U^\top \theta_W\}}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_W\}} - \frac{\exp\{v_U^\top \theta_{W \cup \{i\}}\}}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_{W \cup \{i\}}\}} \right| \\
&\leqslant \left| \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_W\}} - \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_{W \cup \{i\}}\}} \right| \\
&\leqslant \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}}.
\end{aligned}
$$

Here the first inequality holds because $\exp\{v_U^\top \theta_W\} = \exp\{v_U^\top \theta_{W \cup \{i\}}\} \leqslant 1$, since $i \notin U$.
For $j > 0$ corresponding to $v_j = v_U$ and $i \in U$, we have

$$|p_j - q_j| = \left| \frac{\exp\{v_U^\top \theta_W\}}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_W\}} - \frac{\exp\{v_U^\top \theta_{W \cup \{i\}}\}}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_{W \cup \{i\}}\}} \right|$$

$$\leqslant \exp\{v_u^\top \theta_{W \cup \{i\}}\} \cdot \left| \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_W\}} - \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_{W \cup \{i\}}\}} \right|$$

$$+ \left| \exp\{v_u^\top \theta_W\} - \exp\{v_u^\top \theta_{W \cup \{i\}}\} \right| \cdot \left| \frac{1}{1 + \sum_{j \in S_t} \exp\{v_j^\top \theta_W\}} \right|$$

$$\leqslant \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}} + \frac{\epsilon}{\sqrt{d}} \cdot \frac{1}{1 + K/e} \cdot \leqslant \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}} + \frac{2e\epsilon}{K\sqrt{d}}.$$

Combining all upper bounds on $|p_j - q_j|$ and Eq. (4.181), we have

$$\mathrm{KL}(P_W(\cdot|S_t)\|P_{W \cup \{i\}}(\cdot|S_t)) \leqslant 2e^2 K \cdot \left[\frac{128e^4 n_i(S_t)^2 \epsilon^2}{K^2 d}(1 + K) + K n_i(S_t) \cdot \frac{8e^4 \epsilon^2}{K^2 d} \right]$$

$$\lesssim n_i(S_t)\epsilon^2/d.$$

Here the last inequality holds because $n_i(S_t) \leqslant 1$. Note also that $N_i = \sum_{t=1}^{T} n_i(S_t)$ by definition, and subsequently summing over all $t = 1$ to T we have

$$\mathrm{KL}(P_W\|P_{W \cup \{i\}}) \lesssim \mathbb{E}_W[N_i] \cdot \epsilon^2/d,$$

which is to be demonstrated. □

Combining Lemma 88 and Eq. (4.179), we have

$$\frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S_{\theta_W}^*) - R(S_t) \geqslant \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{6} - T \sum_{i=1}^{d} \sqrt{C_{\mathrm{KL}} \mathbb{E}_W[N_i]\epsilon^2/d} \right).$$

Further using Cauchy-Schwartz inequality, we have

$$\sum_{i=1}^{d} \sqrt{C_{\mathrm{KL}} \mathbb{E}_W[N_i]\epsilon^2/d} \leqslant \sqrt{d} \cdot \sqrt{\sum_{i=1}^{d} C_{\mathrm{KL}} \mathbb{E}_W[N_i]\epsilon^2/d},$$

which is further upper bounded by $\sqrt{d} \cdot \sqrt{C_{\mathrm{KL}} T \epsilon^2/4}$ because $\sum_{i=1}^{d} \mathbb{E}_W[N_i] \leqslant dT/4$. Subsequently,

$$\frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S_{\theta_W}^*) - R(S_t) \geqslant \frac{\epsilon}{4K\sqrt{d}} \left(\frac{dT}{6} - T\sqrt{C'_{\mathrm{KL}} dT \epsilon^2} \right), \qquad (4.182)$$

where $C'_{\mathrm{KL}} = C_{\mathrm{KL}}/4$. Setting $\epsilon = \sqrt{d/144C'_{\mathrm{KL}}T}$ we complete the proof of Theorem 20.

Chapter 5

Conclusion and discussion

In this thesis, we study various problmes under the general theme of selective data acquision in learning and decision making. In this ending chapter, we summarize the main conceptual findings from this thesis, and also discuss potential future directions for extending the presented thesis work.

5.1   Benefits of selective or active data acquisition

Using selective or interactive data collection schemes to improve data efficiency has a long history in statistics and machine learning research (Balcan et al., 2009; Cohn et al., 1996; Fedorov, 1972; Hanneke et al., 2014; Pukelsheim, 2006; Settles, 2009; Tong & Koller, 2001; Wu & Hamada, 2011). Intuitively, by focusing data collection to regions that are most informative about the underlying data generation procedure, the efficiency of the data analysis procedures is much improved. Rigorous theoretical justification also exist, mostly targeting active regression or classification problems (Balcan et al., 2010; Balcan & Long, 2013; Castro & Nowak, 2008; Hanneke et al., 2014; Krishnamurthy, 2015; Wang, 2011; Wang & Singh, 2016).

In this thesis work, we study the benefits of selective or active data acquisition schemes beyond their traditional applications in statistical regression and classification. Our theoretical and empirical results re-affirm the benefits of selective data acquisition for problems like nonparametric optimization and/or dynamic assortment optimization.

More specifically, for optimizing an unknown, non-convex smooth function in low dimensions, our theoretical results show a *polynomial* gap of the optimal convergence rates (sample complexity) between passive and interactive query schemes when the objective functions have a non-trivial level set growth (Theorems 4, 5 and 6) via localized minimax analysis, whose importance we will also discuss in the next subsection. For dynamic assortment optimization, our results on regret upper bounds (Theorems 13, 14, 17 and 19) emphasize the importance of combining statistical estimation (of customers' utility model parameters) and combinatorial optimization (of assortments) at the same time in order to achieve the minimum regret possible over T sequentially arriving customers. This latter point is also discussed in Sec. 5.3 later.

5.2 Importance of optimality and minimax analysis

In this thesis, *Optimality* of our proposed methods is studied using the *minimax* framework (Ibragimov & Has'minskii, 1981; Korostelev & Tsybakov, 2012; Tsybakov, 2009) and its variants from the statistics literature, which lower bounds the *worst-case* error or regret any algorithm/policy will incur. While such analysis leads to negative results based on worst-case scenarios, its importance should not be undermined as new insights are generated through such minimax analysis. Below we mention two examples.

First, for the problem of global optimization of smooth non-convex functions, our minimax analysis (together with corresponding upper bounds and algorithms) establishes a clear separation between passive and interactive schemes. Such a separation would not be possible without rigorous optimality analysis of the proposed algorithms. In addition, we remark that for this particular question the classical *global* minimax analysis will not separate passive and interactive schemes either, as the worst-case function over all smooth functions is the same for both query schemes. To overcome this difficulty, a *local* variant of minimax analysis is adopted in order to prove a difference in convergence rates for objective functions with non-trivial level set growth.

Second, analysis of optimal regret plays an important role in revealing some surprising phase transitions of problem complexity when only very subtle changes in the problem settings are present. For example, in Sec. 3.3 where optimization of non-stationary convex function sequences is studied, a single change in the norms of how function variation is measured gives rises to the curse-of-dimensionality phenomenon, which is not visible when the most restrictive $p = \infty$ norm is used. Another example concerns dynamic assortment optimization under the plain MNL model: results in Theorems 13, 14 and 16 together reveal a surprising phase transition between the uncapacitated case ($K = N$) with no N-dependency and the capacitated case ($K \leqslant N/4$) with $\sqrt{N}$-dependency. Such unexpected phenomena would be counter-intuitive without rigorous optimality analysis matching upper bounds of proposed algorithms.

5.3 Unification of data analysis and decision making

Many fundamental tasks in operations and revenue management involve decision making and optimization, such as the optimization of commodity assortments for online or offline display (Davis et al., 2014; Gallego et al., 2004; Kök et al., 2008; Li & Rusmevichientong, 2014; Mahajan & van Ryzin, 2001; Talluri & van Ryzin, 2004; van Ryzin & Mahajan, 1999), the determination of optimal pricing of items (Bitran & Caldentey, 2003; Elmaghraby & Keskinocak, 2003; Talluri & Van Ryzin, 2006) and stochastic assignments of sequentially arriving jobs to workers or kidneys to patients under medical management settings (Bertsimas et al., 2013; Derman et al., 1972; Su & Zenios, 2005; Zenios et al., 2000).

Traditionally, the above-mentioned decision making problems are solved with environmental parameters and settings fully specified, by resorting to greedy or dynamic programming (DP) type methods. Unfortunately, the more common practical scenario is when the modeling parameters are unknown a priori, which have to be estimated either offline from historical data or online simultaneously from sequentially made decisions. The uncertainty in the estimated parameters also brings unique challenges to existing algorithms originally designed for full-information set-

tings. One particular challenge is the instability of most dynamic programming methods, in which even a small estimation error in the model parameters might incur large deviation from the optimal policy in a DP formulation (Goldenshluger & Zeevi, 2017).

To overcome the difficulties DP-type algorithms face, *regret minimization* arises as a powerful *unified* framework for simultaneous estimation and decision making. While the concept of regret (and its optimality) has been very common in multi-armed and contextual bandit problems (Abbasi-Yadkori et al., 2011; Audibert et al., 2011; Auer, 2002; Bubeck & Cesa-Bianchi, 2012; Chu et al., 2011; Filippi et al., 2010; Li et al., 2017b; Rusmevichientong & Tsitsiklis, 2010), its study in more general operations management questions remains a relatively new area and has attracted significant recent research efforts (Agrawal et al., 2017a,b; Caro & Gallien, 2007; El-maghraby & Keskinocak, 2003; Rusmevichientong et al., 2010). Our results in Chapter 4 further extend such efforts to more complex and practical revenue management models, and we envision a much wider range of problems that could benefit from similar ideas and treatments, which we elaborate in more details in the next subsection.

5.4 Future directions

Based on the thesis presented, my future work would extend the learning-while-doing framework for operations research and management problems by combining both the perspectives of asymptotic regret analysis and exact policy optimization. Below are two major directions I would like to pursue in the near future.

5.4.1 Regret analysis for dynamic programming with partial information

Traditionally, when full information about the environment is available, the optimal strategy of a sequential decision making problem can be obtained by solving a (stochastic) dynamic programming. When only partial information is available, however, such an approach becomes less practical as dynamic programming is generally sensitive to small perturbation (estimation error) of the decision process parameters. For such settings, asymptotic regret analysis might be a more appropriate framework.

One such example is the question of assigning sequentially arriving jobs to awaiting workers, which is a classical question in operations research, with important applications in kidney exchange systems (Bertsimas et al., 2013; Derman et al., 1972; Su & Zenios, 2005; Zenios et al., 2000). When the difficulty levels of arriving jobs (or qualities of available kidneys) are stochastic and follow a known probability distribution, exactly optimal allocation policies can be found by dynamic programming (Derman et al., 1972). It is an interesting question to study the allocation problem when the underlying distribution of arriving jobs is unknown and have to be learnt from observations of previously assigned matches.

5.4.2 POMDP and reinforcement learning

The results presented in this thesis on dynamic assortment planning with unknown utility models (Chapter 4), as well as previous works of Agrawal et al. (2017a,b); Rusmevichientong et al.

(2010), all took a regret minimization approach by first developing an online assortment recommendation policy and then proving upper bounds on its regret. When possible, information-theoretical lower bounds are proved to establish the (asymptotic) optimality of the proposed policies.

When prior information about the customers' utility models is available, existing regret based approaches might be too conservative as they typically only consider the *worst-case* regret. Partially observable Markov decision processes (POMDPs, (Astrom, 1965; Kaelbling et al., 1998)) present a more flexible framework for the modeling and solving of dynamic assortment optimization questions, by representing the unknown utility model parameters as unobserved states. Approximate computation techniques such as the ones introduced in the works of Fukuda (2004); Zhang (2010) could also be applied.

Appendix A

Useful inequalities

A.1 Scalar concentration inequalities

Lemma 89 (Hoeffding (1963))**.** *Suppose $X_1, \ldots, X_n$ are i.i.d. random variables such that $a \leqslant X_i \leqslant b$ almost surely. Then for any $t > 0$,*

$$\Pr\left[\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}X\right| > t\right] \leqslant 2\exp\left\{-\frac{nt^2}{2(b-a)^2}\right\}.$$

Lemma 90 (Hsu et al. (2012))**.** *Suppose $x \sim \mathcal{N}_d(0, I_{d\times d})$ and let A be a $d \times d$ positive semi-definite matrix. Then for all $t > 0$,*

$$\Pr\left[x^\top A x > \mathrm{tr}(A) + 2\sqrt{\mathrm{tr}(A^2)t} + 2\|A\|_{\mathrm{op}}t\right] \leqslant e^{-t}.$$

Lemma 91 (Bernstein's inequality)**.** *Suppose X is a sub-exponential random variable with parameters ν and α.*

$$\Pr\left[|X - \mathbb{E}X| > t\right] \leqslant \left\{\begin{array}{ll} 2\exp\left\{-t^2/2\nu^2\right\}, & 0 < t \leqslant \nu^2/\alpha; \\ 2\exp\left\{-t/2\alpha\right\}, & t > \nu^2/\alpha. \end{array}\right.$$

The following lemma is a simplified version of Theorem 1.2A in (Victor, 1999) (note that the original form in (Victor, 1999) is one-sided; the two-sided version below can be trivially obtained by considering $-X_1, \ldots, -X_n$ and applying the union bound).

Lemma 92 (Bernstein's inequality for martingales)**.** *Suppose $X_1, \ldots, X_n$ are random variables such that $\mathbb{E}[X_j|X_1, \ldots, X_{j-1}] = 0$ and $\mathbb{E}[X_j^2|X_1, \ldots, X_{j-1}] \leqslant \sigma^2$ for all $t = 1, \ldots, n$. Further assume that $\mathbb{E}[|X_j|^k|X_1, \ldots, X_{j-1}] \leqslant \frac{1}{2}k!\sigma^2 b^{k-2}$ for all integers $k \geqslant 3$. Then for all $t > 0$,*

$$\Pr\left[\left|\sum_{j=1}^{n} X_j\right| \geqslant t\right] \leqslant 2\exp\left\{-\frac{t^2}{2(n\sigma^2 + bt)}\right\}.$$

The following lemma is the Hoeffding's maximal inequality, by Hoeffding (1963).

Lemma 93 (Hoeffding's maximal inequality)**.** *Let* $X_1, \cdots, X_n$ *be i.i.d. random variables with mean* μ *and satisfy* $a \leqslant X_i \leqslant b$ *almost surely for all* $i \in [n]$. *Then for any* $t > 0$,

$$\Pr\left[\forall i \in [n], X_1 + \cdots + X_i \geqslant i \cdot \mu + t\right] \leqslant \exp\left\{-\frac{2t^2}{n(b-a)^2}\right\}. \tag{A.1}$$

The following result is cited from Theorem 5 of (Agrawal et al., 2017a).

Lemma 94 (Concentration of geometric random variables (Agrawal et al., 2017a))**.** *Suppose* $X_1, \cdots, X_n$ *are i.i.d. geometric random variables with parameters* $p > 0$, *meaning that* $\Pr[X_i = k] = (1-p)^k p$ *for* $k = 0, 1, 2, \cdots$. *Define* $\mu := \mathbb{E}X_i = (1-p)/p$. *Then*

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i > (1+\delta)\mu\right] \leqslant \begin{cases} \exp\left\{-\frac{n\mu\delta^2}{2(1+\delta)(1+\mu)^2}\right\}, & \text{if } \mu \leqslant 1, \\ \exp\left\{-\frac{n\delta^2\mu^2}{6(1+\mu)^2}\left(3 - \frac{2\delta\mu}{1+\mu}\right)\right\}, & \text{if } \mu \geqslant 1, \delta \in (0,1); \end{cases}$$

$$\Pr\left[\frac{1}{n}\sum_{i=1}^n X_i < (1-\delta)\mu\right] \leqslant \begin{cases} \exp\left\{-\frac{n\delta^2\mu}{6(1+\mu)^2}\left(3 - \frac{2\delta\mu}{1+\mu}\right)\right\}, & \text{if } \mu \leqslant 1, \\ \exp\left\{-\frac{n\delta^2\mu^2}{2(1+\mu)^2}\right\}, & \text{if } \mu \geqslant 1. \end{cases}$$

A.2 Matrix/vector concentration inequalities

Lemma 95 (Rudelson & Vershynin (2007))**.** *Let* x *be a* p-*dimensional random vector such that* $\|x\|_2 \leqslant M$ *almost surely and* $\|\mathbb{E}xx^\top\|_2 \leqslant 1$. *Let* $x_1, \cdots, x_n$ *be i.i.d. copies of* x. *Then for every* $t \in (0, 1)$

$$\Pr\left[\left\|\frac{1}{n}\sum_{i=1}^n x_i x_i^\top - \mathbb{E}xx^\top\right\|_2 > t\right] \leqslant 2\exp\left\{-C \cdot \frac{nt^2}{M^2 \log n}\right\},$$

where $C > 0$ *is some universal constant.*

Lemma 96 (Corollary 5.2 of Mackey et al. (2014))**.** *Let* $(Y_k)_{k \geqslant 1}$ *be a sequence of random* d-*dimensional Hermitian matrices that satisfy*

$$\mathbb{E}Y_k = 0 \quad \text{and} \quad \|Y_k\|_2 \leqslant R \quad a.s.$$

Define $X = \sum_{k \geqslant 1} Y_k$. *The for any* $t > 0$,

$$\Pr\left[\|X\|_2 \geqslant t\right] \leqslant d \cdot \exp\left\{-\frac{t^2}{3\sigma^2 + 2Rt}\right\} \quad \text{for} \quad \sigma^2 = \left\|\sum_{k \geqslant 1} \mathbb{E}Y_k^2\right\|_2.$$

Lemma 97 (Corollary 10.3 of Mackey et al. (2014))**.** *Let* $A_1, \cdots, A_n$ *be a sequence of deterministic* d-*dimensional Hermitian matrices that satisfy*

$$\sum_{k=1}^n A_k = 0 \quad \text{and} \quad \sup_{1 \leqslant k \leqslant n} \|A_k\|_2 \leqslant R.$$

Define random matrix $X = \sum_{j=1}^{m} A_{\sigma(j)}$ for $m \leqslant n$, where σ is a random permutation from $[n]$ to $[n]$. Then for all $t > 0$,

$$\Pr\left[\|X\|_2 \geqslant t\right] \leqslant d \exp\left\{-\frac{t^2}{12\sigma^2 + 4\sqrt{2}Rt}\right\} \quad \textit{for } \ \sigma^2 = \frac{m}{n}\left\|\sum_{k=1}^{n} A_k^2\right\|_2.$$

Lemma 98 (Tropp (2015), simplified). *Suppose $A_1, \ldots, A_n$ are i.i.d. positive semidefinite random matrices of dimension d and $\|A_i\|_{\mathrm{op}} \leqslant R$ almost surely. Then for any $t > 0$,*

$$\Pr\left[\left\|\frac{1}{n}\sum_{i=1}^{n} A_i - \mathbb{E}A\right\|_{\mathrm{op}} > t\right] \leqslant 2\exp\left\{-\frac{nt^2}{8R^2}\right\}.$$

A.3 Other inequalities

Lemma 99 (Weyl's inequality). *Let A and $A + E$ be $d \times d$ matrices with $\sigma_1, \ldots, \sigma_d$ and $\sigma'_1, \ldots, \sigma'_d$ being their singular values, sorted in descending order. Then $\max_{1 \leqslant i \leqslant d} |\sigma_i - \sigma'_i| \leqslant \|E\|_{\mathrm{op}}$.*

Bibliography

Abbasi-Yadkori, Y., Pál, D., & Szepesvári, C. (2011). Improved algorithms for linear stochastic bandits. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*. 135, 183

Abe, N., Biermann, A. W., & Long, P. M. (2003). Reinforcement learning with immediate rewards and linear hypotheses. *Algorithmica*, *37*(4), 263–293. 135

Abernethy, J., Hazan, E., & Rakhlin, A. (2008). Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the Conference on Learning Theory (COLT)*. 94

Agarwal, A., Bartlett, P. L., Ravikumar, P., & Wainwright, M. J. (2012). Information-theoretic lower bounds on the oracle complexity of stochastic convex optimization. *IEEE Transactions on Information Theory*, *58*(5), 3235–3249. 49, 57, 99

Agarwal, A., Dekel, O., & Xiao, L. (2010). Optimal algorithms for online convex optimization with multi-point bandit feedback. In *Proceedings of the annual Conference on Learning Theory (COLT)*. 45, 56, 64, 65

Agarwal, A., Foster, D. P., Hsu, D., Kakade, S. M., & Rakhlin, A. (2013). Stochastic convex optimization with bandit feedback. *SIAM Journal on Optimization*, *23*(1), 213–240. 59, 134

Ageev, A. A., & Sviridenko, M. I. (2004). Pipage rounding: A new method of constructing algorithms with proven performance guarantee. *Journal of Combinatorial Optimization*, *8*(3), 307–328. 20

Agrawal, S., Avadhanula, V., Goyal, V., & Zeevi, A. (2017a). MNL-bandit: A dynamic learning approach to assortment selection. *arXiv preprint arXiv:1706.03880*. 3, 107, 113, 114, 115, 117, 120, 121, 131, 133, 183, 186

Agrawal, S., Avandhanula, V., Goyal, V., & Zeevi, A. (2017b). Thompson sampling for MNL-bandit. In *Proceedings of the annual Conference on Learning Theory (COLT)*. 107, 113, 114, 115, 117, 133, 183

Agrawal, S., & Goyal, N. (2013). Thompson sampling for contextual bandits with linear payoffs. In *Proceedings of the International Conference on Machine Learning (ICML)*. 135

Allen-Zhu, Z., & Li, Y. (2017). Follow the compressed leader: Faster online learning of eigenvectors and faster MMWU. In *Proceedings of the International Conference on Machine Learning (ICML)*. 28

Allen-Zhu, Z., Liao, Z., & Orecchia, L. (2015). Spectral sparsification and regret minimization

beyond matrix multiplicative updates. In *Proceedings of Annual Symposium on the Theory of Computing (STOC)*. 11, 28, 29, 32

Anderson, S., de Palma, A., & Thisse, J. F. (1992). *Discrete Choice Theory of Product Differentiation*. MIT Press. 2, 107

Arora, S., & Kale, S. (2007). A combinatorial, primal-dual approach to semidefinite programs. In *Proceedings of the annual ACM Symposium on Theory of Computing (STOC)*. 28

Astrom, K. J. (1965). Optimal control of markov processes with incomplete state information. *Journal of Mathematical Analysis and Applications*, *10*(1), 174–205. 184

Audet, C., & Dennis Jr, J. E. (2006). Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on Optimization*, *17*(1), 188–217. 64

Audibert, J.-Y., & Bubeck, S. (2009). Minimax policies for adversarial and stochastic bandits. In *Proceedings of the Conference on Learning Theory (COLT)*. 112

Audibert, J.-Y., Bubeck, S., & Lugosi, G. (2011). Minimax policies for combinatorial prediction games. In *Proceedings of Conference on Learning Theory (COLT)*. 28, 183

Auer, P. (2002). Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, *3*(Nov), 397–422. 135, 183

Avron, H., & Boutsidis, C. (2013). Faster subset selection for matrices and applications. *SIAM Journal on Matrix Analysis and Applications*, *34*(4), 1464–1499. 10, 20, 21, 25, 26

Balakrishnan, S., Narayanan, S., Rinaldo, A., Singh, A., & Wasserman, L. (2013). Cluster trees on manifolds. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 65

Balcan, M.-F., Beygelzimer, A., & Langford, J. (2009). Agnostic active learning. *Journal of Computer and System Sciences*, *75*(1), 78–89. 65, 181

Balcan, M.-F., Hanneke, S., & Vaughan, J. W. (2010). The true sample complexity of active learning. *Machine Learning*, *80*(2-3), 111–139. 181

Balcan, M.-F., & Long, P. M. (2013). Active and passive learning of linear separators under log-concave distributions. In *Proceedings of Annual Conference on Learning Theory (COLT)*. 181

Bandeira, A. S., Scheinberg, K., & Vicente, L. N. (2012). Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. *Mathematical Programming*, *134*, 223–257. 48, 64

Bandeira, A. S., Scheinberg, K., & Vicente, L. N. (2014). Convergence of trust-region methods based on probabilistic models. *SIAM Journal on Optimization*, *24*(3), 1238–1264. 64

Beck, A., & Teboulle, M. (2003). Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters*, *31*(3), 167–175. 7, 8, 57

Bertsimas, D., Farias, V. F., & Trichakis, N. (2013). Fairness, efficiency, and flexibility in organ allocation for kidney transplantation. *Operations Research*, *61*(1), 73–87. 182, 183

Besbes, O., Gur, Y., & Zeevi, A. (2015). Non-stationary stochastic optimization. *Operations Research*, *63*(5), 1227–1244. 38, 59, 61, 62, 64, 65, 95, 99, 101

Bhatia, R. (1997). *Matrix Analysis*, vol. 169. Springer New York. 32

Bickel, P. J., Ritov, Y., & Tsybakov, A. B. (2009). Simultaneous analysis of lasso and dantzig selector. *The Annals of Statistics*, *37*(4), 1705–1732. 50

Billups, S. C., Larson, J., & Graf, P. (2013). Derivative-free optimization of expensive functions with computational error using weighted regression. *SIAM Journal on Optimization*, *23*(1), 27–53. 64

Bitran, G., & Caldentey, R. (2003). An overview of pricing models for revenue management. *Manufacturing & Service Operations Management*, *5*(3), 203–229. 182

Blanchet, J., Gallego, G., & Goyal, V. (2016). A markov chain approximation to choice modeling. *Operations Research*, *64*(4), 886–905. 134

Börsch-Supan, A. (1990). On the compatibility of nested logit models with utility maximization. *Journal of Econometrics*, *43*(3), 373–388. 108

Bouhtou, M., Gaubert, S., & Sagnol, G. (2010). Submodularity and randomized rounding techniques for optimal experimental design. *Electronic Notes in Discrete Mathematics*, *36*, 679–686. 20

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge university press. 1, 62

Bront, J. J. M., Méndez-Díaz, I., & Vulcano, G. (2009). A column generation algorithm for choice-based network revenue management. *Operations Research*, *57*(3), 769–784. 134

Bubeck, S., & Cesa-Bianchi, N. (2012). Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, *5*(1), 1–122. 120, 148, 183

Bubeck, S., Eldan, R., & Lee, Y. T. (2017). Kernel-based methods for bandit convex optimization. In *Proceedings of annual ACM Symposium on the Theory of Computing (STOC)*. 64, 65

Bubeck, S., Munos, R., & Stoltz, G. (2009). Pure exploration in multi-armed bandits problems. In *Proceedings of the International conference on Algorithmic learning theory (ALT)*. 47, 75

Bubeck, S., Munos, R., Stoltz, G., & Szepesvári, C. (2011). X-armed bandits. *Journal of Machine Learning Research*, *12*(May), 1655–1695. 65

Bull, A. D. (2011). Convergence rates of efficient global optimization algorithms. *Journal of Machine Learning Research*, *12*(Oct), 2879–2904. 65

Caro, F., & Gallien, J. (2007). Dynamic Assortment with Demand Learning for Seasonal Consumer Goods. *Management Science*, *53*(2), 276–292. 107, 133, 183

Cartis, C., Gould, N. I., & Toint, P. L. (2012). On the oracle complexity of first-order and derivative-free algorithms for smooth nonconvex minimization. *SIAM Journal on Optimization*, *22*(1), 66–86. 64

Castro, R. (2014). Adaptive sensing performance lower bounds for sparse signal detection and support estimation. *Bernoulli*, *20*(4), 2217–2246. 75

Castro, R. M., & Nowak, R. D. (2008). Minimax bounds for active learning. *IEEE Transactions on Information Theory*, *54*(5), 2339–2353. 47, 74, 75, 181

Chaloner, K., & Verdinelli, I. (1995). Bayesian experimental design: A review. *Statistical Science*, *10*(3), 273–304. 19

Chaudhuri, K., Dasgupta, S., Kpotufe, S., & von Luxburg, U. (2014). Consistent procedures for cluster tree estimation and pruning. *IEEE Transactions on Information Theory*, *60*(12), 7900–7912. 65, 70

Chaudhuri, K., Kakade, S., Netrapalli, P., & Sanghavi, S. (2015). Convergence rates of active learning for maximum likelihood estimation. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 21

Chen, R., Menickelly, M., & Scheinberg, K. (2018a). Stochastic optimization using a trust-region method and random models. *Mathematical Programming*, *169*(2), 447–487. 64

Chen, S., Varma, R., Singh, A., & Kovačević, J. (2015a). Signal representations on graphs: Tools and applications. *arXiv preprint arXiv:1512.05406*. 21

Chen, X., Ma, W., Simchi-Levi, D., & Xin, L. (2016). Dynamic recommendation at checkout under inventory constraint. *Available at SSRN*. 134

Chen, X., Owen, Z., Pixton, C., & Simchi-Levi, D. (2015b). A statistical learning approach to personalization in revenue management. *Available at SSRN*. 134

Chen, X., & Wang, Y. (2018). A note on a tight lower bound for mnl-bandit assortment selection models. *Operations Research Letters*, *46*(5), 534–537. 163, 175

Chen, X., Wang, Y., & Zhou, Y. (2018b). Dynamic assortment optimization with changing contextual information. *arXiv preprint: arXiv:1810.13069*. 130, 131

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *The Annals of Mathematical Statistics*, *24*(4), 586–602. 15

Cheung, W. C., & Simchi-Levi, D. (2017). Thompson sampling for online personalized assortment optimization problems with multinomial logit choice models. *Available at SSRN*. 134

Chu, W., Li, L., Reyzin, L., & Schapire, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 135, 183

Chung, H., Ahn, H.-S., & Jasin, S. (2019). (Rescaled) multi-attempt approximation of choice model and its application to assortment optimization. *Production and Ooperations Management*, *28*(2), 341–353. 134

Cohen-Addad, V., & Kanade, V. (2017). Online optimization of smoothed piecewise constant functions. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 134

Cohn, D. A., Ghahramani, Z., & Jordan, M. I. (1996). Active learning with statistical models. *Journal of Artificial Intelligence Research*, *4*, 129–145. 181

Combes, R., & Proutiere, A. (2014). Unimodal bandits: Regret lower bounds and optimal algorithms. In *Proceedings of International Conference on Machine Learning (ICML)*. 134

Condat, L. (2015). Fast projection onto the simplex and the L1-ball. *Mathematical Programming*, *158*, 575–585. 8

Conn, A. R., Scheinberg, K., & Vicente, L. N. (2008a). Geometry of interpolation sets in derivative free optimization. *Mathematical Programming*, *111*(1-2), 141–172. 64

Conn, A. R., Scheinberg, K., & Vicente, L. N. (2008b). Geometry of sample sets in derivative-free optimization: polynomial regression and underdetermined interpolation. *IMA Journal of Numerical Analysis*, *28*(4), 721–748. 64

Conn, A. R., Scheinberg, K., & Vicente, L. N. (2009). Global convergence of general derivative-free trust-region algorithms to first-and second-order critical points. *SIAM Journal on Optimization*, *20*(1), 387–415. 64

Cope, E. W. (2009). Regret and convergence bounds for a class of continuum-armed bandit problems. *IEEE Transactions on Automatic Control*, *54*(6), 1243–1253. 134

Cover, T., & Thomas, J. (2006). *Elements of Information Theory*. Wiley. 98

Csiszar, I., & Körner, J. (2011). *Information theory: coding theorems for discrete memoryless systems*. Cambridge University Press. 161

Dani, V., Hayes, T. P., & Kakade, S. M. (2008). Stochastic Linear Optimization under Bandit Feedback. In *Proceedings of the Annual Conference on Learning Theory (COLT)*. 131, 135, 173

Daniely, A., Gonen, A., & Shalev-Shwartz, S. (2015). Strongly adaptive online learning. In *Proceedings of the International Conference on Machine Learning (ICML)*. 66

Dasgupta, S., Hsu, D. J., & Monteleoni, C. (2008). A general agnostic active learning algorithm. In *Proceedings of Advances in neural information processing systems (NIPS)*. 65

Davenport, M. A., Massimino, A. K., Needell, D., & Woolf, T. (2015). Constrained adaptive sensing. *IEEE Transactions on Signal Processing*, *64*(20), 5437–5449. 21

Davis, J. M., Gallego, G., & Topaloglu, H. (2014). Assortment optimization under variants of the nested logit model. *Operations Research*, *62*(2), 250–273. 118, 133, 182

Derman, C., Lieberman, G. J., & Ross, S. M. (1972). A sequential stochastic assignment problem. *Management Science*, *18*(7), 349–355. 182, 183

Désir, A., Goyal, V., Jagabathula, S., & Segev, D. (2016). Assortment optimization under the mallows model. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*. 134

Dhillon, P., Lu, Y., Foster, D., & Ungar, L. (2013). New sampling algorithms for fast least squares regression. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 21

Dodangeh, M., & Vicente, L. N. (2016). Worst case complexity of direct search under convexity. *Mathematical Programming*, *155*(1-2), 307–332. 64

Drineas, P., Mahoney, M. W., & Muthukrishnan, S. (2008). Relative-error CUR matrix decompositions. *SIAM Journal on Matrix Analysis and Applications*, *30*(2), 844–881. 21

Drineas, P., Mahoney, M. W., Muthukrishnan, S., & Sarlós, T. (2011). Faster least squares approximation. *Numerische Mathematik*, *117*(2), 219–249. 21

Duchi, J., Shalev-Shwartz, S., Singer, Y., & Chandra, T. (2008). Efficient projections onto the L1-

ball for learning in high dimensions. In *Proceedings of International Conference on Machine learning (ICML)*. 8

Duchi, J. C., Jordan, M. I., Wainwright, M. J., & Wibisono, A. (2015). Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, *61*(5), 2788–2806. 56, 64, 65

Ebert, D. S. (2003). *Texturing & modeling: a procedural approach*. Morgan Kaufmann. 74

Elmaghraby, W., & Keskinocak, P. (2003). Dynamic pricing in the presence of inventory considerations: Research overview, current practices, and future directions. *Management Science*, *49*(10), 1287–1309. 182, 183

Even-Dar, E., Mannor, S., & Mansour, Y. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, *7*(Jun), 1079–1105. 65

Fan, J., & Gijbels, I. (1996). *Local polynomial modelling and its applications*. CRC Press. 41

Fan, X., Grama, I., & Liu, Q. (2015). Exponential inequalities for martingales with applications. *Electronic Journal of Probability*, *20*(1), 1–22. 168

Farias, V. F., Jagabathula, S., & Shah, D. (2013). A Nonparametric Approach to Modeling Choice with Limited Data. *Management Science*, *59*(2), 305–322. 134

Fedorov, V. V. (1972). *Theory of optimal experiments*. Elsevier. 1, 19, 181

Filippi, S., Cappe, O., Garivier, A., & Szepesvári, C. (2010). Parametric bandits: The generalized linear case. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*. 129, 135, 173, 183

Flaxman, A. D., Kalai, A. T., & McMahan, H. B. (2005). Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings of the Annual ACM-SIAM symposium on Discrete algorithms (SODA)*. 53, 59, 62, 64, 65

Freedman, D. A. (1975). On tail probabilities for martingales. *The Annals of Probability*, *3*(1), 100–118. 168

Fukuda, K. (2004). From the zonotope construction to the minkowski addition of convex polytopes. *Journal of Symbolic Computation*, *38*(4), 1261–1272. 184

Gallego, G., Iyengar, G., Phillips, R., & Dubey, A. (2004). Managing flexible products on a network. Tech. Rep. CORC TR-2004-01, Department of Industrial Engineering and Operations Research, Columbia University. 133, 182

Gallego, G., & Topaloglu, H. (2014). Constrained assortment optimization for the nested logit model. *Management Science*, *60*(10), 2583–2601. 134

Ghadimi, S., & Lan, G. (2012). Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework. *SIAM Journal on Optimization*, *22*(4), 1469–1492. 49

Ghadimi, S., & Lan, G. (2013a). Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, II: shrinking procedures and optimal algorithms. *SIAM Journal on Optimization*, *23*(4), 2061–2089. 49

Ghadimi, S., & Lan, G. (2013b). Stochastic first- and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, *23*(4), 2341–2368. 64

Goldenshluger, A., & Zeevi, A. (2017). Optimal stopping of a random sequence with unknown distribution. Tech. rep., Workig paper, Graduate School of Business, Columbia University. 183

Golrezaei, N., Nazerzadeh, H., & Rusmevichientong, P. (2014). Real-time optimization of personalized assortments. *Management Science*, *60*(6), 1532–1551. 107, 134

Graham, R., & Sloane, N. (1980). Lower bounds for constant weight codes. *IEEE Transactions on Information Theory*, *26*(1), 37–43. 102

Grill, J.-B., Valko, M., & Munos, R. (2015). Black-box optimization of noisy functions with unknown smoothness. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 44, 65

Gupta, A., & Song, D. (1997). Lp-norm spherical distribution. *Journal of Statistical Planning and Inference*, *60*(2), 241–260. 89

Hall, E. C., & Willett, R. M. (2015). Online convex optimization in dynamic environments. *IEEE Journal of Selected Topics in Signal Processing*, *9*(4), 647–662. 66

Hanneke, S. (2007). A bound on the label complexity of agnostic active learning. In *Proceedings of the International Conference on Machine Learning (ICML)*. 65

Hanneke, S., et al. (2014). Theory of disagreement-based active learning. *Foundations and Trends® in Machine Learning*, *7*(2-3), 131–309. 181

Hazan, E. (2016). Introduction to online convex optimization. *Foundations and Trends in Optimization*, *2*(3-4), 157–325. 66

Hazan, E., & Kale, S. (2014). Beyond the regret minimization barrier: Optimal algorithms for stochastic strongly-convex optimization. *Journal of Machine Learning Research*, *15*, 2489–2512. 61, 102

Hazan, E., & Karnin, Z. (2015). Hard margin active linear regression. In *Proceedings of International Conference on Machine Learning (ICML)*. 21

Hazan, E., Klivans, A., & Yuan, Y. (2017). Hyperparameter optimization: A spectral approach. *arXiv preprint arXiv:1706.00764*. 65

Hazan, E., & Levy, K. (2014). Bandit convex optimization: Towards tight bounds. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*. 62, 63, 64, 65, 66, 94

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, *58*(301), 13–30. 78, 110, 112, 185, 186

Hsu, D., Kakade, S., & Zhang, T. (2012). A tail inequality for quadratic forms of subgaussian random vectors. *Electronic Communications in Probability*, *17*(52), 1–6. 66, 185

Hudson, H. M. (1978). A natural identity for exponential families with applications in multiparameter estimation. *The Annals of Statistics*, *6*(3), 473–484. 55

Ibragimov, I. A., & Has'minskii, R. Z. (1981). *Statistical Estimation: Asymptotic Theory*.

Springer-Verlag. 98, 99, 182

Jadbabaie, A., Rakhlin, A., Shahrampour, S., & Sridharan, K. (2015). Online optimization: Competing with dynamic comparators. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 66

Jamieson, K. G., Nowak, R., & Recht, B. (2012). Query complexity of derivative-free optimization. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*. 61, 102

Javanmard, A., & Montanari, A. (2014). Confidence intervals and hypothesis testing for high-dimensional regression. *Journal of Machine Learning Research*, *15*(1), 2869–2909. 50

Kaelbling, L. P., Littman, M. L., & Cassandra, A. R. (1998). Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, *101*(1-2), 99–134. 184

Kan, A. R., & Timmer, G. T. (1987a). Stochastic global optimization methods part I: Clustering methods. *Mathematical Programming*, *39*(1), 27–56. 65

Kan, A. R., & Timmer, G. T. (1987b). Stochastic global optimization methods part II: Multi level methods. *Mathematical Programming*, *39*(1), 57–78. 65

Khuri, A., Mukherjee, B., Sinha, B., & Ghosh, M. (2006). Design issues for generalized linear models: a review. *Statistical Science*, *21*(3), 376–399. 15

Kleinberg, R. D. (2005). Nearly tight bounds for the continuum-armed bandit problem. In *Proceedings of the Advances in Neural Information Processing Systems (NIPS)*. 65

Kök, A. G., Fisher, M. L., & Vaidyanathan, R. (2008). Assortment planning: Review of literature and industry practice. In *Retail supply chain management*, (pp. 99–153). Springer. 3, 107, 108, 182

Kök, A. G., & Xu, Y. (2011). Optimal and competitive assortments with endogenous pricing under hierarchical consumer choice models. *Management Science*, *57*(9), 1546–1563. 134

Kolda, T. G., Lewis, R. M., & Torczon, V. (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, *45*(3), 385–482. 64

Korostelev, A. P., & Tsybakov, A. B. (2012). *Minimax theory of image reconstruction*, vol. 82. Springer Science & Business Media. 44, 182

Krishnamurthy, A. (2015). *Interactive Algorithms for Unsupervised Machine Learning*. Ph.D. thesis, Carnegie Mellon University. 181

Lan, G. (2012). An optimal method for stochastic composite optimization. *Mathematical Programming*, *133*, 365–397. 49, 80

Lebedev, M. A., & Nicolelis, M. A. (2006). Brain–machine interfaces: past, present and future. *TRENDS in Neurosciences*, *29*(9), 536–546. 18

Leeds, D. D., Pyles, J. A., & Tarr, M. J. (2014). Exploration of complex visual feature spaces for object perception. *Frontiers in computational neuroscience*, *8*. 37

Lei, Q., Yen, I. E.-H., Wu, C.-y., Dhillon, I. S., & Ravikumar, P. (2017). Doubly greedy primal-dual coordinate descent for sparse empirical risk minimization. In *Proceedings of the International Conference on Machine Learning (ICML)*. 48

Lepski, O. V., Mammen, E., & Spokoiny, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selectors. *The Annals of Statistics*, *25*(3), 929–947. 43

Li, C., Jegelka, S., & Sra, S. (2017a). Polynomial time algorithms for dual volume sampling. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 21

Li, G., & Rusmevichientong, P. (2014). A greedy algorithm for the two-level nested logit model. *Operations Research Letters*, *42*(5), 319–324. 134, 182

Li, G., Rusmevichientong, P., & Topaloglu, H. (2015). The d-level nested logit model: Assortment and price optimization problems. *Operations Research*, *63*(2), 325–342. 118, 134

Li, L., Lu, Y., & Zhou, D. (2017b). Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of International Conference on Machine Learning (ICML)*. 129, 135, 164, 165, 173, 183

Lieb, E. H. (1973). Convex trace functions and the wigner-yanase-dyson conjecture. *Advances in Mathematics*, *11*(3), 267–288. 32

Liu, Q., & van Ryzin, G. (2008). On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, *10*(2), 288–310. 133

Lounici, K. (2008). Sup-norm convergence rate and sign concentration property of lasso and dantzig estimators. *Electronic Journal of Statistics*, *2*, 90–102. 50

Ma, P., Mahoney, M. W., & Yu, B. (2015). A statistical perspective on algorithmic leveraging. *Journal of Machine Learning Research*, *16*(1), 861–911. 21

Mackey, L., Jordan, M. I., Chen, R. Y., Farrell, B., & Tropp, J. A. (2014). Matrix concentration inequalities via the method of exchangeable pairs. *The Annals of Probability*, *42*(3), 906–945. 186

Mahajan, S., & van Ryzin, G. (2001). Stocking retail assortments under dynamic consumer substitution. *Operations Research*, *49*(3), 334–351. 133, 182

Malherbe, C., Contal, E., & Vayatis, N. (2016). A ranking approach to global optimization. In *Proceedings of the International Conference on Machine Learning (ICML)*. 65

Malherbe, C., & Vayatis, N. (2017). Global optimization of lipschitz functions. In *Proceedings of the International Conference on Machine Learning (ICML)*. 65

McFadden, D. (1980). Econometric models for probabilistic choice among products. *Journal of Business*, *53*(3), S13–S29. 108

McMahan, H. B. (2011). Follow-the-regularized-leader and mirror descent: Equivalence theorems and L1 regularization. In *Procedings of International Conference on Artificial Intelligence and Statistics (AISTATS)*. 11, 27

McMahan, H. B. (2017). A survey of algorithms and analysis for adaptive online learning. *Journal of Machine Learning Research*, *18*, 1–50. 63

Méndez-Díaz, I., Miranda-Bront, J. J., Vulcano, G., & Zabala, P. (2014). A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics*, *164*,

246–263. 134

Minsker, S. (2012). *Non-asymptotic bounds for prediction problems and density estimation*. Ph.D. thesis, Georgia Institute of Technology. 65

Minsker, S. (2013). Estimation of extreme values and associated level sets of a regression function via selective sampling. In *Proceedings of Conferences on Learning Theory (COLT)*. 44, 45, 65

Mokhtari, A., Shahrampour, S., Jadbabaie, A., & Ribeiro, A. (2016). Online optimization in dynamic environments: Improved regret rates for strongly convex problems. In *Proceedings of the IEEE Conference on Decision and Control (CDC)*. 66

Nabaee, M., & Labeau, F. (2012). Quantized network coding for sparse messages. In *Proceedings of the IEEE Statistical Signal Processing Workshop (SSP)*. 18

Nakamura, N., Seepaul, J., Kadane, J. B., & Reeja-Jayan, B. (2017). Design for low-temperature microwave-assisted crystallization of ceramic thin films. *Applied Stochastic Models in Business and Industry*. 37, 48

Naor, A. (2007). The surface measure and cone measure on the sphere of ℓ_p^n. *Transactions of the American Mathematical Society*, *359*(3), 1045–1079. 89, 93

Naor, A., & Romik, D. (2003). Projecting the surface measure of the sphere of ℓ_p^n. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics*, *39*(2), 241–261. 89, 91, 93

Nemirovski, A., Juditsky, A., Lan, G., & Shapiro, A. (2009). Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, *19*(4), 1574–1609. 49, 87

Nesterov, Y., & Nemirovskii, A. (1994). *Interior-point polynomial algorithms in convex programming*. SIAM. 62

Nesterov, Y., & Spokoiny, V. (2017). Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, *17*(2), 527–566. 64, 65

Newey, W. K. (1997). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics*, *79*(1), 147–168. 66

Newey, W. K., Hsieh, F., & Robins, J. M. (2004). Twicing kernels and a small bias property of semiparametric estimators. *Econometrica*, *72*(3), 947–962. 52

Nikolov, A. (2015). Randomized rounding for the largest simplex problem. In *Proceedings of the annual ACM symposium on Theory of computing (STOC)*. 20

Nikolov, A., & Singh, M. (2016). Maximizing determinants under partition constraints. In *Proceedings of the annual ACM symposium on Theory of computing (STOC)*. 20

Nikolov, A., Singh, M., & Tantipongpipat, U. T. (2018). Proportional volume sampling and approximation algorithms for A-optimal design. *arXiv preprint arXiv:1802.08318*. 20, 21

Pan, S. J., & Yang, Q. (2010). A survey on transfer learning. *IEEE Transactions on Knowledge and Data Engineering*, *22*(10), 1345–1359. 17

Pataki, G. (1998). On the rank of extreme matrices in semidefinite programs and the multiplicity of optimal eigenvalues. *Mathematics of Operations Research*, *23*(2), 339–358. 26

Polonik, W. (1995). Measuring mass concentrations and estimating density contour clusters-an excess mass approach. *The Annals of Statistics*, *23*(3), 855–881. 65

Powell, M. J. (2003). On trust region methods for unconstrained minimization without derivatives. *Mathematical Programming*, *97*(3), 605–623. 64

Powell, M. J. (2004). Least frobenius norm updating of quadratic models that satisfy interpolation conditions. *Mathematical Programming*, *100*(1), 183–215. 64

Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM. 1, 5, 6, 19, 181

Rachev, S. T., & Ruschendorf, L. (1991). Approximate independence of distributions on spheres and their stability properties. *The Annals of Probability*, *19*(3), 1311–1337. 89

Rakhlin, A. (2009). Lecture notes on online learning. *Draft*. 29

Raskutti, G., & Mahoney, M. (2015). Statistical and algorithmic perspectives on randomized sketching for ordinary least-squares. In *Proceedings of International Conference on Machine Learning (ICML)*. 21

Raskutti, G., Wainwright, M. J., & Yu, B. (2011). Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Transactions on Information Theory*, *57*(10), 6976–6994. 99

Rasmussen, C. E., & Williams, C. K. (2006). *Gaussian processes for machine learning*, vol. 1. MIT press Cambridge. 65

Reeja-Jayan, B., Harrison, K. L., Yang, K., Wang, C.-L., Yilmaz, A., & Manthiram, A. (2012). Microwave-assisted low-temperature growth of thin films in solution. *Scientific reports*, *2*. 37, 48

Rigollet, P., & Vert, R. (2009). Optimal rates for plug-in estimators of density level sets. *Bernoulli*, *15*(4), 1154–1178. 65

Rudelson, M., & Vershynin, R. (2007). Sampling from large matrices: An approach through geometric functional analysis. *Journal of the ACM*, *54*(4). 186

Rusmevichientong, P., Shen, Z.-J., & Shmoys, D. (2010). Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. *Operations Research*, *58*(6), 1666–1680. 107, 114, 115, 122, 133, 173, 183

Rusmevichientong, P., Shmoys, D., Tong, C., & Topaloglu, H. (2014). Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, *23*(11), 2023–2039. 134

Rusmevichientong, P., & Topaloglu, H. (2012). Robust assortment optimization in revenue management under the multinomial logit choice model. *Operations Research*, *60*(4), 865–882. 3, 107, 109, 134

Rusmevichientong, P., & Tsitsiklis, J. N. (2010). Linearly parameterized bandits. *Mathematics of Operations Research*, *35*(2), 395–411. 135, 183

Sabato, S., & Munos, R. (2014). Active regression by stratification. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 21

Saha, A., & Tewari, A. (2011). Improved regret guarantees for online smooth convex opti-

mization with bandit feedback. In *Proceedings of the International Conference on Artificial Intelligence and Statistics (AISTATS)*. 94

Sandryhaila, A., & Moura, J. M. (2014). Big data analysis with signal processing on graphs: Representation and processing of massive data sets with irregular structure. *IEEE Signal Processing Magazine*, *31*(5), 80–90. 18

Saure, D., & Zeevi, A. (2013). Optimal dynamic assortment planning with demand learning. *Manufacturing & Service Operations Management*, *15*(3), 387–404. 3, 107, 133

Schechtman, G., & Zinn, J. (1990). On the volume of the intersection of two l_p^n balls. *Proceedings of the American Mathematical Society*, *110*(1), 217–224. 89

Scheinberg, K., & Toint, P. L. (2010). Self-correcting geometry in model-based algorithms for derivative-free unconstrained optimization. *SIAM Journal on Optimization*, *20*(6), 3512–3532. 64

Schuchman, L. (1964). Dither signals and their effect on quantization noise. *IEEE Transactions on Communication Technology*, *12*(4), 162–165. 19

Settles, B. (2009). Active learning literature survey. Tech. rep., University of Wisconsin-Madison. 181

Shamir, O. (2013). On the complexity of bandit and derivative-free stochastic convex optimization. In *Proceedings of the annual Conference on Learning Theory (COLT)*. 64, 65

Shamir, O. (2017). An optimal algorithm for bandit and zero-order convex optimization with two-point feedback. *Journal of Machine Learning Research*, *18*(52), 1–11. 56, 59, 64, 65

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., & Vandergheynst, P. (2013). The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, *30*(3), 83–98. 18

Singh, A., Scott, C., & Nowak, R. (2009). Adaptive hausdorff estimation of density level sets. *The Annals of Statistics*, *37*(5B), 2760–2782. 65

Singh, M., & Xie, W. (2017). Approximate positive correlated distributions and approximation algorithms for d-optimal design. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 20

Sinz, F., & Bethge, M. (2010). Lp-nested symmetric distributions. *Journal of Machine Learning Research*, *11*(Dec), 3409–3451. 89

Snoek, J., Larochelle, H., & Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In *Proceedings of Advances in Neural Information Processing Systems (NIPS)*. 37

Song, D., & Gupta, A. (1997). l_p-norm uniform distribution. *Proceedings of the American Mathematical Society*, *125*(2), 595–601. 55, 89

Spielman, D. A., & Srivastava, N. (2011). Graph sparsification by effective resistances. *SIAM Journal on Computing*, *40*(6), 1913–1926. 22, 23

Srebro, N., Sridharan, K., & Tewari, A. (2011). On the universality of online mirror descent. In *Proceedings of Advances in neural information processing systems (NIPS)*. 49

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *The Annals of Statistics*, *9*(6), 1135–1151. 55

Su, X., & Zenios, S. A. (2005). Patient choice in kidney allocation: A sequential stochastic assignment model. *Operations Research*, *53*(3), 443–455. 182, 183

Summa, M. D., Eisenbrand, F., Faenza, Y., & Moldenhauer, C. (2015). On largest volume simplices and sub-determinants. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 20

Talluri, K., & van Ryzin, G. (2004). Revenue management under a general discrete choice model of consumer behavior. *Management Science*, *50*(1), 15–33. 133, 182

Talluri, K. T., & Van Ryzin, G. J. (2006). *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media. 182

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, *58*(1), 267–288. 50

Tibshirani, R. J. (2013). The lasso problem and uniqueness. *Electronic Journal of Statistics*, *7*, 1456–1490. 26

Tong, S., & Koller, D. (2001). Support vector machine active learning with applications to text classification. *Journal of Machine Learning Research*, *2*(Nov), 45–66. 181

Torczon, V. (1997). On the convergence of pattern search algorithms. *SIAM Journal on Optimization*, *7*(1), 1–25. 64

Toulias, T. L., & Kitsos, C. P. (2014). On the properties of the generalized normal distribution. *Discussiones Mathematicae Probability and Statistics*, *34*(1-2), 35–49. 55

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge University Press, 2nd ed. 116, 133

Tropp, J. (2015). An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, *8*(1-2), 1–230. 187

Tsybakov, A. B. (2009). *Introduction to nonparametric estimation.*. Springer Series in Statistics. Springer, New York. 44, 61, 75, 76, 98, 99, 148, 161, 175, 182

Ulu, E., Mccann, J., & Kara, L. B. (2017). Lightweight structure design under force location uncertainty. *ACM Transactions on Graphics*, *36*(4), 158. 5, 12, 13, 14

Van de Geer, S., Bühlmann, P., Ritov, Y., & Dezeure, R. (2014). On asymptotically optimal confidence regions and tests for high-dimensional models. *The Annals of Statistics*, *42*(3), 1166–1202. 50

van de Geer, S. A. (2000). *Empirical Processes in M-estimation*, vol. 6. Cambridge university press. 47, 68, 168

Van der Vaart, A. W. (1998). *Asymptotic statistics*, vol. 3. Cambridge university press. 14, 17, 38, 130, 170

van Ryzin, G., & Mahajan, S. (1999). On the relationships between inventory costs and variety benefits in retail assortments. *Management Science*, *45*(11), 1496–1509. 133, 182

Vandenberghe, L., & Boyd, S. (1996). Semidefinite programming. *SIAM Review*, *38*(1), 49–95.

7, 8

Vershynin, R. (2012). How close is the sample covariance matrix to the actual covariance matrix? *Journal of Theoretical Probability*, *25*(3), 655–686. 165

Vicente, L. N., & Custódio, A. (2012). Analysis of direct searches for discontinuous functions. *Mathematical Programming*, *133*(1-2), 299–325. 64

Victor, H. (1999). A general class of exponential inequalities for martingales and ratios. *The Annals of Probability*, *27*(1), 537–564. 81, 185

Wang, L. (2011). Smoothness, disagreement coefficient, and the label complexity of agnostic active learning. *Journal of Machine Learning Research*, *12*(Jul), 2269–2292. 181

Wang, R. (2013). Assortment management under the generalized attraction model with a capacity constraint. *Journal of Revenue and Pricing Management*, *12*(3), 254–270. 134

Wang, Y., & Singh, A. (2016). Noise-adaptive margin-based active learning and lower bounds under tsybakov noise condition. In *Proceedings of AAAI Conference on Artificial Intelligence (AAAI)*. 102, 181

Wang, Y., & Singh, A. (2018). Linear quantization by effective-resistance sampling. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. 19

Wang, Y., Ulu, E., Singh, A., & Kara, L. B. (2018). Efficient load sampling for worst-case structural analysis under force location uncertainty. In *ASME IDETC-DAC*. 12, 13, 14

Wang, Y., Yu, W. A., & Singh, A. (2017). On computationally tractable selection of experiments in regression models. *Journal of Machine Learning Research*, *18*(1–40). 10, 14

Williams, H. C. W. L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A: Economy and Space*, *9*, 285–344. 108

Woodruff, D. (2014). Sketching as a tool for numerical linear algebra. *Foundations and Trends in Theoretical Computer Science*, *10*(1–2), 1–157. 21

Wu, C. J., & Hamada, M. S. (2011). *Experiments: planning, analysis, and optimization*, vol. 552. John Wiley & Sons. 181

Yu, B. (1997). Assouad, fano, and le cam. In *Festschrift for Lucien Le Cam*, (pp. 423–435). Springer. 98, 99

Yu, J. Y., & Mannor, S. (2011). Unimodal bandits. In *Proceedings of International Conference on Machine Learning (ICML)*. 134

Yudin, D., & Nemirovskii, A. (1983). *Problem complexity and method efficiency in optimization*. John Wiley and Sons. 1, 49, 57, 62, 64, 65

Zabinsky, Z. B., & Smith, R. L. (1992). Pure adaptive search in global optimization. *Mathematical Programming*, *53*(1), 323–338. 65

Zenios, S. A., Chertow, G. M., & Wein, L. M. (2000). Dynamic allocation of kidneys to candidates on the transplant waiting list. *Operations Research*, *48*(4), 549–569. 182, 183

Zhang, C.-H., & Zhang, S. S. (2014). Confidence intervals for low dimensional parameters in

high dimensional linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *76*(1), 217–242. 50

Zhang, H. (2010). Partially observable markov decision processes: A geometric technique and analysis. *Operations Research*, *58*(1), 214–228. 184

Zhang, L., Yang, T., Jin, R., & Zhou, Z.-H. (2018). Strongly adaptive regret implies optimally dynamic regret. In *Proceedings of the International Conference on Machine Learning (ICML)*. 66

Zhou, Y., Huang, C., Jiang, T., & Cui, S. (2013). Wireless sensor networks and the internet of things: Optimal estimation with nonuniform quantization and bandwidth allocation. *IEEE Sensors Journal*, *13*(10), 3568–3574. 18

Zinkevich, M. (2003). Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the International Conference on Machine Learning (ICML)*. 29, 66