Selective Data Acquisition in Learning and Decision Making Problems

Yining Wang

MARCH 2019
CMU-ML-19-101

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.

This research was sponsored by the National Science Foundation awards IIS1252412 and CCF1563918, the Air Force Office of Scientific Research award FA95501410285, and the Air Force Research Laboratory award FA87501720212.

Copyright © 2019 Yining Wang
To my family
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.
Acknowledgments

First and foremost, I would like to thank my advisor Aarti Singh, who has placed an irreplaceable role in shaping my research interests, ability and methodology. Under her guidance and encouragement, I have learnt a lot and developed a passion for the field of statistical machine learning. She encourages me to pursue research questions that truly interest me, and is always there when I need helps or additional guidance in identifying and approaching some challenging topics. Her support, guidance, wisdom, and encouragement all helped shape this thesis and much of my work, and they were all instrumental to my success in my PhD study.

I am also thankful to many other wonderful faculty members at Machine Learning Department and Department of Statistics, including Sivaraman Balakrishnan, Larry Wasserman, who served on my thesis committee, and Ruslan Salakhutdinov, Pradeep Ravikumar, Nina Balcan, Ryan Tibshirani, with whom I had collaborations. I am grateful to the wonderful instructors of the PhD courses I took: Eric Xing, Barnabas Poczos, Roy Maxion, Javier Pena, Jiashun Jin, Alessandro Rinaldo.

One chapter of this thesis is devoted to sequential decision making and its applications in online revenue management problems. I thank Xi Chen from New York University, Assaf Zeevi from Columbia University and Yuan Zhou from University of Illinois at Urbana-Champaign for pointing me to this exciting new area and helping identifying and solving interesting questions in online revenue management with dynamic environments.

Throughout the course of my PhD study, I am truly fortunate to have many collaborators both in the machine learning department and outside, whose help to my research cannot be over-emphasized. They are Simon Du, Yu-Xiang Wang, Zeyuan Allen-Zhu, Yuanzhi Li, Bo Li, Adams Wei Yu, Anima Anandkumar, Alex Smola, Hsiao-Yu Fish Tung, and many others.

I am also grateful to my mentors of my summer internships during my PhD study. They are Kevin Roundy for my first-year internship at Symantec Research Labs, and Denny Zhou, Zhong Wang for my third-year internship at Microsoft Research Redmond. I have learnt a lot from their experiences and the potential application of statistical machine learning in industrial security and artificial intelligence systems.

Finally, I am eternally thankful my parents and my wife for their enduring and unconditional love and support. You are truly the best I have ever had in my life.
# Contents

1 Introduction .................................................. 1
   1.1 Computationally tractable experimental design methods ............... 1
   1.2 Selective queries in nonparametric optimization ....................... 2
   1.3 Dynamic assortment planning .................................. 2

2 Computationally tractable experimental design methods ................. 5
   2.1 Backgrounds and optimality criteria ................................ 5
   2.2 The continuous relaxation framework ................................ 7
      2.2.1 The semidefinite programming (SDP) approach .................... 7
      2.2.2 The entropic mirror descent approach ............................ 8
   2.3 Rounding techniques .......................................... 8
      2.3.1 Sampling based techniques .................................... 9
      2.3.2 Greedy removal techniques for A-optimality .................... 10
      2.3.3 Greedy swapping techniques .................................. 11
   2.4 Application: 3D lightweight structure optimization .................. 12
      2.4.1 Background .................................................. 12
      2.4.2 Method .................................................... 13
      2.4.3 Experimental settings ....................................... 14
      2.4.4 Results and discussion ...................................... 14
   2.5 Extensions .................................................. 14
      2.5.1 Generalized linear models ................................... 14
      2.5.2 Transfer learning and Delta’s method .......................... 17
      2.5.3 Quantized linear regression .................................. 18
   2.6 Summary and related works ...................................... 19
   2.7 Proofs ...................................................... 21
      2.7.1 Proof of Theorem 1 .......................................... 21
      2.7.2 Proof of Theorem 2 .......................................... 25
      2.7.3 Proof of Lemma 1 ............................................ 27
      2.7.4 Proof of Theorem 3 .......................................... 32
      2.7.5 Proof of Lemma 2 ............................................ 36

3 Selective queries in nonparametric optimization ......................... 37
   3.1 Nonparametric optimization: local minimax rates ...................... 38
      3.1.1 Local minimax rates ........................................ 38
4 Dynamic assortment planning

4.1 The plain multinomial logit model
   4.1.1 Popular assortments, level sets, and a potential function
   4.1.2 The trisection algorithm and its regret
   4.1.3 Improved regret with adaptive confidence intervals
   4.1.4 Lower bound: the uncapacitated setting
   4.1.5 Lower bound: the capacitated setting
   4.1.6 Numerical results

4.2 The nested multinomial logit model
   4.2.1 Assortment space reductions
   4.2.2 A nested singleton model
   4.2.3 UCB-based dynamic assortment planning policies
   4.2.4 Policies with an improved \( N \) dependency
   4.2.5 A regret lower bound
   4.2.6 Numerical results

4.3 The linear contextual logit model
   4.3.1 An MLE-UCB policy and its regret
   4.3.2 A regret lower bound
   4.3.3 Numerical results

4.4 Summary and related works

4.5 Proofs of results in Sec. 4.1
   4.5.1 Proof of Lemma 52
   4.5.2 Proof of Lemma 53
   4.5.3 Proof of Lemma 54
   4.5.4 Proof of Theorem 13
   4.5.5 Proof of Lemma 55
   4.5.6 Proof of Theorem 14
   4.5.7 Proof of Theorem 15
   4.5.8 Proof of Theorem 16

4.6 Proofs of results in Sec. 4.2
   4.6.1 Proof of Theorem 17
   4.6.2 Proof of Lemma 59
   4.6.3 Proof of Theorem 18

4.7 Proofs of results in Sec. 4.3
   4.7.1 Proof of Theorem 17
   4.7.2 Proof of Theorem 20

5 Conclusion and discussion

5.1 Benefits of selective or active data acquisition
5.2 Importance of optimality and minimax analysis
5.3 Unification of data analysis and decision making
5.4 Future directions
   5.4.1 Regret analysis for dynamic programming with partial information
   5.4.2 POMDP and reinforcement learning
List of Figures

2.1 Stress distributions on designed structures under external forces. 

2.2 Example test structures with complex geometries. Fixed boundary conditions and contact regions are indicated in blue and red, respectively.

2.3 Sampled force locations \( S \) using the K-MEANS algorithm (top row) versus our proposed algorithm (bottom row), for \( k = 100 \).

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

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, suboptimal 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.

3.2 Sparse quadratic optimization with identity quadratic term.

3.3 Sparse quadratic optimization with polynomial decay of eigenvalues.

3.4 Sparse fourth-degree polynomial optimization with identity quadratic term.
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 $F_\lambda$ with intermediate values $\lambda = 0.25$ and $\lambda = 0.75$. For the sake of better visualization, the regions $0 \leq x \leq \sqrt{h} (I_1)$ and $\sqrt{h} \leq x \leq 2\sqrt{h} (I_2)$ 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 $\lambda$ first increase and then decrease, or vice versa, depending on the coding $i_j \in \{0, 1\}$ for the particular batch. Also note that $\lambda$ will never be over 0.75 nor under 0.25 throughout the entire construction of the function sequence.

Illustration of the potential function $F(\theta)$, the important quantities $F^*$ and $\hat{\theta}$, and their properties.

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.

Accumulated regret of our policy with $M = 5$ nests, varying the number of items per nest $N$ and the granularity parameter $\delta$.

Illustration of the performance of MLE-UCB.

Average regret of MLE-UCB for various $d$'s.

Average regret of MLE-UCB for various $N$'s.
List of Tables

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. . . . . . . . . . . . . . 15

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. . . . . . . . . . . 115

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. . . . . . . . . . . . . . . . . . . . . . . . . . . . 127

4.3 Adversarial construction of two types of nests. The revenue parameter $\rho$ is set to $\rho = \frac{9\sqrt{2}}{1 + \sqrt{2}} \approx 0.694774$. . . . . . . . . . . . . . . . . . . . . . . . . . . . 160
# List of Algorithms

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The projected entropic mirror descent algorithm for solving Eq. (2.5).</td>
<td>7</td>
</tr>
<tr>
<td>2</td>
<td>Projection onto the probabilistic simplex with box constraint</td>
<td>9</td>
</tr>
<tr>
<td>3</td>
<td>Sampling based experiment selection.</td>
<td>9</td>
</tr>
<tr>
<td>4</td>
<td>Greedy experiment selection.</td>
<td>10</td>
</tr>
<tr>
<td>5</td>
<td>A swapping algorithm for rounding</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>Bit allocation algorithm by leverage score sampling</td>
<td>20</td>
</tr>
<tr>
<td>7</td>
<td>Successive elimination for low-dimensional nonparametric optimization</td>
<td>41</td>
</tr>
<tr>
<td>8</td>
<td>The trisection algorithm.</td>
<td>111</td>
</tr>
<tr>
<td>9</td>
<td>EXPLORE Subroutine: exploring a certain revenue level $\theta$</td>
<td>111</td>
</tr>
<tr>
<td>10</td>
<td>The meta-policy $\pi$ built upon policy $\pi^d$ under the nested singleton model.</td>
<td>120</td>
</tr>
<tr>
<td>11</td>
<td>The upper confidence bound (UCB) policy for dynamic assortment planning under the nested singleton model in Eq. (4.17).</td>
<td>121</td>
</tr>
<tr>
<td>12</td>
<td>The meta-policy $\pi$ built upon policy $\pi^d$ and the discretization argument under the nested singleton model.</td>
<td>125</td>
</tr>
<tr>
<td>13</td>
<td>The MLE-UCB policy for dynamic assortment optimization</td>
<td>129</td>
</tr>
</tbody>
</table>
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 a subset $X \subseteq \mathcal{X}$ under resource constraints such as $|X| \leq k$ 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, under 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 \( i \)th 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 probabilistic model

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

where \( \theta_0 \) 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

\[ v_i = \exp \{x_i^T \theta_0\}. \]

Unlike stationary settings where \( \theta_0 \) is perfectly known and the assortment planning problem is merely a combinatorial optimization one (see, e.g., Anderson et al. (1992);
Kök et al. (2008)), under dynamic settings the preference parameters \( \theta_0 \) are unknown and have to be inferred or estimated on the fly from customers’ purchasing actions \( \{ i_t \} \) (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

$$y = Z\beta_0 + \xi,$$

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 $\beta_0$ is the ordinary least squares (OLS) estimator, of the form $\hat{\beta} = (Z^T Z)^{-1} Z^T y$. By simple algebra it is easy to check that the estimation error $\hat{\beta} - \beta_0$ is a centered
Gaussian random vector:

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

If the variance of $\hat{\beta} - \beta$ is to be minimized (i.e., minimizing $\mathbb{E}\|\hat{\beta} - \beta\|_2$), one should seek design subset $Z$ with the smallest $\text{tr}(\Sigma^{-1})$. Another popular objective is to maximize the determinant of $\Sigma$, 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 $f : \mathbb{S}_p^+ \rightarrow \mathbb{R}^+$ that maps a $p$-dimensional positive definite matrix $\Sigma$ 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): $f_A(\Sigma) = \frac{1}{p} \text{tr}(\Sigma^{-1})$;
- **D-optimality** (Determinant): $f_D(\Sigma) = (\det |\Sigma|)^{-1/p}$;
- **T-optimality** (Trace): $f_T(\Sigma) = p/\text{tr}(\Sigma)$;
- **E-optimality** (Eigenvalue): $f_E(\Sigma) = \|\Sigma^{-1}\|_{op} = \lambda_{\max}(\Sigma^{-1})$;
- **V-optimality** (Variance): $f_V(\Sigma) = \frac{1}{n} \sum_{i=1}^n x_i^T \Sigma^{-1} x_i$;
- **G-optimality**: $f_G(\Sigma) = \max_i x_i^T \Sigma^{-1} x_i$.

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

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

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 $f : \mathbb{S}_p^+ \rightarrow \mathbb{R}$ is regular if it satisfies the following properties:

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

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

6
Algorithm 1: The projected entropic mirror descent algorithm for solving Eq. (2.5).

**Require:** function \( \min_\omega \tilde{F}(\omega) \) 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 \( g^{(t)} \in \partial \tilde{F}(\omega^{(t)}) \);
4:   Update: \( \omega^{(t+1/2)} \propto \omega^{(t)} \exp \{-\eta_t g^{(t)} \} \), 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 \( \hat{\omega} := \frac{1}{T} \sum_{t=0}^{T-1} \omega^{(t)} \).

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

It can be verified that 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 \( F(\pi^*) \leq F(s^*) \) 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 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):

\[
\pi^* \in \arg \min_{\pi} F(\pi) = \arg \min_{\pi_1, \cdots, \pi_n} \left( \sum_{i=1}^{n} \pi_i x_i x_i^T \right) \quad \text{s.t.} \quad 0 \leq \pi_i \leq b, \quad \sum_{i=1}^{n} \pi_i \leq k. \tag{2.4}
\]

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 \( F(\pi^*) \leq F(s^*) \) 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 \( A(\pi) = \sum_{i=1}^{n} \pi_i x_i x_i^T \), which is a \( p \times p \) positive semidefinite matrix. By definition, \( f(\pi; X) = \sum_{j=1}^{p} e_j A(\pi)^{-1} e_j \), where \( e_j \) is the \( p \)-dimensional vector with only \( p \)th
coordinate being 1. Subsequently, Eq. (2.4) is equivalent to the following SDP problem:

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

where

\[
B_j = \begin{bmatrix} A(\pi) & e_j \\ e_j^T & t_j \end{bmatrix}, \quad j = 1, \cdots, p.
\]

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

\[
\min_{\omega} \tilde{F}(\omega) := \min_{\omega} f \left( \sum_{i=1}^{n} \omega_i x_i x_i^T \right) \quad \text{s.t.} \quad 0 \leq \omega_i \leq b/k, \quad \sum_{i=1}^{k} \omega_i = 1, \quad (2.5)
\]

by the change of variables \( \omega_i = \pi_i / k \) and noting that the \( \sum_{i=1}^{n} \pi_i \leq k \) constraint in Eq. (2.4) must bind, meaning that the optimal solution \( \pi^* \) 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 Kullback-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 \( \omega_i \leq 1/k \), 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 \( \pi^* \) 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 \( b \in [1/n, 1] \).
Ensure: an output \( \omega' \in \Delta_n \) such that \( \|\omega'\|_\infty \leq b \).
\( \Rightarrow \Delta_n = \{ x \in \mathbb{R}^n : x_i \geq 0, \sum_{i} x_i = 1 \} \)
\( \Rightarrow \omega' = \arg\min_{\omega \in \Delta_n, y_i \leq b}KL(y\|\omega) \)
\( \Rightarrow \) where \( KL(y\|\omega) := \sum_{i} y_i \log \frac{y_i}{\omega_i} \)

1: Sort \( \omega \) in descending order: \( \omega_1 \geq \omega_2 \geq \cdots \geq \omega_n \geq 0 \);
2: if \( \omega_1 \leq b \) then return \( \omega \);
3: \( KL_1 \leftarrow b \log(b/\omega_1), \ Z_1 \leftarrow 1 - \omega_1, \ KL_{\text{opt}} \leftarrow \infty, \ \eta_{\text{opt}} \leftarrow 0, \ \text{and} \ C_{\text{opt}} \leftarrow 0; \)
4: for \( q \) from 2 to \( n \) do
5: \( C \leftarrow (1 - b(q - 1))/Z_{q-1}; \)
6: if \( C > 0 \) and \( C w_q \leq 0 \) and \( (KL_{q-1} + C \log(C) \cdot Z_{q-1} - KL_{\text{opt}}) \) then
7: \( KL_{\text{opt}} \leftarrow KL_{q-1} + C \log(C) \cdot Z_{q-1}, \ \eta_{\text{opt}} \leftarrow w_q, \ \text{and} \ C_{\text{opt}} \leftarrow C; \)
8: end if
9: \( KL_q \leftarrow b \log(b/\omega_q), \ Z_q \leftarrow Z_{q-1} - \omega_q; \)
10: end for
11: Set \( \omega'_i \leftarrow b \) if \( \omega_i \geq \eta_{\text{opt}} \), and \( \omega'_i \leftarrow C_{\text{opt}} \omega_i \) if \( \omega_i < \eta_{\text{opt}} \);
12: return \( \omega'. \)

Algorithm 3: Sampling based experiment selection.

Input: \( X \in \mathbb{R}^{n \times p} \), optimal solution \( \pi^* \), target subset size \( k \).
Output: \( \tilde{S} \subseteq [n] \), a selected subset of size at most \( k \).
Initializations: \( t = 0, S_0 = \emptyset, R_0 = \emptyset. \)
1. With replacement: sample \( i_t \sim P^{(1)} \); set \( w_t = [\pi^*_i/(kp^{(1)}_i)]; \)
   Without replacement: pick random \( i_t \notin R_{t-1} \); sample \( w_t \sim \text{Bernoulli}(kp^{(2)}_i). \)
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 \( \tilde{S} = S_T. \)

2.3.1 Sampling based techniques

Perhaps the simplest idea of rounding \( \pi^* \) into a discrete solution is to sample according to an appropriately normalized categorical distribution based on \( \pi^* \). More specifically, define \( \Sigma_* := X^T \text{diag}(\pi^*)X = \sum_{i=1}^n \pi_i^* x_i x_i^T \) and
\[
P^{(1)}: \quad p_j^{(1)} = \pi_j^* x_j^T \Sigma_* ^{-1} x_j / p, \quad \text{with replacement (} b = k); \]
\[
P^{(2)}: \quad p_j^{(2)} = \pi_j^*/k, \quad \text{without replacement (} b = 1). \]

Note that both \( \{p_j^{(1)}\}_{j=1}^n \) and \( \{p_j^{(2)}\}_{j=1}^n \) sum to one because \( \sum_{j=1}^n \pi_j^* = k \) and \( \sum_{j=1}^n \pi_j^* x_j x_j^T \Sigma_* ^{-1} x_j = \text{tr}((\sum_{j=1}^n \pi_j^* x_j x_j^T) \Sigma_* ^{-1}) = \text{tr}(\Sigma_* \Sigma_* ^{-1}) = p. \)

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 \( \mathbb{E}X_{i_t}^T X_{i_t} = X^T \text{diag}(\pi^*/k)X \) and \( \mathbb{E}[w_{i_t}] = 1 \), for both with and without replacement settings. Note that \( \|p^{(2)}\|_\infty \leq 1/k \) by feasibility constraints and hence \( \text{Bernoulli}(kp^{(2)}_i) \) is a valid distribution for all \( i_t \in [n]. \) For the with replacement setting, by weak law of large numbers, \( X_{\tilde{S}}^T X_{\tilde{S}} \xrightarrow{P} X^T \text{diag}(\pi^*)X \) 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 \( k \leq |S_0| \).

**Output:** \( \hat{S} \subseteq [n] \), a selected subset of size \( k \).

Initialization: \( t = 0 \).
1. Find \( j^* \in S_t \) such that \( \text{tr}[ (X_{S_t \setminus \{j^*\}} X_{S_t \setminus \{j^*\}})^{-1} ] \) 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 \( \hat{S} = S_t \).

and hence \( \text{tr}[ (X_{S_t}^\top X_S)^{-1} ] \xrightarrow{p} \mathbb{E}[f(\pi^*; X)] \) by continuous mapping theorem.

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

**Theorem 1.** Suppose the following conditions hold:

\[
\begin{align*}
\text{with replacement} & : \quad p \log k / k = O(1); \\
\text{without replacement} & : \quad \|\Sigma^{-1}\|_2 \kappa(\Sigma) \|X\|_\infty^2 \log p = O(1). 
\end{align*}
\]

Then with probability at least 0.8 the subset \( \hat{S} \) output by Algorithm 3 satisfies

\[
F(\hat{S}) \leq O(1) \cdot \min_{|s|_1 \leq k, |s|_\infty \leq b, \pi \in \mathbb{N}^n} F(s), \quad b \in \{1, k\}
\]

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 \( f_A(\Sigma) = \text{tr}(\Sigma^{-1}) \) 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 \( \hat{S} \) 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

\[
F_A(\hat{S}) \leq \left( 1 + \frac{p(p+1)}{2(k-p+1)} \right) \cdot \min_{|s|_1 \leq k, |s|_\infty \leq 1, \pi \in \mathbb{N}^n} F_A(s).
\]

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 \( W := \sum_{i=1}^{n} \pi_i x_i x_i^T \). The pre-processing step is to whiten the design points \( \{x_i\}_{i=1}^{n} \) by taking

\[
\tilde{x}_i := W^{-1/2} x_i. \tag{2.6}
\]

Note that after pre-processing, the transformed data points \( \{\tilde{x}_i\}_{i=1}^{n} \) satisfies \( \sum_{i=1}^{n} \pi_i \tilde{x}_i \tilde{x}_i^T = I_{p \times p} \), 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

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

While the definitions of the potential function \( \phi \) 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 \( \phi \) 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 \( Z_t := Z_0 + \sum_{t'=1}^{t} u_{t'} u_{t'}^T - v_{t'} v_{t'}^T \). If further \( v_t^T A_{Z_{t-1}}^{1/2} v_t < 1/2\alpha \) holds for all \( t \), then

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

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 \( \phi(\tilde{x}_j, \tilde{x}_i; Z) \), where \( Z = \sum_{i \in S} \tilde{x}_i \tilde{x}_i^T \).
Algorithm 5 A swapping algorithm for rounding

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

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 \( \phi(\tilde{x}_{j_t}, \tilde{x}_i; Z_{t-1}) \) can be found in \( O(n + k) \) instead of \( O(nk) \) time by separately maximizing and minimizing \( \phi_+(\tilde{x}_{j_t}; Z_{t-1}) \) and \( \phi_-(\tilde{x}_i; Z_{t-1}) \) as shown in Steps 6 and 7 in Algorithm 5, because the potential \( \phi \) decomposes additively. In step 5 of Algorithm 5, the unique real number \( c_t \in \mathbb{R} \) such that \( \text{tr}(A_{Z_{t-1}}) = 1 \) can be found by a binary search, because \( \text{tr}[(c_t I + \alpha Z_{t-1})^{-2}] \) 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 \( k \geq \frac{5p}{\varepsilon^2} \) for some \( \varepsilon \in (0, 1/6] \). Then for any regular \( f \), \( \hat{s} \in \{0, 1\}^n \) output by Algorithm 5 has size \( \sum^n_{i=1} \hat{s}_i \leq k \) and satisfies

\[
F(\hat{s}) \leq (1 + 6\varepsilon)F(s^*) = (1 + 6\varepsilon) \max_{\hat{s} \in \{0, 1\}^n, \|s\|_1 = k} F(s).
\]

### 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.
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 \( S \subseteq [n] \), \( |S| \leq k \ll n \) 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

\[
y_i = x_i^T \beta_0 + \xi_i,
\]

where \( \beta_0 \) is a \( p \)-dimensional unknown regression model and \( \{\xi_i\}_{i=1}^n \) are noise variables.

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

\[
\min_s f (\sum_{i=1}^n s_i x_i x_i^T) \quad s.t. \quad s_i \in \{0, 1\}, \sum_{i=1}^n s_i \leq k.
\]

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 \( \beta_0 \) is then estimated by ordinary least squares \( \hat{\beta} = (\sum_{i \in S} x_i x_i^T)^{-1} (\sum_{i \in S} y_i x_i) \), and predictions on the other external force locations are produced by \( \hat{y}_i = x_i^T \hat{\beta} \) for \( i \notin S \). The force locations \( i \in [n] \) are then ranked in descending order according to \( \{\hat{y}_i\}_{i=1}^n \), 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).
2.4.3 Experimental settings

We evaluate the performance of our algorithm on three test structures (FERTILITY, ROCKINGCHAIR 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 $S \subseteq [n]$, $|S| \leq k$. 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 $\{\hat{y}_i\}_{i=1}^n$ 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 $\mu(x) = E[Y|x]$ satisfies $g(\mu(x)) = \eta = x^T \beta_0$ for some known link function $g : \mathbb{R} \rightarrow \mathbb{R}$. Under regularity conditions (Van der Vaart, 1998), the maximum-likelihood
Note that under regularity conditions $-\mathbb{E} \frac{\partial^2 \log p(y_i; \tilde{\eta}_i)}{\partial \tilde{\eta}_i^2} = \mathbb{E} \left( \frac{\partial \log p(y_i; \tilde{x}_i)}{\partial \tilde{\eta}_i} \right)^2$ 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 \( k = 100 \).

\[
\tilde{X} = [\tilde{x}_1, \ldots, \tilde{x}_n]^\top
\]

for generalized linear models. Below we consider two generalized linear model examples and derive explicit forms of \( \tilde{X} \).

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

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

Simple algebra yields

\[
\tilde{x}_i = \sqrt{\frac{e^{\tilde{\eta}_i}}{(1 + e^{\tilde{\eta}_i})^2}}x_i,
\]

where \( \tilde{\eta}_i = x_i^\top \tilde{\beta} \).

**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 \( \lambda = e^{\eta_i} = e^{x_i^\top \beta} \). The likelihood model is formally defined as

\[
p(y_i = r; \eta_i) = \frac{e^{\eta_i}r e^{-\eta_i}}{r!}, \quad r = 0, 1, 2, \ldots
\]

Simple algebra yields

\[
\tilde{x}_i = \sqrt{e^{\tilde{\eta}_i}}x_i,
\]

where \( \tilde{\eta}_i = x_i^\top \tilde{\beta} \).
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 $\tilde{\beta}_n = (X^TX)^{-1}X^Ty$ be the OLS estimate of $\beta_0$. If $\nabla g$ is continuously differentiable and $\tilde{\beta}_n$ is consistent, then by the classical delta’s method (Van der Vaart, 1998)

$$E\{(g(\tilde{\beta}_n) - g(\beta_0))^2\} = \sigma^2 \text{tr}(G_0(X^TX)^{-1} \nabla g(\beta_0)^\top) \approx (1 + o(1)) \sigma^2 \text{tr}(G_0(X^TX)^{-1})$$

where $G_0 = \nabla g(\beta_0)^\top \nabla g(\beta_0)$. If $G_0$ depends on the unknown parameter $\beta_0$ then the design dependence problem again exists, and a locally optimal solution can be obtained by replacing $G_0$ in the objective function with $\tilde{G} = \nabla g(\tilde{\beta})^\top \nabla g(\tilde{\beta})$ for some initial estimate $\tilde{\beta}$ of $\beta_0$.

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

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

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

**Example:** prediction error. In some application scenarios the prediction error $\|Z\tilde{\beta} - Z\beta_0\|^2_2$ rather than the estimation error $\|\tilde{\beta} - \beta_0\|^2_2$ is of interesting, either because the linear model is used mostly for prediction or component of the underlying model $\beta_0$ 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 $\beta_0$.

Suppose $Z \in \mathbb{R}^{mxp}$ is a known full-rank data matrix upon which predictions are seeked, and define $\hat{\Sigma}_Z = \frac{1}{m}Z^TZ > 0$ to be the sample covariance of $Z$. Our algorithmic framework
as well as its corresponding analysis remain valid for such prediction problems with transform \( x_i \mapsto \tilde{\Sigma}_Z^{-1/2} x_i \). 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 \( x_i \mapsto \tilde{\Sigma}_Z^{-1/2} x_i \).

### 2.5.3 Quantized linear regression

Consider the noiseless linear signal model

\[
y = X \beta_0
\]

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 \( \beta_0 \) 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 \( k \in \mathbb{N}, k \geq n \), find a bit allocation \( k = (k_1, \ldots, k_n) \in \mathbb{N}^n \), \( k_1 + \cdots + k_n \leq k \) such that the mean square error between the recovered signal \( \hat{\beta}_k \) and the true signal \( \beta_0 \) is minimized.

Suppose a bit allocation strategy \( k = (k_1, \ldots, k_n) \in \mathbb{N}_+^n \) is given, such that \( \sum_{i=1}^n k_i \leq n \). Let \( \text{round}(\cdot) \) 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 \( y_i = x_i^T \beta_0 \) with \( k_i \in \mathbb{N}_+ \) binary bits of measurement can then be expressed as

\[
\tilde{y}_i = 2^{-(k_i - 1)} \cdot \text{round} \left[ 2^{k_i - 1} \left( \frac{y_i}{M} + \delta_i \right) \right]
\]

where \( M := \max_{1 \leq i \leq n} |x_i^T \beta_0| \) 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).
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 $\tilde{y}_i$ 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 $\tilde{y}_i$ 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 $\beta_0$ which we define as follows:

$$
\hat{\beta}_k \in \arg\min_{\beta \in \mathbb{R}^p} \sum_{i:k_i > 0} 4^{k_i+1}(\tilde{y}_i - x_i^T \beta)^2.
$$

(2.14)

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

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

$$
\mathbb{E}\|\hat{\beta}_k - \beta_0\|^2_2 \leq M^2 \cdot \text{tr} \left[ \sum_{i:k_i > 0} 4^{k_i+1}x_i x_i^T \right]^{-1}.
$$

(2.15)

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

$$
\min_{\pi = (\pi_1, \ldots, \pi_n) \in \mathbb{R}^n} \text{tr} \left[ \left( \sum_{i=1}^n (4^{\pi_i} - 1)x_i x_i^T \right)^{-1} \right] \quad \pi_i \geq 0, \|\pi\|_1 \leq k_0.
$$

(2.16)

After the continuous optimization is solved, with (approximately) optimal solution $\pi$, 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.
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 rise 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 $P = NP$. 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-optimality 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 $f(\Sigma^n_{i=1} x_i x_i^\top)$. This ratio is
tightly 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 \(\beta_0\), 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 \(\hat{\Sigma}_S = X_S^T X_S\). Suppose the following conditions hold:

\[
\begin{align*}
\text{With replacement} : & \quad p \log T/T = O(1); \\
\text{Without replacement} : & \quad \|\Sigma_a^{-1}\|_2 K(\Sigma_a)\|X\|_\infty^2 \log p = O(T/n).
\end{align*}
\]

Then with probability at least 0.9 the following holds:

\[
\begin{align*}
z^T \hat{\Sigma}_S z & \geq K_T z^T \Sigma_a z, \quad \forall z \in \mathbb{R}^p, \quad (2.17)
\end{align*}
\]

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 \(\min_{|S| \leq k} f(\sum_{i \in S} x_i x_i^T)\) can be indeed \(O(n/k)\) times larger than \(f(\sum_{i=1}^n x_i x_i^T)\) (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 $T \geq \delta k$ for with replacement and $T \geq \delta n$ for without replacement.

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

Combining Lemmas 3 and 4 with $\delta = 0.1$ and note that $T \leq k$ almost surely (because $w_t \geq 1$), 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 $\Phi = \text{diag}(\pi^*)$ and $\Pi = \Phi^{1/2}X\Sigma^{-1}_a X^\top \Phi^{1/2} \in \mathbb{R}^{n \times n}$. The following proposition lists properties of $\Pi$:

Proposition 1 (Properties of projection matrix). The following properties for $\Pi$ hold:

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

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

$$\Pi^2 = \Phi^{1/2}X\Sigma^{-1}_a X^\top \Phi^{1/2} \Phi^{1/2}X\Sigma^{-1}_a X^\top \Phi^{1/2} = \Phi^{1/2}X(X^\top \Phi X)^{-1}X^\top \Phi X(X^\top \Phi X)^{-1}\Phi^{1/2} = \Phi^{1/2}X(X^\top \Phi X)^{-1}X^\top \Phi^{1/2} = \Pi.$$

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

$$\Pi u = \Phi^{1/2}X\Sigma^{-1}_a X^\top \Phi^{1/2}u = \Phi^{1/2}X(X^\top \Phi X)^{-1}X^\top \Phi^{1/2}Xv = \Phi^{1/2}Xv = u$$

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

Proof of 3: Because $\Sigma_a = X^\top \Phi X$ 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, $\dim(\text{Range}(\Pi)) = \dim(\text{Range}(\Phi^{1/2}X)) = p - \dim(\ker(\Phi^{1/2}X)) = p$. On the other hand, the eigenvalues of $\Pi$ must be either 0 or 1 because $\Pi$ is a projection matrix. So the eigenvalues of $\Pi$ are 1 with multiplicity $p$ and 0 with multiplicity $n - p$.

Proof of 4: By definition,

$$\Pi_{ii} = \sqrt{\pi_i^*} x_i^\top \Sigma^{-1}_a x_i \sqrt{\pi_i^*} = \pi_i^* x_i^\top \Sigma^{-1}_a x_i.$$
In addition, $\Pi$ is a symmetric projection matrix. Therefore,
\[
\Pi_{ii} = [\Pi^2]_{ii} = \Pi_i^T \Pi_i = \|\Pi_i\|^2.
\]

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 $\Pi = \Phi^{1/2}X\Sigma^{-1}X^T\Phi^{1/2}$ and $W$ be an $n \times n$ non-negative diagonal matrix. If $\|\Pi W\Pi - \Pi\|_2 \leq \epsilon$ for some $\epsilon \in (0, 1/2)$ then
\[
(1 - \epsilon)u^T \Sigma_a u \leq u^T \tilde{\Sigma}_a u \leq (1 + \epsilon)u^T \Sigma_a u, \quad \forall u \in \mathbb{R}^p,
\]
where $\Sigma_a = X^T \Phi X$ and $\tilde{\Sigma}_a = X^T W^{1/2} \Phi W^{1/2} X$.

We next proceed to find an appropriate diagonal matrix $W$ and validate Lemma 5. Define $\tilde{\Sigma}_T = \sum_{t=1}^T \tilde{\Sigma}_t$ and let $\tilde{\Sigma}_T = \frac{k}{T} \tilde{\Sigma}_W \leq \tilde{\Sigma}_S$, we have that $\tilde{\Sigma}_S \geq k \tilde{\Sigma}_T$ and hence $z^T \tilde{\Sigma}_S z \geq \frac{T}{k} z^T \tilde{\Sigma}_T z$ for all $z \in \mathbb{R}^p$. Therefore, to lower bound the spectrum of $\tilde{\Sigma}_S$ it suffices to lower bound the spectrum of $\tilde{\Sigma}_T$.

Define diagonal matrix $\tilde{W}$ as
\[
\tilde{W}_{jj} = \frac{\sum_{t=1}^T \tilde{w}_t^* j_t = j}{\pi_j^*}, \quad j = 1, \ldots, n.
\]

We have that $\tilde{\Sigma}_a = \tilde{\Sigma}_T$ for this particular choice of $W$.

**Lemma 6.** For any $\epsilon > 0$,
\[
\Pr\left[\|\Pi \tilde{W} \Pi - \Pi\|_2 > \epsilon\right] \leq 2 \exp\left\{-C \cdot \frac{T \epsilon^2}{p \log T}\right\},
\]
where $C > 0$ is an absolute constant.

**Proof.** Define $n$-dimensional random vector $v$ as
\[
\Pr\left[v = \sqrt{\frac{T \tilde{w}_j^*}{\pi_j^*}} \Pi_j \right] = p_j^{(1)}, \quad j = 1, \ldots, n.
\]
Let $v_1, \ldots, v_k$ be i.i.d. copies of $v$ and define $A_t = v_t v_t^T$. By definition, $\Pi W^{(1)} \Pi$ is equally distributed with $\frac{1}{k} \sum_{t=1}^k A_t$. In addition,
\[
\mathbb{E} A_t = \sum_{j=1}^n \frac{T \tilde{w}_j^*}{\pi_j^*} p_j^{(1)} \Pi_j \Pi_j^T = \Pi^2 = \Pi,
\]
\[\text{For those } j \text{ with } \pi_j^* = 0, \text{ we have by definition that } p_j^{(1)} = 0.\]
which satisfies \( \|E A_t\|_2 = 1 \), and
\[
\|A_t\|_2^2 = \|v_t^*\|^2 \leq \sum_{1 \leq j \leq n} \frac{T \tilde{w}_j^*}{\pi_j^*} \Pi_j \leq \sum_{1 \leq j \leq n} P_j \Pi_j \|J\|_2^2 \leq \sup_{1 \leq j \leq n} \frac{p}{x_j^T \Sigma^{-1} x_j} \cdot x_j^T \Sigma^{-1} x_j = p.
\]
Applying Lemma 95 we have that
\[
\Pr\left[ \|\hat{\Pi}_{\hat{W}} \Pi - \Pi\|_2 > \epsilon \right] \leq 2 \exp \left\{ -C \cdot \frac{T \epsilon^2}{p \log T} \right\}.
\]

Set \( t = O(1) \) and equate the right-hand side of Eq. (2.18) with \( O(1) \). We then have
\[
\Pr\left[ \hat{\Sigma}_T \geq \Omega(1) \cdot \hat{\Sigma}_* \right] = \Omega(1) \quad \text{if} \quad p \log T / T = O(1).
\]
Subsequently, under the condition that \( p \log T / T = O(1) \), with probability at least 0.9 it holds that
\[
\hat{\Sigma}_S \geq \Omega(T / k) \cdot \Sigma_*,
\]
which completes the proof of Lemma 3 for the with replacement setting.

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

More specifically, define independently distributed random matrices \( A_1, \ldots, A_T \) as
\[
A_t = (w_t - \pi_{i_t}^*) x_{i_t} x_{i_t}^T, \quad t = 1, \ldots, T.
\]
Note that \( w_t \) is a random Bernoulli variable with \( \Pr[w_{i_t} = 1] = kp_{i_t}^{(2)} = \pi_{i_t}^* \). Therefore, \( E A_t = 0 \).

In addition,
\[
\sup_{1 \leq t \leq T} \|A_t\| \leq \sup_{1 \leq j \leq n} \|x_j\|_{\infty}^2 \leq \|X\|_{\infty}^2 \quad a.s.,
\]
and
\[
\left\| \sum_{t=1}^T E A_t^2 \right\|_2 \leq \left\| \sum_{t=1}^T \pi_{i_t}^* (1 - \pi_{i_t}^*) \|x_{i_t}\|_{\infty}^2 x_{i_t} x_{i_t}^T \right\|_2 \leq \|X\|_{\infty}^2 \|\hat{\Sigma}_R\|_2.
\]
Noting that \( \sum_{t=1}^T A_t = \hat{\Sigma}_S - \hat{\Sigma}_{R} \) and invoking Lemma 96, we have that
\[
\Pr\left[ \|\hat{\Sigma}_S - \hat{\Sigma}_T\|_2 > t | R_T \right] \leq 2p \cdot \exp \left\{ -\frac{t^2}{3 \|\hat{\Sigma}_T\|_2 \|X\|_{\infty}^2 + 2 \|X\|_{\infty}^2 t} \right\}.
\]

Setting \( t = O(1) \cdot \lambda_{\min} (\Sigma_T) \) we have that, if \( \|\hat{\Sigma}_T\|_2 \kappa(\hat{\Sigma}_T) \|X\|_{\infty}^2 \log p = O(1) \), then with probability at least 0.95 conditioned on \( \hat{\Sigma}_T \)
\[
\Omega(1) \cdot \hat{\Sigma}_T \leq \hat{\Sigma}_S \leq O(1) \cdot \hat{\Sigma}_T.
\] (2.19)
It remains to establish spectral similarity between $\hat{\Sigma}_T$ and $\frac{T}{n}\Sigma_\star$, a scaled version of $\Sigma_\star$. Define deterministic matrices $A_1, \cdots, A_n$ as

$$A_j = \pi_j^\ast x_j^\top x_j - \frac{1}{n}\Sigma_\star, \quad j = 1, \cdots, n.$$ 

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

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

and

$$\frac{T}{n} \left\| \sum_{j=1}^n A_j \right\|_2 \leq \frac{2T}{n} \left( \left\| \sum_{j=1}^n (\pi_j^\ast)^2 \|x_i\|_2^2 x_i^\top x_i \right\|_2 + \frac{1}{n^2} \|\Sigma_\star\|_2^2 \right)$$

$$\leq \frac{2T}{n} \left( \|X\|_\infty^2 \left\| \sum_{j=1}^n \pi_j^\ast x_i x_i^\top \right\|_2 + \frac{1}{n^2} \|\Sigma_\star\|_2^2 \right)$$

$$\leq \frac{2T}{n} \left( \|X\|_\infty^2 \|\Sigma_\star\|_2 + \frac{1}{n^2} \|\Sigma_\star\|_2^2 \right)$$

$$\leq \frac{4T}{n} \|X\|_\infty^2 \|\Sigma_\star\|_2.$$ 

Invoking Lemma 97, we have that

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

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

$$\Omega(T/n) \cdot \Sigma_\star \leq \hat{\Sigma}_T \leq O(T/n) \cdot \Sigma_\star.$$  \hspace{1cm} (2.20)

Combining Eqs. (2.19,2.20) and noting that $\|\hat{\Sigma}_T^{-1}\|_2 \leq O(\frac{\lambda}{\lambda}) \|\Sigma_\star^{-1}\|_2$, $\kappa(\Sigma_T) \leq O(1)\kappa(\Sigma_\star)$, 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 $\hat{S} \subseteq [n]$ of size $k$ is obtained by running algorithm in Algorithm 4 with an initial subset $S_0 \subseteq [n]$, $|S_0| \geq k$. Both $\hat{S}$ and $S_0$ are standard sets (i.e., without replacement). Then

$$\tr \left[ (X_\hat{S}^\top X_\hat{S})^{-1} \right] \leq \frac{|S_0| - p + 1}{k - p + 1} \tr \left[ (X_{S_0}^\top X_{S_0})^{-1} \right].$$
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 
\[
\text{tr}[(X_S^\top X_S)^{-1}] \leq \frac{n^{p+1}}{k-p+1} \text{tr}[(X^\top X)^{-1}].
\]
This results in an approximation ratio of \( \frac{n^{p+1}}{k-p+1} \) by applying the trivial bound 
\[
\text{tr}[(X^\top X)^{-1}] \leq \min_{s_i \in \{0,1\}} \sum_{i=1}^k F(s),
\]
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 \( \pi^* \); that is, \( S_0 = \{j \in [n] : \pi^*_j > 0\} \). Because \( \|\pi^*\|_\infty \leq 1 \) under the without replacement setting, we have the following corollary:

**Corollary 1.** Let \( S_0 \) be the support of \( \pi^* \) and suppose \( \|\pi^*\|_\infty \leq 1 \). Then

\[
\text{tr}[(X_S^\top X_S)^{-1}] \leq \frac{\|\pi^*\|_0 - p + 1}{k-p+1} F(\pi^*) \leq \frac{\|\pi^*\|_0 - p + 1}{k-p+1} \min_{s_i \in \{0,1\}, \sum_i s_i \leq k} F(s).
\]

It is thus important to upper bound the support size \( \|\pi^*\|_0 \). With the trivial bound of \( \|\pi^*\|_0 \leq n \) 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 \( \phi : \mathbb{R}^p \rightarrow \mathbb{R}^{\frac{p(p+1)}{2}} \) as \( \phi(x) = (\xi_{ij}x(i)x(j))_{1 \leq i \leq j \leq p} \) where \( x(i) \) denotes the \( i \)th coordinate of a \( p \)-dimensional vector \( x \) and \( \xi_{ij} = 1 \) if \( i = j \) and \( \xi_{ij} = 2 \) otherwise. Denote \( \tilde{\phi}(x) = (\phi(x), 1) \in \mathbb{R}^{\frac{p(p+1)}{2}+1} \) as the affine version of \( \phi(x) \). For any \( \frac{p(p+1)}{2} + 1 \) distinct rows of \( X \), their mappings under \( \tilde{\phi} \) 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 \( \pi^* \).

**Lemma 8.** \( \|\pi^*\|_0 \leq k + \frac{p(p+1)}{2} \) 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 \( f(\pi; \lambda, \bar{\lambda}, \mu) \) be the Lagrangian multiplier function of the without replacement formulation \( (b = 1) \) of Eq. (2.4):

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

Here \( \{\lambda_i\}_{i=1}^n \geq 0 \), \( \{\bar{\lambda}_i\}_{i=1}^n \geq 0 \) and \( \mu \geq 0 \) are Lagrangian multipliers for constraints \( \pi_i \geq 0 \), \( \pi_i \leq 1 \), and \( \sum_i \pi_i \leq k \), respectively. By KKT condition, \( \frac{\partial f}{\partial \pi_i} \bigg|_{\pi^*} = 0 \) and hence

\[
-\frac{\partial f}{\partial \pi_i} \bigg|_{\pi^*} = x_i^\top \sum_{a}^2 x_i = \bar{\lambda}_i - \lambda_i + \mu, \quad i = 1, \cdots, n.
\]
where $\Sigma_* = X^T \text{diag}(\pi^*) X$ 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 $|A| \leq k$.

Therefore, to upper bound $\|\pi^*\|_0$ it suffices to upper bound $|B|$. By complementary slackness, for all $i \in B$ we have that $\tilde{\lambda}_i = \lambda_i = 0$; that is,

$$x_i^T \Sigma_*^{-2} x_i = \langle \phi(x_i), \psi(\Sigma_*^{-2}) \rangle = \mu, \quad \forall i \in B,$$

(2.21)

where $\phi : \mathbb{R}^p \to \mathbb{R}^{(p+1)/2}$ 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 $x_1, \ldots, x_{p(p+1)/2+1}$ 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:

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

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

### 2.7.3 Proof of Lemma 1

To prove Lemma 1 we consider the following online matrix game: let $\Delta_{p \times p} = \{ A \in \mathbb{R}^{p \times p} : A \succeq 0, \text{tr}(A) = 1 \}$ 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 $\langle F_t, A_t \rangle = \text{tr}(F_t^T A_t)$. The goal of the player is to minimize his/her regret:

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

(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 “hind-sight” (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:

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

(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 \( \psi \).

Perhaps the most famous choice of \( \psi \) 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 
\( \psi(A) = -2\text{tr}(A^{1/2}) \) 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,

\[
A_t = \left( c_t I + \alpha Z_0 + \alpha \sum_{\ell=0}^{t-1} F_\ell \right)^{-2}, \quad t = 1, 2, \ldots,
\]

(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 \( \text{tr}(A_t) = 1 \).

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 \( F_t = u_t u_t^T - v_t v_t^T \), then we prove the following lemma which upper bounds the total regret of the \( \ell_{1/2} \) strategy:

**Lemma 9** (main regret lemma). Suppose \( F_t = u_t u_t^T - v_t v_t^T \) 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 \( \alpha \langle A_t^{1/2}, v_t v_t^T \rangle < 1/2 \) for all \( t \), we have for any \( U \in \Delta_{p \times p} \),

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

(2.26)

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

\[
\sum_{t=0}^{T-1} \langle F_t, A_t - U \rangle \leq 2\alpha \sum_{t=0}^{T-1} \left( -\frac{\langle A_t, u_t u_t^T \rangle \cdot \langle A_t^{1/2}, u_t u_t^T \rangle}{1 + 2\alpha \langle A_t^{1/2}, u_t u_t^T \rangle} + \frac{\langle A_t, v_t v_t^T \rangle \cdot \langle A_t^{1/2}, v_t v_t^T \rangle}{1 - 2\alpha \langle A_t^{1/2}, v_t v_t^T \rangle} \right) + \frac{\Delta_{\psi}(A_0, U)}{\alpha}.
\]
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 \( F_t = u_t u_t^T \) (thus of rank 1) was originally presented in Theorem 3.2 of (Allen-Zhu et al., 2015). The involvement of the extra \(-v_t v_t^T\) 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:

\[
\tilde{A}_t = \arginf_{A \succeq 0} \{ \Delta_\psi(A_{t-1}, A) + \alpha \langle F_{t-1}, A \rangle \}; \quad A_t = \arginf_{A \in \Delta_{p \times p}} \Delta_\psi(\tilde{A}_t, A).
\]

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 \( \arginf_{A \in \Delta_{p \times p}} \{ \Delta_\psi(A_{t-1}, A) + \alpha \langle F_{t-1}, A \rangle \} \). This can also be verified by writing \( \tilde{A}_t \) explicitly using the following claim, and verifying that \( A_t \) (in its closed form by Proposition 2) is indeed a minimizer of \( \Delta_\psi(\tilde{A}_t, A) \) over \( A \in \Delta_{p \times p} \) by taking its gradient.

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

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

\[
\langle \alpha F_{t-1}, A_{t-1} - U \rangle = \langle \nabla \psi(A_{t-1}) - \nabla \psi(\tilde{A}_t), A_{t-1} - U \rangle = \Delta_\psi(A_{t-1}, U) - \Delta_\psi(\tilde{A}_t, U) + \Delta_\psi(\tilde{A}_t, A_{t-1}) \leq \Delta_\psi(\tilde{A}_{t-1}, U) - \Delta_\psi(\tilde{A}_t, U) + \Delta_\psi(\tilde{A}_t, A_{t-1}).
\]

(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 \( \Delta_\psi(\tilde{A}_t, A_{t-1}) \) by its definition gives

\[
\Delta_\psi(\tilde{A}_t, A_{t-1}) = \psi(A_{t-1}) - \psi(\tilde{A}_t) - \langle \nabla \psi(\tilde{A}_t), A_{t-1} - \tilde{A}_t \rangle \\
= -2\text{tr}(A_{t-1}^{1/2}) + 2\text{tr}(\tilde{A}_t^{1/2}) + \langle \tilde{A}_t^{1/2}, A_{t-1} - \tilde{A}_t \rangle \\
= \langle \tilde{A}_t^{1/2}, A_{t-1} \rangle - \text{tr}(\tilde{A}_t^{1/2}) + 2\text{tr}(A_{t-1}^{1/2}) \\
= \langle A_{t-1}^{1/2} + \alpha F_{t-1}, A_{t-1} \rangle + \text{tr}(\tilde{A}_t^{1/2}) - 2\text{tr}(A_{t-1}^{1/2}) \\
= \langle \alpha F_{t-1}, A_{t-1} \rangle + \text{tr}(\tilde{A}_t^{1/2}) - \text{tr}(A_{t-1}^{1/2}).
\]

(2.28)

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

\[
-\alpha \sum_{t=0}^{T-1} \langle F_t, U \rangle \leq \Delta_\psi(A_0, U) - \Delta_\psi(\tilde{A}_T, U) + \sum_{t=0}^{T-1} \text{tr}(\tilde{A}_t^{1/2}) - \text{tr}(A_t^{1/2}) \\
\leq \Delta_\psi(A_0, U) + \sum_{t=0}^{T-1} \text{tr}(\tilde{A}_t^{1/2}) - \text{tr}(A_t^{1/2}),
\]

(2.29)
where the second inequality holds because Bregman divergence $\Delta_\psi(\tilde{A}_t, U)$ is always non-negative.

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

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

It is crucial to spectrally lower bound the core $2 \times 2$ 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 $P_t^\top A_t^{1/2} P_t = [b \ d \ c] \in \mathbb{R}^{2 \times 2}$ and $2\alpha\langle A_t^{1/2}, v_t v_t^\top\rangle < 1$. Then

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

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 $P A P^\top$, 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 $\text{tr}(\tilde{A}_{t+1}^{1/2}) - \text{tr}(A_t^{1/2})$ can be bounded as

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

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

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

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

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 $U \in \Delta_{p \times p}$:

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

**Proof.** By definition of $\Delta_\psi, \psi$ and $A_0$, we have

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

\(^5(A + UCV)^{-1} = A^{-1} - A^{-1}U(C^{-1} + VA^{-1}U)^{-1}VA^{-1},\) provided that all inverses exist.
where the last inequality holds because \( \text{tr}(U^{1/2}) \geq 0 \) and \( \text{tr}(A_0^{1/2}) \leq \sqrt{p} \cdot \text{tr}(A_0) = \sqrt{p} \). Note also that \( \langle I, U \rangle = \text{tr}(U) = 1 \). Therefore,

\[
\Delta_\psi(A_0, U) \leq \alpha \langle Z_0, U \rangle + c_0 + \sqrt{p}.
\]

Because \( \text{tr}(A_0) = 1 \), the constant \( c_0 \) (if positive) must be upper bounded by \( \sqrt{p} \) because otherwise \( \text{tr}(A_0) \leq \text{tr}((c_0I)^{-2}) = p \cdot c_0^{-2} < 1 \). Therefore, it is proved that \( \Delta_\psi(A_0, U) \leq \alpha \langle Z_0, U \rangle + 2\sqrt{p} \).

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 \( \text{tr}[(\alpha Z + cI)^{-2}] = 1 \). By simple asymptotic analysis, \( \lim_{c \to (-\alpha \lambda_{\min}(Z))^{-1}} \text{tr}[(\alpha Z + cI)^{-2}] = +\infty \) and \( \lim_{c \to +\infty} \text{tr}[(\alpha Z + cI)^{-2}] = 0 \). Because \( \text{tr}[(\alpha Z + cI)^{-2}] \) 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 \( \text{tr}[(\alpha Z + cI)^{-2}] = 1 \). 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 \( t \geq 1 \), and try to prove the same for \( A_t \).

The KKT condition of the optimization problem and the gradients of the Bregman divergence \( \Delta_\psi \) yields

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

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

\[
A_t = (\nabla \psi)^{-1} (\nabla \psi(A_{t-1}) + d_t I - \alpha F_{t-1}) \]

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

Because \( d_t \) is the unique number that ensures \( A_t \geq 0 \) and \( \text{tr}(A_t) = 1 \), and \( Z_0 + \sum_{\ell=0}^{t-1} F_\ell \geq 0 \), it must hold that \(-c_{t+1} + d_t = c_t \). Subsequently, \( \nabla \psi(A_t) = -(A_t^{-1/2}) = -c_tI - \alpha Z_0 - \alpha \sum_{\ell=0}^{t-1} F_\ell \). 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 \( \alpha \langle v_{t-1} v_{t-1}^\top, v_{t-1} v_{t-1}^\top \rangle < 1 \), and hence \(-\alpha F_{t-1} \leq -\alpha v_{t-1} v_{t-1}^\top \leq A_{t-1}^{-1/2} \). This is because for any matrices \( A \geq 0 \) and \( B \geq 0 \), we have \( \langle A, B \rangle = \text{tr}(A^\top B) < 1 \implies A^\top B \leq I \implies B \leq A^{-1} \). Consequently, we have \( A_{t-1}^{-1/2} + \alpha F_{t-1} > 0 \) and its inverse exists.
Next, to prove \( \tilde{A}_t = (A_{t-1}^{-1/2} + \alpha F_{t-1})^{-2} \) 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 \( \tilde{A}_t \) is zero.\(^6\) Indeed,

\[
\nabla \left( \Delta_{\psi}(A_{t-1}, \tilde{A}_t) + \alpha\langle F_{t-1}, \tilde{A}_t \rangle \right) = \nabla \psi(\tilde{A}_t) - \nabla \psi(A_{t-1}) + \alpha F_{t-1} = -\tilde{A}_t^{-1/2} + \tilde{A}_t^{-1/2} + \alpha F_{t-1} = 0.
\]

**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:

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

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 \( Q^\top (J + \text{diag}(2b, 2c))^{-1} Q < I \). By standard matrix analysis and the fact that \( J = \text{diag}(1, -1), \)

\[
Q^\top (J + \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix})^{-1} Q = Q^\top \begin{bmatrix} (1 + 2b)^{-1} & 0 \\ 0 & -(1 - 2c)^{-1} \end{bmatrix} Q \leq \frac{\|QQ^\top\|_{op}}{1 + 2b} \cdot I.
\]

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 \( QQ^\top = \begin{bmatrix} b & -d \\ -d & c \end{bmatrix} \leq \begin{bmatrix} 2b & 0 \\ 0 & 2c \end{bmatrix} \); finally, (c) holds because \( b = 2\alpha\langle A_t^{1/2}, u_t u_t^\top \rangle \geq 0, \frac{2b}{1 + 2b} < 1 \) 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:

\( \footnote{The convexity of this objective follows from Lieb’s concavity theorem (Bhatia, 1997; Lieb, 1973), and is already a known fact in matrix regret minimization literatures (Allen-Zhu et al., 2015).} \)
Lemma 11 (main averaging lemma). For every $\varepsilon > 0$ and subset $\Lambda \subseteq [n]$ of cardinality $k$, suppose $\lambda_{\text{min}}(\sum_{i \in \Lambda} x_i x_i^\top) \leq 1 - 15\varepsilon$ 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 $\text{tr}(A) = 1$. Then, the following statements are true:

\begin{align}
\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} &\leq \frac{1 - \varepsilon}{k}; \\
(2.33) \\
\max_{j \in [n] \setminus \Lambda} \frac{\langle A, x_j x_j^\top \rangle}{1 + 2 \alpha \langle A^{1/2}, x_j x_j^\top \rangle} &\geq \nu + \frac{\varepsilon}{k}. \\
(2.34)
\end{align}

Furthermore, if $\alpha = \sqrt{p}/\varepsilon$ and $k \geq 6p/\varepsilon^2$ 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 $\lambda_{\text{min}}(\sum_{i \in \Lambda} x_i x_i^\top) \leq 1 - 15\varepsilon$, we can simply choose $i_t$ to be the index $i \in \Lambda_t$ which minimizes $\frac{\langle A, x_i x_i^\top \rangle}{1 - 2 \alpha \langle A_{i_t}^{1/2}, x_i x_i^\top \rangle}$, and $j_t$ to be the index $j \notin \Lambda_t$ which maximizes $\frac{\langle A, x_j x_j^\top \rangle}{1 + 2 \alpha \langle A_{i_t}^{1/2}, x_j x_j^\top \rangle}$. Eqs. (2.33) and (2.34) together imply that

\begin{equation}
\left( - \frac{\langle A_{i_t}, x_{j_t} x_{j_t}^\top \rangle}{1 + 2 \alpha \langle A_{i_t}^{1/2}, x_{j_t} x_{j_t}^\top \rangle} + \frac{\langle A_{i_t}, x_i x_i^\top \rangle}{1 - 2 \alpha \langle A_{i_t}^{1/2}, x_i x_i^\top \rangle} \right) \leq \frac{\varepsilon}{k}. \\
(2.35)
\end{equation}

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

\[-\lambda_{\text{min}} \left( \sum_{j \in \Lambda_T} x_j x_j^\top \right) \leq \frac{T - 1}{k} \varepsilon + \frac{2\sqrt{p}}{\alpha} = -\frac{T \varepsilon}{k} + 2\varepsilon .
\]

Here in the last inequality we apply the choice $\alpha = \sqrt{p}/\varepsilon$. In other words, as long as $T \geq k/\varepsilon$, it must satisfy $\lambda_{\text{min}}(\sum_{j \in \Lambda_T} x_j x_j^\top) \geq 1 - 2\varepsilon$, 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 $\lambda_{\text{min}}(Z) \leq 1$. Let $A = (\alpha Z + cI)^{-2}$, where $c \in \mathbb{R}$ is the unique real number such that $A \succeq 0$ and $\text{tr}(A) = 1$. Then

1. $\alpha \langle A^{1/2}, Z \rangle \leq p + \alpha \sqrt{p}$;
2. $\langle A, Z \rangle \leq \sqrt{\alpha/\alpha + \lambda_{\text{min}}(Z)}$.

**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 $Z = \text{diag}(\sigma_1, \ldots, \sigma_p)$ for $\lambda_1 \geq \cdots \lambda_p \geq 0$, because $Z \succeq 0$. Subsequently,

\[\alpha \langle A^{1/2}, Z \rangle = \sum_{i=1}^{p} \frac{\alpha \lambda_i}{\alpha \lambda_i + c} = p - c \cdot \sum_{i=1}^{p} \frac{1}{\alpha \lambda_i + c}.\]
If \( c \geq 0 \), then \( \alpha \langle A^{1/2}, Z \rangle \leq p \) 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,

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

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

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

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

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

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

\[
\sum_{i \in \Lambda} 2\alpha \langle A^{1/2}, x_i x_i^T \rangle = 2\alpha \langle A^{1/2}, Z \rangle \geq |\Lambda| = k.
\]

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

\[
2\alpha \langle A^{1/2}, Z \rangle \leq 2p + 2\alpha \sqrt{p}
\]

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

We next proceed to prove Eq. (2.33). By definition of \( \nu \), we must have that \( (1 - 2\alpha \langle A^{1/2}, x_i x_i^T \rangle) \nu \leq \langle A, x_i x_i^T \rangle \) for all \( i \in \Lambda \), because if \( 2\alpha \langle A^{1/2}, x_i x_i^T \rangle \geq 1 \) 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,

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

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 \( \nu \leq (1 - \varepsilon)/k \).
Finally we prove Eq. (2.34). Define $t = \nu + \varepsilon/k \leq 1/k$. To prove Eq. (2.34) it suffices to show that $\sum_{j \in [n] \setminus \Lambda} \lambda_j \langle A, x_j x_j^\top \rangle \geq t \sum_{j \in [n] \setminus \Lambda} \pi_j (1 + 2\alpha \langle A^{1/2}, x_j x_j^\top \rangle)$, because $\pi_j \geq 0$ 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

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

Similarly,

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

Note that for any $p \times p$ positive semi-definite matrix $Z \succeq 0$, $\text{tr}(Z^{1/2}) \leq \sqrt{p} \cdot \text{tr}(Z)$ thanks to the Hölder’s inequality \footnote{If $x_1 + \cdots + x_d \leq \sqrt{d} \cdot \sqrt{x_1^2 + \cdots + x_d^2}$ for any sequences of $d$ real numbers $x_1, \ldots, x_n$.} applied to the non-negative spectrum of $Z^{1/2}$, and that $\text{tr}(A) = 1$ by definition. Subsequently,

$$
\sum_{j \in [n] \setminus \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \cdot \sum_{j \in [n] \setminus \Lambda} \pi_j (1 + 2\alpha \langle A^{1/2}, x_j x_j^\top \rangle) \\
\geq \text{tr}(A) - \sum_{j \in \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \left( k - \sum_{j \in \Lambda} \pi_j \right) - 2\alpha t \cdot \text{tr}(A^{1/2}) \\
\geq 1 - \sum_{j \in \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \left( k - \sum_{j \in \Lambda} \pi_j \right) - 2\alpha t \sqrt{p} \\
= 1 - tk - 2t\alpha \sqrt{p} - \sum_{j \in \Lambda} \pi_j \langle A, x_j x_j^\top \rangle - t \\
\geq 1 - tk - 2t\alpha \sqrt{p} - \sum_{j \in \Lambda} \max\{\langle A, x_j x_j^\top \rangle - t, 0\} \tag{2.37}
$$

$$
\geq 1 - tk - 2t\alpha \sqrt{p} - \sum_{j \in \Lambda} (\langle A, x_j x_j^\top \rangle - t) - \sum_{j \in \Lambda} \max\{(t - \langle A, x_j x_j^\top \rangle), 0\} \\
\geq 1 - 2t\alpha \sqrt{p} - \sqrt{p}/\alpha - \lambda_{\min} \left( \sum_{j \in \Lambda} x_j x_j^\top \right) - \sum_{j \in \Lambda} \max\{(t - \langle A, x_j x_j^\top \rangle), 0\}. \tag{2.38}
$$

Here Eq. (2.37) holds because $\pi_j \leq 1$ for all $j$, and in Eq. (2.38) we apply the property as established in Proposition 5. By the conditions that $\alpha =
\(\sqrt{p}/\varepsilon, t \leq 1/k\) and \(\lambda_{\min}(\sum_{j=1}^{n} x_j x_j^T) \leq 1 - 3\varepsilon\), the right-hand side of Eq. (2.38) can be lower bounded by the simplified form of

\[
\sum_{j \in [n] \setminus \Lambda} \pi_j \left\{ \langle A, x_j x_j^T \rangle - t(1 + 2\alpha \langle A^{1/2}, x_j x_j^T \rangle) \right\} \geq 2\varepsilon - \frac{2p}{\varepsilon k} \max_{j \in \Lambda} \{ t - \langle A, x_j x_j^T \rangle, 0 \}. \tag{2.39}
\]

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

\[
\sum_{i \in A'} (\nu - \langle A, x_i x_i^T \rangle) \leq \sum_{i \in A'} 2\nu \alpha \langle A^{1/2}, x_j x_j^T \rangle \leq 2\nu(p + \alpha \sqrt{p}) \quad \text{for all} \ A' \subseteq \Lambda.
\]

Consider \(\Lambda = \{ i \in \Lambda : t - \langle A, x_i x_i^T \rangle \geq 0 \}\). We then have

\[
\sum_{j \in \Lambda} \max_{j \in \Lambda} \{ t - \langle A, x_j x_j^T \rangle, 0 \} = \sum_{j \in \Lambda} (t - \langle A, x_j x_j^T \rangle) = (t - \nu) |\Lambda| + \sum_{j \in \Lambda} (\nu - \langle A, x_j x_j^T \rangle) \leq \varepsilon + \frac{4p/\varepsilon}{k}, \tag{2.40}
\]

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

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

If \(k \geq 6p/\varepsilon^2\), 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 \(W := \text{diag}(w_1, \ldots, w_n)\). The weighted OLS estimator \(\hat{\beta}_k\) then admits a closed-form expression

\[
\hat{\beta}_k = (X^T W X)^{-1} X^T W \tilde{y}, \tag{2.41}
\]

where \(\tilde{y} = (\tilde{y}_1, \ldots, \tilde{y}_n) \in \mathbb{R}^n\). Define \(\varepsilon := \tilde{y} - y\). Using the linear model that \(y = X\beta_0\), we have \(\hat{\beta}_k - \beta_0 = (X^T W X)^{-1} X^T W \varepsilon\). On the other hand, by the quantized error model Eq. (2.13), it holds that \(\mathbb{E}[\varepsilon_i | X] = 0\) and \(\mathbb{E}[\varepsilon_i^2 | X] \leq 4^{-2k_1+1} M^2 = w_i^{-1} M^2\). Subsequently,

\[
\mathbb{E}\|\hat{\beta}_k - \beta_0\|^2_2
\]

\[
= \text{tr} \left[ (X^T W X)^{-1} X^T W \mathbb{E}(\varepsilon \varepsilon^T) W^T X (X^T W X)^{-1} \right]
\]

\[
\leq M^2 \cdot \text{tr} \left[ (X^T W X)^{-1} X^T W X (X^T W X)^{-1} \right]
\]

\[
= M^2 \cdot \text{tr} \left[ (X^T W X)^{-1} \right].
\]

36
Chapter 3

Selective queries in nonparametric optimization

Let \( f : \mathcal{X} \rightarrow \mathbb{R} \) 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
\[
\min_{x \in \mathcal{X}} f(x),
\]
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 \( \hat{x}_n \in \mathcal{X} \) 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
\[
y_t = f(x_t) + \xi_t,
\]
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 \( \mathbb{E}f(\hat{x}_n) - f^* \) where \( f^* = \min_{x \in \mathcal{X}} f(x) \), reflecting the gap in function values between the estimated minimizer \( \hat{x}_n \) and the true minimizer \( x^* \in \arg\min_{x \in \mathcal{X}} f(x) \).

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...
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 $f : \mathcal{X} \to \mathbb{R}$ over the unit cube $\mathcal{X} = [0, 1]^d$. The optimization error of $\hat{x}_n \in \mathcal{X}$ is evaluated by

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

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 $\hat{x}_n$ such that

$$\sup_{f_0 \in \Theta} \sup_{f \in \Theta', f \neq f_0} \Pr_{|f-f_0| \leq \varepsilon_n(f_0)} \mathcal{L}(\hat{x}_n; f) \geq C_1 \cdot R_n(f_0) \leq 1/4, \quad (3.4)$$

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 $\mathcal{L}(\hat{x}_n; f)$ 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 $\hat{x}_n$ has perfect knowledge about the reference function $f_0 \in \Theta$. We then prove locally minimax lower bounds of the form

$$\inf_{\hat{x}_n \in \Theta, |f-f_0| \leq \varepsilon_n(f_0)} \sup_{f \in \Theta'} \Pr_{|f-f_0| \leq \varepsilon_n(f_0)} \mathcal{L}(\hat{x}_n; f) \geq C_2 \cdot R_n(f_0) \geq 1/3, \quad (3.5)$$

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 $\Theta$ and $\Theta'$**: The reference function $f_0$ and the true functions $f$ are assumed to belong to different but closely related function classes $\Theta$ and $\Theta'$. 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 $\varepsilon_n$ are based.
2. Upper Bounds: It is worth emphasizing that the estimator \( \hat{x}_n \) has no knowledge of the reference function \( f_0 \). From the perspective of upper bounds, we can consider the simpler task of producing \( f_0 \)-dependent bounds (eliminating the second supremum) to instead study the (already interesting) quantity:

\[
\sup_{f_0} \Pr_{f_0} \left[ \mathcal{L}(\hat{x}_n; f_0) \geq C_1 R_n(f_0) \right] \leq 1/4.
\]

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 \( \hat{x}_n \) 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 \( \mathcal{X}_{f,\kappa} := \{ x \in \mathcal{X} : f(x) \leq f^* + \kappa \} \) belongs to the Hölder class \( \Sigma^\alpha(M) \), meaning that \( f \) is \( k \)-times differentiable on \( \mathcal{X}_{f,\kappa} \) and furthermore for any \( x, x' \in \mathcal{X}_{f,\kappa} \),

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

Here \( k = \lfloor \alpha \rfloor \) is the largest integer lower bounding \( \alpha \) and \( f^{(\alpha,j)}(x) := \partial^j f(x)/\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d} \).

We use \( \Sigma^\alpha(M) \) 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 concerns the “regularity” of the level sets of the “reference” function \( f_0 \).

Define \( L_{f_0}(\epsilon) := \{ x \in \mathcal{X} : f_0(x) \leq f^*_0 + \epsilon \} \) as the \( \epsilon \)-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 \( \ell_2 \)-balls of radius \( \delta \) that cover \( L_{f_0}(\epsilon) \).

\[1\]the particular \( \ell_2 \) 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 \( N(L_{f_0}(\epsilon), \delta) \leq C_0[1 + \mu_{f_0}(\epsilon)\delta^{-d}] \) for all \( \epsilon, \delta \in (0, c_0] \).

We use \( \Theta_C \) 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 \( \ell_2 \) ball of radius \( r \): \( L_{f_0}(\epsilon) = \{ x \in \mathcal{X} : \| x - x^a \|_2 \leq r \} \). Clearly, \( \mu_{f_0}(\epsilon) \approx r^d \). In addition, the \( \delta \)-covering number in \( \ell_2 \) of \( L_{f_0}(\epsilon) \) is on the order of \( 1 + (r/\delta)^d \approx 1 + \mu_{f_0}(\epsilon)\delta^{-d} \), 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 standalone and well-spaced points in \( \mathcal{X} \): the Lebesgue measure of \( L_{f_0}(\epsilon) \) would be zero, but at least \( \Omega(n) \) 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 \( N(L_{f_0}(\epsilon), \delta) \geq n \) as \( \delta \to 0^+ \) but \( \mu_{f_0}(\epsilon) = 0 \).
Algorithm 7 Successive elimination for low-dimensional nonparametric optimization

1: Parameters: $\alpha, M, \delta, n$
2: Output: $\hat{x}_n = x_n$, the final prediction
3: Initialization: $S_0 = G_n$, $\varrho_0(x) \equiv \infty$, $T = \lceil \log_2 n \rceil$, $n_0 = \lceil n/T \rceil$
4: for $\tau = 1, 2, \ldots, T$ do
5:     Compute “extended” sample set $S^0_{\tau-1}(\varrho_{\tau-1})$ defined in Eq. (3.7);
6:     for $t = (\tau - 1)n_0 + 1$ to $\tau n_0$ do
7:         Sample $x_t$ uniformly at random from $S^0_{\tau-1}(\varrho_{\tau-1})$ and observe $y_t = f(x_t) + \omega_t$;
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:    $S_\tau := \{ x \in S_{\tau-1} : \ell_t(x) \leq \min_{x \in S_{\tau-1}} u_t(x') \}$, $\varrho_t(x) := \min \{ \varrho_{\tau-1}(x), h_t(x) \}$.
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 $\alpha$-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 $p_0 \leq p_X(x) \leq \bar{p}_0$ for all $x \in \mathcal{X}$, where $0 < p_0 \leq \bar{p}_0 < \infty$ are uniform constants;

(B2) $|G_n| \gtrsim n^{\alpha/d} \min(\alpha, 1)$ 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 $S^0(\varrho)$ of $S$ with respect to $\varrho$ as

$$S^0(\varrho) := \bigcup_{x \in S} B_{\varrho(x)}^\infty(x; G_n) \quad \text{where} \quad B_{\varrho(x)}^\infty(x; G_n) = \{ z \in G_n : \| z - x \|_\infty \leq \varrho(x) \}. \quad (3.7)$$

The algorithm can then be formulated as two level of iterations, with the outer loop shrinking the “active set” $S_\tau$ 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 $\hat{f}(x)$. In particular, for any $x \in G_n$ and a bandwidth parameter $h > 0$,
consider a least square polynomial estimate

\[ \hat{f}_h \in \arg \min_{g \in \mathcal{P}_h} \sum_{t=1}^{T} \mathbb{I}[x_t \in B^\infty_h(x)] \cdot (y_t - g(x_t))^2, \]  

(3.8)

where \( B^\infty_h(x) := \{ x' \in \mathcal{X} : \| x' - x \|_\infty \leq h \} \) and \( \mathcal{P}_h \) denotes all polynomials of degree \( k \) on \( \mathcal{X} \).

To analyze the performance of \( \hat{f}_h \), define mapping \( \psi_{x,h} : z \mapsto (1, \psi^1_{x,h}(z), \ldots, \psi^k_{x,h}(z)) \) where \( \psi^j_{x,h} : z \mapsto \{ \prod_{t=1}^{T} h^{-1}(z_{x_t} - x_{x_t}) \} \) is the degree-\( j \) polynomial mapping from \( \mathbb{R}^d \) to \( \mathbb{R}^{d^j} \). Also define \( \Psi_{t,h} := (\psi_{x,h}(x_t))_{1 \leq t \leq t, x \in B^\infty_h(x)} \) as the aggregated design matrix, where \( y = \sum_{t=1}^{T} \mathbb{I}[x_t \in B^\infty_h(x)] \) and \( D = 1 + d + \ldots + d^k \). The estimate \( \hat{f}_h \) defined in Eq. (3.8) then admits the following closed-form expression:

\[ \hat{f}_h(z) \equiv \psi_{x,h}(z)^\top (\Psi_{t,h}^\top \Psi_{t,h})^{-1} \Psi_{t,h} Y_{t,h}, \]  

(3.9)

where \( Y_{t,h} = (y_t)_{1 \leq t \leq t, x \in B^\infty_h(x)} \) and \( A^\top \) is the Moore-Penrose pseudo-inverse of \( A \).

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

**Lemma 12.** Suppose \( f \) satisfies Eq. (3.6) on \( B^\infty_h(x; \mathcal{X}) \), \( \max_{x \in B^\infty_h(x; \mathcal{X})} \| \psi_{x,h}(z) \|_2 \leq b \) 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 \( 1 - \delta \)

\[ |\hat{f}_h(x) - f(x)| \leq \frac{b^2}{\sigma} M d^k h^\alpha + b \sqrt{\frac{5D \ln(1/\delta)}{\sigma m}} =: \eta_{h,\delta}(x). \]  

(3.10)

**Remark 4.** \( b_{h,\delta}(x) \), \( s_{h,\delta}(x) \) and \( \eta_{h,\delta}(x) \) depend on \( x \) because \( \sigma \) depends on \( \Psi_{t,h} \), which further depends on the sample points in the neighborhood \( B^\infty_h(x; \mathcal{X}) \) of \( x \).

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

Notice that when bandwidth \( h \) increases, the bias term \( b_{h,\delta}(x) \) is likely to increase too because of the \( h^\alpha \) term; on the other hand, with \( h \) increasing the local neighborhood \( B^\infty_h(x; \mathcal{X}) \) enlarges and would potentially contain more samples, implying a larger \( m \) and smaller standard deviation term \( s_{h,\delta}(x) \). A careful selection of bandwidth \( h \) balances \( b_{h,\delta}(x) \) and \( s_{h,\delta}(x) \) 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 \( b_{h,\delta}(x) \) and standard deviation \( s_{h,\delta}(x) \) in Eq. (3.10), the bandwidth \( h_t(x) > 0 \) at epoch \( t \) and point \( x \) is selected as

\[ h_t(x) := \frac{j_t(x)}{n^2} \text{ where } j_t(x) := \arg \max \{ j \in \mathbb{N}, j \leq n^2 : b_{j/n^2,\delta}(x) \leq s_{j/n^2,\delta}(x) \}. \]  

(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 \( \hat{f}_h(x) \) 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

\[
\ell_t(x) := \max_{1 \leq i \leq t} \left\{ \tilde{f}_{h_t}(x) - \eta_{h_t}(x), \delta(x) \right\} \quad \text{and} \quad u_t(x) := \min_{1 \leq i \leq t} \left\{ \tilde{f}_{h_t}(x) + \eta_{h_t}(x), \delta(x) \right\}.
\]

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 \( \tilde{f}(x) \) 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 \( \tilde{f}(x) \geq \min_{z \in G_n} \tilde{f}(z) + 1 / \log n \).

**Remark.** The \( 1 / \log n \) term in removal condition \( \tilde{f}(x) \geq \min_{z \in G_n} \tilde{f}(z) + 1 / \log n \) 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 \( f \in \Sigma_\kappa^n(M) \) 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

\[
\min_{x \in S_0'} f(x) = \min_{x \in G_n} f(x) \quad \text{and} \quad S_0' \subseteq \bigcup_{x \in L_f(\kappa/2)} B_{h_0}^{\kappa n}(x ; \mathcal{X}),
\]

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

To interpret \( \kappa/2 \), note that for sufficiently large \( n \), \( f \in \Sigma_\kappa^n(M) \) implies \( f \) being \( \alpha \)-H"older smooth (i.e., \( f \) satisfies Eq. (3.6)) on \( \bigcup_{x \in L_f(\kappa/2)} B_{h_0}^{\kappa n}(x ; \mathcal{X}) \), 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 \( f \in \Sigma_\kappa^n(M) \), 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 \( f_0 \in \Sigma_0^\alpha(M) \cap \Theta_C \), where \( C = (c_0, C_0) \), define

\[
\varepsilon_n^U(f_0) := \sup \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geq n/ \log^\omega n \right\},
\]

where \( \omega > 5 + d/\alpha \) is a large constant. Suppose also that \( \varepsilon_n^U(f_0) \to 0 \) as \( n \to \infty \). Then for sufficiently large \( n \), there exists an estimator \( \hat{x}_n \) 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 \( \alpha \) and \( d \) such that

\[
\sup_{f_0 \in \Sigma_0^\alpha(M) \cap \Theta_C} \sup_{f \in \Sigma_0^\alpha(M), |f - f_0| = \varepsilon_n^U(f_0)} \Pr \left[ \Omega(\hat{x}_n, f) > C_R \log^n n \cdot (\varepsilon_n^U(f_0) + n^{-1/2}) \right] \leq 1/4.
\]

**Remark 6.** Unlike the (local) smoothness class \( \Sigma_0^\alpha(M) \), the additional function class \( \Theta_C \) 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 \( \hat{x}_n \) does not require knowledge of parameters \( \kappa, c_0, C_0 \) or \( \varepsilon_n^U(f_0) \), and automatically adapts to them, as shown in the next section. While the knowledge of smoothness parameters \( \alpha \) and \( M \) seems to be necessary, we remark that it is possible to adapt to \( \alpha \) and \( M \) by running \( O(\log^2 n) \) parallel sessions of \( \hat{x}_n \) on \( O(\log n) \) grids of \( \alpha \) 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 \( \mu_{f_0}(\varepsilon) \) does not change abruptly with \( \varepsilon \) the expression of \( \varepsilon_n^U(f_0) \) can be significantly simplified. In particular, if for all \( \varepsilon \in (0, c_0] \) it holds that

\[
\mu_{f_0}(\varepsilon/ \log n) \geq \mu_{f_0}(\varepsilon)/[\log n]^{O(1)},
\]

then \( \varepsilon_n^U(f_0) \) can be upper bounded as

\[
\varepsilon_n^U(f_0) \leq [\log n]^{O(1)} \cdot \sup \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geq n \right\}.
\]

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

The quantity \( \varepsilon_n^U(f_0) = \inf \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geq n/ \log^\omega n \right\} \) is crucial in determining the convergence rate of optimization error of \( \hat{x}_n \) locally around the reference function \( f_0 \). While the definition of \( \varepsilon_n^U(f_0) \) is mostly implicit and involves solving an inequality concerning the distribution function \( \mu_{f_0}(\varepsilon) \), we remark that it admits a simple form when \( \mu_{f_0} \) 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 \( \mu_{f_0}(\epsilon) \leq \epsilon^3 \) for some constant \( \beta \in [0, 2 + d/\alpha) \). Then \( \varepsilon_n^U(f_0) = \tilde{O}(n^{-\alpha/(2\alpha+d-\alpha \beta)}) \). In addition, if \( \beta \in [0, d/\alpha] \) then \( \varepsilon_n^U(f_0) + n^{-1/2} \leq \varepsilon_n^U(f_0) = \tilde{O}(n^{-\alpha/(2\alpha+d-\alpha \beta)}) \).

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 \( \varepsilon_n^U(f_0) \geq n/\log^dn \) with the condition \( \mu_{f_0}(\epsilon) \leq \epsilon^3 \). We therefore omit its proof. The following two examples give some simple reference functions \( f_0 \) that satisfy the \( \mu_{f_0}(\epsilon) \leq \epsilon^3 \) condition in Proposition 7 with particular values of \( \beta \).

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

Example 2. \( f_0 \in \Sigma_n^2(M) \) 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 \( \mu_{f_0}(\epsilon) \leq \epsilon^3 \). 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 \( \hat{x}_n \in \arg\min_{x \in \mathcal{X}} \hat{f}(x) \). Since \( \beta \geq 0 \), 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 \( \beta \), 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 \( \beta \) 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 \( R_n(f_0) \approx n^{-\alpha/(2\alpha+d)} \), 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 \( R_n(f_0) \approx n^{-1/2} \), 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 \( \nabla^2 f_0(x) \succeq \sigma I, \forall x \in \mathcal{X} \).

\(^3\)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 \( M(L_{f_0}(\epsilon), \delta) \geq C'_0 \mu_{f_0}(\epsilon) \delta^{-d} \) for all \( \epsilon, \delta \in (0, c'_0] \), where \( M(L_{f_0}(\epsilon), \delta) \) is the maximum number of disjoint \( \ell_2 \) balls of radius \( \delta \) 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

\[
\varepsilon_n^L(f_0) := \sup \left\{ \varepsilon > 0 : \varepsilon^{-(2+d/\alpha)} \mu_{f_0}(\varepsilon) \geq n \right\}.
\]

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

\[
\inf_{\tilde f_n} \sup_{f \in \Sigma_{\kappa}^\alpha(M)} \Pr \left[ \mathcal{L}(\tilde x_n; f) \geq \varepsilon^L_n(f_0) \right] \geq \frac{1}{3}.
\]

**Remark 9.** For any \( f_0 \) and \( n \) it always holds that \( \varepsilon^L_n(f_0) \leq \varepsilon^U_n(f_0) \).

**Remark 10.** If the distribution function \( \mu_{f_0}(\epsilon) \) satisfies Eq. (3.16) in Remark 8, then \( \varepsilon^L_n(f_0) \geq \varepsilon^U_n(f_0)/[\log n]^{O(1)} \).

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 \( \epsilon \), the gap between \( \varepsilon^U_n(f_0) \) and \( \varepsilon^L_n(f_0) \) is only a poly-logarithmic term in \( n \). Additionally, the following proposition derives explicit expression of \( \varepsilon^L_n(f_0) \) 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 \( \mu_{f_0}(\epsilon) \geq \epsilon^\beta \) for some \( \beta \in [0, 2+d/\alpha] \). Then \( \varepsilon^L_n(f_0) = \Omega(n^{-\alpha/(2\alpha+d-\alpha\beta)}) \).

The following proposition additionally shows the existence of \( f_0 \in \Sigma_{\kappa}^\alpha(M) \cap \Theta_C \cap \Theta'_{C'} \) that satisfies \( \mu_{f_0}(\epsilon) \approx \epsilon^\beta \) 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 \( \alpha, \beta, M \) and \( d \) such that \( \mu_{f_0}(\epsilon) \approx \epsilon^\beta \).
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 $\alpha$-H"older 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 $x \in \calX$ 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 $\calA = (\varphi_1, \ldots, \varphi_n, \theta_n)$ 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 $\calA$.

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., $x_1, \ldots, x_n$ drawn independently at uniform from $\calX$) 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 $\alpha, c_0, C_0, \epsilon_0', C_0' > 0$ and $\kappa = \infty$. Denote $C = (c_0, \calC)$ and $C' = (c_0', \calC')$. Then there exist constant $M > 0$ depending on $\alpha, d, C, C'$ and $N$ depending on $M$ such that, for any $f_0 \in \Sigma_n^\alpha(M/2) \cap \Theta_C \cap \Theta_{C'}$ satisfying $\epsilon_n^L(f_0) \leq \epsilon_n^L$, where $\epsilon_n^L(f_0) = \log n/n^{\alpha/(2\alpha+d)}$,

\[
\inf_{\bar{x}_n} \sup_{f \in \Sigma_n^\alpha(M)} \Pr_{\bar{x}_n, f} \left[ \mathcal{L}(\bar{x}_n; f) \geq \epsilon_n^L \right] \geq \frac{1}{3} \quad \text{for all } n \geq N. \tag{3.20}
\]

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 $\calA$ 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 $\beta$), 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 $f: \mathbb{R}^d \rightarrow \mathbb{R}$ 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 $\lambda \in [0, 1]$, $f(\lambda x + (1 - \lambda)x') \leq \lambda f(x) + (1 - \lambda)f(x')$.

(A2) *(Minimizer of bounded $\ell_1$-norm):* We assume there exists $x^* \in \mathcal{X}$ such that $f(x^*) = f^* = \inf_{x \in \mathcal{X}} f(x)$ and $\|x^*\|_1 \leq B$; $x^*$ 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 $\|\nabla f(x)\|_0 \leq s$ and $\|\nabla f(x)\|_1 \leq H$ for all $x \in \mathcal{X}$, where $\|\cdot\|_0$ and $\|\cdot\|_1$ are the $\ell_0$ and $\ell_1$ 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 $\|\nabla^2 f(x)\|_1 \leq H$ for all where $\|A\|_1 := \sum_{i,j=1}^d |A_{ij}|$ is the entry-wise $\ell_1$ norm of matrix $A$.

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

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

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

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

With this setup, the MD method iteratively computes

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

where $\{ \eta_t \}_{t=1}^T$ is a sequence of step sizes and $\tilde{\mathcal{X}} \subseteq \mathcal{X}$ 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 $\tilde{g}_t \approx \nabla f(x_t)$ 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 $\tilde{g}_t$.

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 $n < d$ samples to estimate the unknown gradient $g_t := \nabla f(x_i)$. The high-level idea is to consider $n < d$ 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 $\tilde{y}_i := y_i/\delta$ can be written as

$$\tilde{y}_i = \frac{f(x_t + \delta z_i) + \xi_i}{\delta} = \frac{f(x_t)}{\delta} + g_t^T z_i + \frac{\delta}{2} z_i^T H_t(\kappa_i, z_i) z_i + \delta^{-1} \xi_i := \mu_t + g_t^T z_i + \varepsilon_i, \quad (3.22)$$
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., $\mathbb{E} [\varepsilon_i | z_i, x_i]$) 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 $\mu_t$:

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

(3.23)

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

The following lemma shows that with a carefully chosen $\lambda$, $\hat{g}_t$ is a good estimate of $g_t$ in both $\ell_\infty$ and $\ell_1$ norms.

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

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

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

Lemma 13 follows by the standard $\ell_1$ and $\ell_\infty$ 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 $\mathbb{E} [\varepsilon_i | z_i, x_i]$ are not exactly zero. and we provide a detailed proof in the Appendix.

**Remark 11.** The penalization of $\mu$ 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 $\lambda$ depend on knowledge of $\mu_t$.

**Remark 12.** Lemma 13 reveals an interesting bias-variance tradeoff controlled by the “probing” parameter $\delta > 0$. When $\delta$ is close to 0, the bias (reflected by $\mathbb{E} [\varepsilon_i | z_i, x_i]$) 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 $\hat{g}_t$ is large because the variance of the “stochastic” noise term $\xi_i/\delta$ increases as $\delta \to 0$; on the other hand, for large $\delta$ 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 $\tilde{Y}_t = (\tilde{y}_1, \ldots, \tilde{y}_m) \in \mathbb{R}^m$ and $Z_t = (z_1, \ldots, z_m) \in \mathbb{R}^{n \times d}$ be the vector forms of $\{\tilde{y}_i\}_{i=1}^m$ and $\{z_i\}_{i=1}^m$. Since the design points $z_i$ are i.i.d. Rademacher variables, the de-biased gradient estimator $\tilde{g}_t$ takes a particularly simple form:

$$\tilde{g}_t := \hat{g}_t + \frac{1}{m} Z_t^\top (\tilde{Y}_t - Z_t \hat{g}_t - \hat{\mu}_t \cdot 1_m).$$  

(3.24)
Theorem 7. Theorem 7. Theorem 7. \textcolor{red}{order mirror descent algorithm with sparse gradient estimates.} The following theorem is the main result on the rates of convergence of our proposed zeroth-order mirror descent algorithm with \textcolor{red}{sparse gradient estimates.} Theorem 7. Theorem 7. Theorem 7. \textcolor{red}{order mirror descent algorithm with sparse gradient estimates.}

3.2.4 Rates of convergence

Comparing Lemma 14 with the error bound obtained for the Lasso estimator \( \hat{g}_t \) in Lemma 13, it is clear that the entry-wise bias (i.e., \( \| \gamma_t \|_\infty \)) is reduced from \( O(\delta H + \sqrt{\log d/\delta m}) \) to \( O(\delta H + s \log d/\delta m) \). Such de-biasing is at the cost of inflated stochastic error \( \zeta_t \), which means that unlike \( \hat{g}_t, \tilde{g}_t \) is not a good estimator of \( g_t \) in the \( \ell_1 \) or \( \ell_2 \) 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 \( n = \Omega(s^3 \log^2 d + s(1 + H)^2(1 + B^4H^4 \log^2 d)) \), \( n \leq d \) and that we choose the parameters \( m := \lfloor (1 + H) \sqrt{sn} \rfloor \), \( \eta_t = B \sqrt{\frac{n \log d}{n}} \), and \( \delta_t = \sqrt{s \log d/m} \). Then with probability \( 1 - O(d^{-1}) \)

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

where \( \xi_{\sigma,s} = 1 + \sigma + \sigma^2/s \), and \( \tilde{x}_n \) 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 \( \sigma, s, H, B \) 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) \textcolor{red}{(Hessian smoothness).} There exists \( L > 0 \) such that for all \( x, x' \in \mathcal{X} \),

\[
\| \nabla^2 f(x) - \nabla^2 f(x') \|_1 \leq L \| x - x' \|_\infty
\]

Recall that \( \| A \|_1 = \sum_{i,j} |A_{ij}| \) denotes the entry-wise \( \ell_1 \) norm of a matrix \( A \).

If \( f \) is three-times differentiable, then (A6) is implied by the condition that \( \| \nabla^3 f(x) \|_1 \leq L \) for all \( x \in \mathcal{X} \), where \( \| A \|_1 := \sum_{i,j,k} |A_{ijk}| \) is the entry-wise \( \ell_1 \) 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 \( \tilde{g}_t(\delta) \) in Eqs. (3.23,3.24) corresponding to a probing step size of \( \delta \). Under the additional condition (A6), the analysis in Lemma 14 can be strengthened as below:

51
Lemma 15. Suppose (A1) through (A4) and (A6) hold. Suppose also that $n = \Omega(s^3 \log d)$, $n \leq d$ and $\lambda = \delta^{-1} \sigma \sqrt{\log d/n + \delta H}$. Then with probability $1 - O(d^{-2})$

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

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

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

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

The twice de-biased estimator:

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

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

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

where $\tilde{\zeta}_t = 2\zeta_t(\delta/2) - \zeta_t(\delta)$, $\tilde{\beta}_t = 2\beta_t(\delta/2) - \beta_t(\delta)$ and $\tilde{\gamma}_t = \gamma_t(\delta/2) - \gamma_t(\delta)$.

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 $\frac{\delta}{2} \mathbb{E}[(z^\top H_t z)]$ bias term is cancelled by the “twicing” trick, and the remaining bias term $\tilde{\gamma}$ is an order of magnitude smaller than $\gamma$ in the bias term before twicing (e.g., Lemma 14). We also remark that the twice de-biased estimator $\tilde{g}_t^{\text{tw}}$ 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 $\tilde{g}_t^{\text{tw}}$ into the stochastic mirror descent procedure and choosing tuning parameters $n, \lambda, \delta$ and $\eta$ 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 $T \leq d$. Let $\eta := Bn^{2/3} \sqrt{\log d / T}$, $n := [(1 + L) s^{2/3} \sqrt{T}]$ and $\delta := (s \log d / n)^{1/3}$. Then with probability $1 - O(d^{-1})$

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

52
where $\tilde{\xi}_{s,s} = (1 + \sigma + \sigma^2/s^{2/3})$ and $\tilde{x}_n$ 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 $f(x) = \frac{1}{2}(x - x^*)^TQ(x - x^*)$ with (unknown) $Q \succeq 0$ being positive semi-definite and supported on $S \subseteq [d]$ with $|S| \leq s$, 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 ($\tilde{g}_t^{tw}$) 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-
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 \( f_S(x_S) = x_S^T Q x_S + b^T x_S \) 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 \( f_S(x_S) = x_S^T Q x + b^T x_S \) where we set \( Q_{ii} = i^{-\gamma} \) where \( \gamma \) 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 \( f_S(x_S) = |(x_S - b)^T Q (x_S - b)|^2 + (x_S - b)^T Q (x_S - b) \) 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 \( t \leq n \).

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

(B3) (Bounded gradients): $f$ is differentiable on $\mathbb{R}^d$ and furthermore $\sup_x \|\nabla f(x)\|_q \leq L$.

The constraint $\|x^*\|_p \leq B$ for $p \in [1, 2]$ is a weaker form of constraining the optimal solution $x^*$ to be sparse. More specifically, as the norm $p$ moves from 2 to 1, the optimal solution $x^*$ has to be sparser in order to satisfy the more stringent constraint $\|x^*\|_p \leq B$. Nevertheless, the condition is considerably weaker than the exact sparsity constraint (e.g., $\|x^*\|_0 \leq s$), 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 $\tilde{f}(x) := \mathbb{E}_{u \sim \mu}[f(x + u)]$, where $\mu$ is a distribution supported on $\mathcal{X} = \mathbb{R}^d$. The smoothed function $\tilde{f}$ is constructed so that $|\tilde{f}(x) - f(x)|$ is small, and furthermore the gradient $\nabla \tilde{f}(x)$ can be unbiasedly estimated;

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

3. The mirror descent update is preformed:

$$x_{t+1} \in \arg \min_{x \in \mathcal{X}} \{\eta_t g_t + D_\psi(x, x_t)\},$$

(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 $\psi$, which is $\kappa$-strongly convex with respect to $\|\cdot\|_p$ (i.e., $D_\psi(x, y) \geq \kappa \|x - y\|_p^2/2$).

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

$$\mathbb{E}[h(u) \nabla \log p(u)] = -\mathbb{E}[\nabla h(u)] \quad \text{where} \quad u \sim \mu.$$  

(3.27)

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

$$\nabla \tilde{f}(x_t) = \mathbb{E}_{\mu}[f(x_t + u) \nabla \log p(u)].$$

(3.28)

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

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

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

$$p(x) = p(x_1, \ldots, x_d) = \prod_{i=1}^d p_i(x_i) \quad \text{where} \quad p_i(x_i) = \frac{p^{1-1/p}}{2\delta_0^{2\Gamma(1/p)}} \exp \left\{-\frac{|x_i|^p}{\delta_0^p} \right\}.$$  

(3.30)
When \( p \in (1, 2] \), \( \log p \) is differentiable everywhere, and the gradient of \( \log p \) takes the form of \(-\nabla \log p(x) = \tilde{x}/\delta^p\), where \( \tilde{x} = [\text{sgn}(x_1)|x_1|^{p-1}, \ldots, \text{sgn}(x_d)|x_d|^{p-1}] \). The gradient estimate \( \tilde{g}_t \) can then be written as

\[
\tilde{g}_t = (y_t - \eta_t)\delta^{-p}\tilde{u}_t \quad \text{where} \quad \tilde{u}_t = [\text{sgn}(u_{t1})|u_{t1}|^{p-1}, \ldots, \text{sgn}(u_{td})|u_{td}|^{p-1}].
\] (3.31)

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

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

**Lemma 17.** \( |\tilde{f}(x) - f(x)| \leq L\mathbb{E}[\|u\|_p] \leq L(1 + \delta)\delta^{1/p} \log d \) for all \( x \in \mathcal{X} \).

**Lemma 18.** \( \mathbb{E}[\|\tilde{g}_t\|_\infty^2] \leq \mathbb{E}[\|2\delta^2 + L^2\|_p^2]\|\nabla \log p(u)\|_q^2] \leq \sigma^2(1 + \delta)\delta^{-2}d^2q\log^2 d + (1 + \delta)d^2\log^4 d \).

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 (\( \delta \)) are carefully selected:

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

\[
\mathbb{E}f(\tilde{x}_n) - f(x^*) = \tilde{O}\left(\left[\frac{\sigma^2 d^2 BL}{n}\right]^{1/4} + \sqrt{\frac{d^2 BL^2}{n}}\right).
\] (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 \( \delta \ll 1 \). Also, in the \( \tilde{O}(\cdot) \) notation we omit poly-logarithmic dependency on \( d \).

**Remark 15.** Regardless of the values of \( p \) and \( q \) appearing in the conjugate norms that define the boundedness of \( x^* \) and \( \nabla f \), 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 \( \ell_p \) geometry: the constrained case

One disadvantage of the Stein-Hudson’s gradient estimator is that the support of the probing distribution \( \mu \) 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 \( \ell_p \) 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\footnote{A convex set $K \subset \mathbb{R}^d$ is symmetric if for all $x \in \mathbb{R}^d$, $x \in K \iff -x \in K$.} convex set in $\mathbb{R}^d$. Let $\nu_K$ and $\sigma_{\partial K}$ be the uniform measure on $K$ and its boundary $\partial K$, respectively. For any convex, differentiable function $f : \mathbb{R}^d \to \mathbb{R}$, a "smoothed version" $\tilde{f} : \mathbb{R}^d \to \mathbb{R}$ is defined as

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

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

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

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

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

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

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

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

Combining Lemmas 19, 20 and using the potential function $\psi(x) = \frac{1}{p-1}\|x\|_p^2$, 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 $|f(x)| \leq C$ for all $x \in \mathcal{X}$. Suppose also that $\|x^*\|_p \leq b$, and define $\bar{x}_n = (\sum_{t=1}^n x_t)/n$. Then

$$
\mathbb{E} f(\bar{x}_n) - f(x^*) \lesssim \frac{b}{\eta n} + \eta \cdot \max_{t \leq n} \sqrt{\mathbb{E} \|\hat{g}_t(x_t)\|_q^2} + \delta L \cdot \mathbb{E}_{u \sim \nu(K)}[\|u\|_p]
$$

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

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

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

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

$$
\kappa(p, 2(p - 1)) \quad \text{where} \quad \kappa(p, r) := \frac{\Gamma((1 + r)/p)}{\Gamma(1/p)}.
$$
**Lemma 23.** For any pair of conjugate norms $p, q \in (1, \infty)$, $1/p + 1/q = 1$, there exists a finite positive constant $C_p$ such that for sufficiently large $d$, for $K = \mathbb{R}^d_p$,

$$
E_{\nu \sim \sigma_E} [\|\ell(v)\|_q^2] \leq C_p \cdot d^{1/q - 1/p}.
$$

(3.36)

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

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

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

### 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) $\pi$ 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 $\{x_t\}_{t=1}^T$:

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

(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 $\sup_{x, x' \in \mathcal{X}} \|x - x'\|_2 \leq D$;

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

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

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

(A5) (**Strong convexity**): there exists constant $M > 0$ such that $f_t(x') \geq f_t(x) + \nabla f_t(x)^\top (x' - x) + \frac{M}{2} \|x' - x\|_2^2$ 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 $f : \mathcal{X} \rightarrow \mathbb{R}$, define

$$
\|f\|_p := \left\{ \left( \frac{1}{\text{vol}(\mathcal{X})} \int_{\mathcal{X}} |f(x)|^p \text{d}x \right)^{1/p} \right\} \quad 1 \leq p < \infty; \quad p = \infty.
$$

(3.38)

Here $\text{vol}(\mathcal{X}) = \int_{\mathcal{X}} 1 \text{d}x$ 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 $\left( \int_{\mathcal{X}} |f(x)|^p \text{d}x \right)^{1/p}$. 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 $f_1, \cdots, f_T : \mathcal{X} \rightarrow \mathbb{R}$, define the $L_{p,q}$-variation functional of $f = (f_1, \cdots, f_T)$ as

$$
\text{Var}_{p,q}(f) := \left\{ \left( \frac{1}{T} \sum_{t=1}^{T-1} \|f_{t+1} - f_t\|_p^q \right)^{1/q} \right\} \quad 1 \leq p \leq \infty, 1 \leq q < \infty; \quad 1 \leq p \leq \infty, q = \infty.
$$

(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

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

(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 $f_i(x) = \sum_{i=1}^m \mathbb{I}[x \in \mathcal{X}_i] \cdot (a_{ti}x^3 + b_{ti}x^2 + c_{ti}x + d_{ti})$, where $\mathbb{I}[:]$ 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 $f_i$ 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 $x \in \mathcal{X} \setminus (\mathcal{X}_i \cup \mathcal{X}_{i+1})$ and
\[
\sup_{x \in X_i \cup X_{i+1}} |f_t(x) - f_{t+1}(x)| \leq \delta < \infty. \quad \text{We then have that (noting that } \text{vol}(X) = 1 \text{ and } \text{vol}(X_i \cup X_{i+1}) = 2/m) \\
\|f_t - f_{t+1}\|_p = \left(\int_{X_i \cup X_{i+1}} |f_t(x) - f_{t+1}(x)|^p dx\right)^{1/p} \leq \delta \cdot (m/2)^{-1/p}.
\]

Because \( \delta \cdot (m/2)^{-1/p} \) is an increasing function of \( p \), the parameter \( p \) controls the \textit{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 \textit{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 \( \mathcal{F}_{\infty,q}(V_T) \).

\textit{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 \( \Delta \) 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 \textit{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.

\textbf{Theorem 11} (Upper bound). \textit{Fix arbitrary } \( 1 \leq p < \infty \) \text{ and } \( 1 \leq q < \infty \). \textit{Suppose (A1) through (A5) hold and } \( 0 \leq V_T \leq 1 \). \textit{Then there exists a computationally efficient policy } \( \pi \) \text{ and } \( C_1 > 0 \) as a polynomial function of } \( \log T \) \text{ and } \( \log V_T \) \text{ such that}
\[
\sup_{f \in \mathcal{F}_{p,q}(V_T)} \text{R}^{\pi}_f (f) \leq C_1 \cdot T \cdot V_T^{2p/(6p+d)}.
\]

\textit{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 \textit{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 $q \geq 1$ 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 evenly across all $t \in T$ (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 $V_T \leq 1$). The condition $V_T \leq 1$ 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 $x_0 \in X$ 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:

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

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 $\Delta_T$ is tuned as a function of $p$.

2. Within each interval/batch of the meta-policy, a proper sub-policy $\pi_s$ 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 $\Delta_T$; sub-policy $\pi_s$.

1. Divide epochs $\{1, \ldots, T\}$ into $J = \lceil T/\Delta_T \rceil$ batches $B_1, \ldots, B_J$ such that $B_1 = \{b_1, \ldots, \bar{b}_1\}$, $B_2 = \{b_2, \ldots, \bar{b}_2\}$, etc., with $b_1 = 1$, $\bar{b}_J = T$ and $b_{\ell+1} = \bar{b}_\ell + 1$ for $\ell = 1, \ldots, J-1$. The epochs are divided as evenly as possible, so that $|B_\ell| \in \{\Delta_T, \Delta_T+1\}$ for all $\ell = 1, \ldots, J$.

2. For each batch $B_\ell$, $\ell = 1, \ldots, J$, do the following:
   
   (a) Run sub-policy $\pi_s$ with $\bar{b}_\ell$ and $\bar{b}_\ell$, corresponding to $f_{\bar{b}_\ell}, f_{\bar{b}_\ell+1}, \ldots, f_{\bar{b}_J}$.

The key idea behind the meta-policy is to “restart” certain sub-policy $\pi_s$ after $\Delta_T$ epochs. This strategy ensures that the sub-policy $\pi_s$ has sufficient number of epochs to exploit feedback information, while at the same time avoids usage of outdated feedback. Scalings of $\Delta_T$ 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 $\pi_s$ 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 $e_j = (0, \ldots, 1, \ldots, 0) \in \mathbb{R}^d$. This idea originates from the seminal work of Yudin & Nemirovskii (1983) and was applied to convex bandits problems (e.g., Bèsbes 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 $\mathcal{X}^o \neq \emptyset$. A convex function $\varphi : \mathcal{X}^o \to \mathbb{R}$ is a $\kappa$-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 $\lim_{n \to \infty} x_n \in \partial \mathcal{X}$ 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 $|\nabla^3 \varphi(x)[z, z, z]| \leq 2|z^T \nabla^2 \varphi(x) z|^{3/2}$ and $|z^T \nabla \varphi(x)| \leq \kappa^{1/2}|z^T \nabla^2 \varphi(x) z|^{1/2}$, where $\nabla^3 \varphi(x)[z, z, z] = \frac{\partial^3}{\partial t_1 \partial t_2 \partial t_3} \varphi(x + t_1 z + t_2 z + t_3 z)|_{t_1 = t_2 = t_3 = 0}$.

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 $\kappa$-self-concordant barrier function $\varphi$ with $\kappa = O(d)$, and furthermore for bounded $\mathcal{X}$ the barrier $\varphi$ 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 $\mathcal{X} = \{x : Ax \leq b\}$ 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.
Suppose Proposition 10. \(X\) belongs to the domain any 3.3.5 Minimax lower bounds

A in (Hazan & Levy, 2014). An important aspect of step 2(c) is the clever choice of the matrix regularity conditions. Finally, step 2(c) is a random perturbation step originally considered in McMahan (2017) (Sec. 6) that the FTRL step is equivalent to mirror descent under minimal convergence rates of optimization algorithms measured in non-standard metric. It was shown a regularization term (\(X\) being too close to the boundary of \(\nabla f(x)\), change-of-variable formula and the smoothness of \(f\) is very close to the boundary of \(\mathcal{X}\). The FTRL step is essentially a mirror descent, which uses a regularization term (\(\phi(\cdot)\) 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 \(x^*\) is very close to the boundary of \(\mathcal{X}\). More specifically, the following proposition shows that \(x_{b,t} = z_t + A_t u_t\) always belongs to the domain \(\mathcal{X}\), justifying the correctness of policy \(\pi_s\).

**Proposition 10.** Suppose \(\phi\) is strictly convex on \(\mathcal{X}^o\). Then for any \(x \in \mathcal{X}^o\), \(\delta \geq 0\) and \(u \in \mathbb{S}_d\), a change of variable formula is needed to obtain feedback.

\[ x + (\nabla^2 \phi(x) + \delta I_d)^{-1/2} u \in \mathcal{X}. \]

### 3.3.5 Minimax lower bounds

We prove the following result, establishing a lower bound of worst-case regret possible for any policy \(\pi\).

**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

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

**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 question. 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 $\alpha \geq 1$ 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 $\pi$ is compared against the optimal stationary benchmark $\inf_{x \in X} \{ \sum_{t=1}^T f_t(x) \}$ 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 \( x^*_t, \ldots, x^*_T \) subject to certain smoothness constraints like \( \sum_t \| x^*_{t+1} - x^*_t \| \leq C \). Jadabaie 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 \( f_t(x_i) \) or \( \nabla f_t(x_i) \) 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 \( x \sim N_0(0, I_{d+x}) \) and let \( A \) be a \( d \times d \) positive semi-definite matrix. Then for all \( t > 0 \),

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

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 \( B^\infty_{h_t}(x; \mathcal{X}) \).

**Lemma 25.** Suppose \( f \) satisfies Eq. (3.6) on \( B^\infty_{h_t}(x; \mathcal{X}) \). Then there exists \( \tilde{f}_x \in \mathcal{P}_k \) such that

\[
\sup_{z \in B^\infty_{h_t}(x; \mathcal{X})} |f(z) - \tilde{f}_x(z)| \leq Md^k h^\alpha.
\]

**Proof.** Consider

\[
\tilde{f}_x(z) := f(x) + \sum_{j=1}^k \sum_{\alpha_1 + \cdots + \alpha_d = j} \frac{\partial^j f(x)}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \prod_{\ell=1}^d (z_\ell - x_\ell)^{\alpha_\ell}.
\]

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

\[
|f(z) - f(x)| \leq \sum_{\alpha_1 + \cdots + \alpha_d = k} |f^{(\alpha)}(x + \xi(z - x)) - f^{(\alpha)}(x)| \cdot \prod_{\ell=1}^d |z_\ell - x_\ell|^{\alpha_\ell}.
\]
Because $f$ satisfies Eq. (3.6) on $B_{h}^{c} (x; \mathcal{X})$, we have that $|f^{(c)}(x + \xi (z - x)) - f^{(c)}(x)| \leq M \cdot \|z - x\|_{\infty}^{-k}$. Also note that $|z_{t} - x_{t}| \leq \|z - x\|_{\infty} \leq h$ for all $z \in B_{h}^{c} (x; \mathcal{X})$. The lemma is thus proved.

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

$$
\tilde{\theta}_{h} = (\Psi_{t,h}^{\top} \Psi_{t,h})^{-1} \Psi_{t,h}^{\top} Y_{t,h}.
$$

(3.45)

In addition, because $\tilde{f}_{x} \in \mathcal{P}_{k}$, there exists $\tilde{\theta} \in \mathbb{R}^{D}$ such that $\tilde{f}_{x}(z) \equiv \psi_{x,h}(z)^{\top} \tilde{\theta}$. Denote also that $F_{t,h} := (f(x))_{1 \leq t \leq n} \in B_{h}^{c} (x)$, $\Delta_{t,h} := (f(x_{t}) - \tilde{f}_{x}(x_{t}))_{1 \leq t \leq n}$, and $W_{t,h} := (w_{t})_{1 \leq t \leq n}$. Eq. (3.45) can then be re-formulated as

$$
\tilde{\theta}_{h} = (\Psi_{t,h}^{\top} \Psi_{t,h})^{-1} \Psi_{t,h}^{\top} \Delta_{t,h} + W_{t,h}.
$$

(3.46)

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

(3.47)

Because $\frac{1}{m} \Psi_{t,h}^{\top} \Psi_{t,h} \geq \sigma I_{D \times D}$ and $\sup_{z \in B_{h}^{c} (x)} \|\psi_{x,h}(z)\|_{2} \leq b$, we have that

$$
\|\tilde{\theta}_{h} - \tilde{\theta}\|_{2} \leq \frac{b}{\sigma} \|\Delta_{t,h}\|_{\infty} + \left\| \left[ \frac{1}{m} \Psi_{t,h}^{\top} \Psi_{t,h} \right]^{-1} \left[ \frac{1}{m} \Psi_{t,h}^{\top} W_{t} \right] \right\|_{2} .
$$

(3.48)

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

$$
\left[ \frac{1}{m} \Psi_{t,h}^{\top} \Psi_{t,h} \right]^{-1} \frac{1}{m} \Psi_{t,h}^{\top} W_{t} \sim \mathcal{N}_{D} \left( 0, \frac{1}{m} \left[ \frac{1}{m} \Psi_{t,h}^{\top} \Psi_{t,h} \right]^{-1} \right) .
$$

(3.49)

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

$$
\left\| \left[ \frac{1}{m} \Psi_{t,h}^{\top} \Psi_{t,h} \right]^{-1} \frac{1}{m} \Psi_{t,h}^{\top} W_{t} \right\|_{2} \leq \sqrt{\frac{5D \log(1/\delta)}{\sigma m}} .
$$

(3.50)

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

$$
\|\tilde{\theta}_{h} - \tilde{\theta}\|_{2} \leq \frac{b}{\sigma h} M d^{k} h^{\alpha} + \sqrt{\frac{5D \log(1/\delta)}{\sigma m}} .
$$

(3.51)

Finally, noting that

$$
|\hat{f}_{h}(x) - f(x)| = |\hat{f}_{h}(x) - \tilde{f}_{x}(x)| = |\psi(x)^{\top} (\tilde{\theta}_{h} - \tilde{\theta})| \leq b \|\tilde{\theta}_{h} - \tilde{\theta}\|_{2}
$$

(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. $\sup_{x \in \mathcal{X}} \min_{x \in G_n} \|x - x^\prime\|_\infty = \tilde{O}(n^{-3/\min(a,1)})$;

2. $f_n^* - f^* = \tilde{O}(n^{-3})$.

**Proof.** Let $H_N \subseteq \mathcal{X}$ be the finite subset of $\mathcal{X}$ such that $|H_N| = N$ and $\sup_{x \in \mathcal{X}} \min_{x \in H_n} \|x - x^\prime\|_\infty$ 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 $\sup_{x \in \mathcal{X}} \min_{x \in H_n} \|x - x^\prime\|_\infty \leq N^{-1/d}$.

For any $x \in H_n$, consider an $\ell_\infty$ 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 $P_X(x \in B_{r_n}^\infty(x)) \geq r_n^d$. Therefore, applying union bound over all $x \in H_n$ we have that

$$P_X \left( \exists x \in H_n, G_n \cap B_{r_n}^\infty(x) = \emptyset \right) \leq N(1 - r_n^d)^{|G_n|} \leq \exp\left\{-r_n^d|G_n| + \log N\right\}. \quad (3.53)$$

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

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

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

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 $O(\sqrt{n} \log^2 n)$ uniform samples in $B_{r_0}^\infty(x; \mathcal{X})$. Subsequently, by standard Gaussian concentration inequality, with probability $1 - O(n^{-1})$ we have

$$\inf_{x \in B_{r_0}^\infty(x; \mathcal{X})} f(z) - O(n_0^{-1/4}) \leq \tilde{f}(x) \leq \sup_{x \in B_{r_0}^\infty(x; \mathcal{X})} f(z) + O(n_0^{-1/4}) \quad \forall x \in G_n. \quad (3.55)$$

Fix arbitrary $\tilde{x}^* \in \arg \min_{x \in G_n} f(x)$. Because $f \in \Sigma_n^0(M)$ for constant $\kappa$ and $h_0 \to 0$, $f$ is smooth on $B_{r_0}^\infty(\tilde{x}^*; \mathcal{X})$ and therefore $\sup_{x \in B_{r_0}^\infty(\tilde{x}^*; \mathcal{X})} f(z) \leq f(\tilde{x}^*) + O(h_0^{\min(a,1)}) \leq f(\tilde{x}^*) + O(1/\log^2 n) \leq f^* + O(1/\log^2 n)$, where the last inequality holds due to Lemma 26. On the other hand, for all $x \in G_n$, $\tilde{f}(x) \geq f^* - O(n_0^{-1/4})$. Therefore, for sufficiently large $n$ we must have $\tilde{f}(\tilde{x}^*) \leq \min_{x \in G_n} \tilde{f}(z) + 1 + \log n$ and subsequently $\tilde{x}^* \in S_0^\circ$.

We next prove the statement that $S_0^\circ \subseteq \bigcup_{x \in L_\upsilon(\kappa/2)} B_{r_0}^\infty(x; \mathcal{X})$. Consider arbitrary $z \in G_n$ and $z \notin \bigcup_{x \in L_\upsilon(\kappa/2)} B_{r_0}^\infty(x; \mathcal{X})$. By definition, $f(z^\prime) \geq f^* + \kappa/2$ for all $z^\prime \in B_{r_0}^\infty(z; \mathcal{X})$. Subsequently, $\tilde{f}(z) \geq f^* + \kappa/2 - O(n_0^{-1/4}) > f^* + 1/\log n$ for constant $\kappa > 0$ and sufficiently large $n$, which implies $z \notin S_0^\circ$.
3.5.3 Proof of Theorem 4

We prove the theorem by considering every reference function \( f_0 \in \Sigma_\kappa^\Theta(M) \cap \Theta_C \) 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 \( d, \alpha, M, C, p \) in \( O(\cdot), \Omega(\cdot), \Theta(\cdot) \). With \( n \) being sufficiently large, we further suppress logarithmic terms of \( n \) in \( \tilde{O}(\cdot) \) and \( \tilde{\Omega}(\cdot) \) notations.

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

**Lemma 27.** For \( \tau = 1, \ldots, T \) define \( \epsilon_\tau := \max\{\tilde{c}_0 \cdot 2^{-\tau}, C_3 [\epsilon_n^U(f_0) + n^{-1/2}] \log^2 n\} \), where \( C_3 > 0 \) is a constant depending only on \( d, \alpha, M, p_0, p_0 \) and \( C \). Then for sufficiently large \( n \), with probability \( 1 - O(n^{-1}) \) the following holds uniformly for all outer iterations \( \tau = 1, \ldots, T \):

\[
S_\tau \subseteq L_f(\epsilon_\tau). \tag{3.56}
\]

Lemma 27 shows that the level \( \epsilon_\tau \) in \( L_f(\epsilon_\tau) \) that contains \( S_{\tau-1} \) shrinks geometrically, until the condition \( \epsilon_\tau \geq C_3 [\epsilon_n^U(f_0) + n^{-1/2}] \log^2 n \) is violated. If the condition is never violated, then at the end of the last epoch \( \tau^* \) we have \( \epsilon_{\tau^*} = O(n^{-1}) \) 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 \( \epsilon_{\tau^*} \leq [\epsilon_n^U(f_0) + n^{-1/2}] \log^2 n \) 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 \( 1 \leq t \leq n \) and \( x \in G_n \);
2. \( x^*_t \in S_\tau \) for all \( 0 \leq \tau \leq n \).

**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 \( \ell_t(x^*_n) \leq f_n^* \) and \( \min_{x \in S_{\tau-1}} u_t(x) \geq f_n^* \) for all \( \tau \).

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 \( x \in G_n \), \( 1 \leq m \leq n \) and \( h > 0 \), let \( K_{h,m}^0(x), \ldots, K_{h,m}^n(x) \) be \( n \) independent point sets, where each point set consists of \( m \) points sampled i.i.d. uniformly at random from \( B_n^\kappa(x; G_n) \). With probability \( 1 - O(n^{-1}) \) the following holds true uniformly for all \( x \in G_n \), \( h \in \{j/n^2 : j \in \mathbb{N}, j \leq n^2\} \) and \( K_{h,m}^\ell(x), \ell \in [n] \) as \( n \to \infty \):

1. \( \sup_{h > 0} \sup_{z \in B_n^\kappa(x)} \| \psi_{x,h}(z) \|_2 \geq \Theta(1) \);
2. \( |B_n^\kappa(x; G_n)| \approx h^d|G_n| \).
3. \( \sigma_{\min}(K_{h,m}^\ell(x)) = \Theta(1) \) for all \( m \geq \Omega(\log^2 n) \) and \( m \leq |G_n| \), where \( \sigma_{\min}(K_{h,m}^\ell(x)) \) is the least eigenvalue of \( \frac{1}{m} \sum_{x \in K_{h,m}^\ell(x)} \psi_{x,h}(z) \psi_{x,h}(z)^T \).

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 \( h_\tau \) selected at the end of each outer iteration \( \tau \) is near-optimal, being sandwiched between two quantities determined by the size of the active sample grid \( \tilde{S}_{\tau-1} := S_{\tau-1}^0(\tau_{\tau-1}) \).

**Lemma 29.** There exist constants \( C_1, C_2 > 0 \) depending only on \( d, \alpha, M, \beta_0, \tau_0 \) and \( C \) such that with probability \( 1 - O(n^{-1}) \), the following holds for every outer iteration \( \tau \in \{1, \ldots, T\} \) and all \( x \in S_{\tau-1}^0 \):

\[
C_1[\tilde{v}_{\tau-1} n_0]^{-1/(2\alpha+d)} - \tau/n \leq \varrho_\tau(x) \leq C_2[\tilde{v}_{\tau-1} n_0]^{-1/(2\alpha+d)} \log n + \tau/n, \tag{3.57}
\]

where \( \tilde{v}_{\tau-1} := |G_n|/|\tilde{S}_{\tau-1}| \).

We are now ready to state the proof of Lemma 27, which is based on an inductive argument over the epochs \( \tau = 1, \ldots, T \).

**Proof.** We use induction to prove this lemma. For the base case \( \tau = 1 \), because \( \|f - f_0\|_\infty \leq \varepsilon_n^U(f_0) \) and \( \varepsilon_n^U(f_0) \to 0 \) as \( n \to \infty \), it suffices to prove that \( S_1 \subseteq L_{f_0}(c_\tau/4) \) for sufficiently large \( n \). Because \( \tilde{S}_0 = S_0 = G_n \), invoking Lemmas 29 and 12 we have that \( |u_\tau(x) - \ell_\tau(x)| = \tilde{O}(n^{-\alpha/(2\alpha+d)}) \) 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 \( \sup_{x \in G_n} |u_\tau(x) - \ell_\tau(x)| \leq c_\tau/8 \) and hence \( S_1 \subseteq L_{f_0}(c_\tau/4) \).

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

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

\[
S_{\tau-1} \subseteq L_{f_0}(\varepsilon_{\tau-1}) \subseteq L_{f_0}(\varepsilon_{\tau-1} + \|f - f_0\|_\infty) \subseteq L_{f_0}(2\varepsilon_{\tau-1}). \tag{3.58}
\]

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

\[
\tilde{S}_{\tau-1} = S_{\tau-1}^0 \subseteq L_{f_0}(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*). \tag{3.59}
\]

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 \( B_{\rho_{\tau-1}^*}^2(x) = \{ z \in X : \| z - x \|_2 \leq \rho_{\tau-1}^* \} \) is the \( \ell_2 \) ball of radius \( \rho_{\tau-1}^* \) centered at \( x \). By (A2), we know that \( |H_n| \leq 1 + [\rho_{\tau-1}^*]^{-d} |\mu_{f_0}(2\varepsilon_{\tau-1})| \). In addition, the enlarged level set satisfies \( \mu_{f_0}^0(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*) \subseteq \bigcup_{x \in H_n} B_{2\rho_{\tau-1}^*}^2(x) \). Subsequently,

\[
\mu_{f_0}^0(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*) \leq |H_n| \cdot [\rho_{\tau-1}^*]^{-d} \leq |\mu_{f_0}(2\varepsilon_{\tau-1})| + [\rho_{\tau-1}^*]^{-d}. \tag{3.60}
\]
By Lemma 29, the monotonicity of $|\tilde{S}_{\tau-1}|$ and the fact that $p_0 \leq p_X(z) \leq \bar{p}_0$ for all $z \in \mathcal{X}$, we have

$$
\rho_{\tau-1}^* \leq [\mu_f(\varepsilon_{\tau-1}, \rho_{\tau-1}^*)]^{1/(2a+d)} n_0^{-1/(2a+d)} \log n
$$

(3.61)

$$
\leq [\mu_f(2\varepsilon_{\tau-1}, \rho_{\tau-1}^*)]^{1/(2a+d)} n_0^{-1/(2a+d)} \log n
$$

(3.62)

$$
\leq \left( \mu_f(2\varepsilon_{\tau-1}) + \left[ \rho_{\tau-1}^* \right]^{d} \right)^{1/(2a+d)} n_0^{-1/(2a+d)} \log n.
$$

(3.63)

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

$$
\rho_{\tau-1}^* \leq \max \left\{ \left[ \mu_f(2\varepsilon_{\tau-1}) \right]^{\frac{1}{2a+d}} n_0^{-\frac{a}{2a+d}}, n_0^{-\frac{1}{2a+d}} \log n \right\}.
$$

(3.64)

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

$$
\eta_{\delta_t(x), \delta}(x) \leq b_{\delta_t(x), \delta}(x) \leq [h_t(x)]^\alpha
$$

(3.65)

$$
\leq \left[ \nu_{\tau-1} n_0 \right]^{-\alpha/(2a+d)} \log n
$$

(3.66)

$$
\leq \left[ \nu_{\tau-2} n_0 \right]^{-\alpha/(2a+d)} \log n
$$

(3.67)

$$
\leq \left[ \rho_{\tau-1}^* \right]^\alpha \log n.
$$

(3.68)

Here Eq. (3.66) holds by invoking the upper bound on $h_t(x)$ in Lemma 29, Eq. (3.67) holds because $\nu_{\tau-1} \geq \nu_{\tau-2}$, and Eq. (3.68) holds by again invoking the lower bound on $\rho_{\tau-1}(x)$ in Lemma 29. Combining Eqs. (3.64,3.65) we have

$$
\max_{x \in S_{\tau-1}} \eta_{\delta_t(x), \delta}(x) \leq \max \left\{ \left[ \mu_f(2\varepsilon_{\tau-1}) \right]^{\frac{1}{2a+d}} n_0^{-\frac{a}{2a+d}} \log^2 n, n_0^{-\frac{1}{2a+d}} \log n \right\}.
$$

(3.69)

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

$$
\left[ \mu_f(2\varepsilon_{\tau-1}) \right]^{\frac{1}{2a+d}} n_0^{-\frac{a}{2a+d}} \log^2 n \leq 2\varepsilon_{\tau-1} n^{\frac{a}{2a+d}} \log^{\frac{2\omega+a}{2a+d}} n \cdot n_0^{-\frac{a}{2a+d}} \log^2 n
$$

(3.70)

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

(3.71)

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

$$
\max_{x \in S_{\tau-1}} \eta_{\delta_t(x), \delta}(x) \leq \varepsilon_{\tau-1}/4.
$$

(3.72)

Lemma 27 is thus proved. $\square$

---

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

---

71
Proof of Lemma 28  We first show that the first property holds almost surely. Recall the definition of \( \psi_{x,h} \), we have that \( 1 \leq \| \psi_{x,h}(z) \|_2 \leq D \cdot [\max_{1 \leq j \leq d} h^{-1} |z_j - x_j|]^k \). Because \( \| z - x \|_\infty \leq h \) for all \( z \in B^c_h(x) \), \( \sup_{z \in B^c_h(x)} \| \psi_{x,h}(z) \|_2 \leq O(1) \) for all \( h > 0 \). Thus, \( \sup_{h>0} \sup_{z \in B^c_h(x)} \| \psi_{x,h}(z) \|_2 = \Theta(1) \) 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

\[
\max_{x,h} \left| \frac{\| B^c_h(x; G_n) \|}{|G_n|} - P_X(z \in B^c_h(x)) \right| \leq \sqrt{\frac{\log n}{|G_n|}}. \tag{3.73}
\]

In addition, note that \( P_X(z \in B^c_h(x; \mathcal{X})) \geq p_1 \lambda(B^c_h(x; \mathcal{X})) \geq h^d \) and \( P_X(z \in B^c_h(x; \mathcal{X})) \leq p_0 \lambda(B^c_h(x; \mathcal{X})) \leq h^d \), where \( \lambda(\cdot) \) denotes the Lebesgue measure on \( \mathcal{X} \). Subsequently, \( |B^c_h(x; G_n)| \) 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 \( h_d \geq n^{-1} \) and \( |G_n| \geq n^{3d/\min(\alpha,1)} \).

We next prove the third property. Because \( p_0 \leq p_X(z) \leq p_0 \) for all \( z \in \mathcal{X} \), we have that

\[
P_0 \int_{B^c_h(x; \mathcal{X})} \psi_{x,h}(z) \psi_{x,h}(z)^\top dU(z, h, z) \leq E \left[ \frac{1}{m} \sum_{z \in K^l_{h,m}} \psi_{x,h}(z) \psi_{x,h}(z)^\top \right] \leq p_0 \int_{B^c_h(x; \mathcal{X})} \psi_{x,h}(z) \psi_{x,h}(z)^\top dU(z, h, z), \tag{3.74}
\]

where \( U(z, h, z) \) is the uniform distribution on \( B^c_h(x; \mathcal{X}) \). Note also that

\[
\int_{\mathcal{X}} \psi_{0,1}(z) \psi_{0,1}(z)^\top dU(z) \leq \int_{B^c_h(x; \mathcal{X})} \psi_{x,h}(z) \psi_{x,h}(z)^\top dU(z) \tag{3.76}
\]

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

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 \( 0 < \psi_0 \leq \Psi_0 < \infty \) 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

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

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

\[
\max_{x,h,m,l} \left\| \frac{1}{m} \sum_{z \in K^l_{h,m}(x)} \psi_{x,h}(z) \psi_{x,h}(z)^\top - E \left[ \psi_{x,h}(z) \psi_{x,h}(z)^\top | z \in B_h(x) \right] \right\|_\infty \leq \sqrt{\frac{\log n}{m}}. \tag{3.80}
\]
Combining Eqs. (3.79,3.80) and applying Weyl’s inequality (Lemma 99) we have
\[ \Omega(1) - O(\sqrt{\log n/m}) \leq \sigma_{\min}(K_{h,m}^{\tau}(x)) \leq O(1) - O(\sqrt{\log n/m}). \] (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 \( \tilde{S}_0 = S_0 = G_n \) and therefore \( \tilde{\nu}_{\tau-1} = 1 \). Furthermore, applying Lemma 28 we have that for all \( h = j/n^2 \)
\[ b_{h,\delta}(x) \asymp h^\alpha \quad \text{and} \quad s_{h,\delta}(x) \asymp \sqrt{\frac{\log n}{h^d n_0}}. \] (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 \( b_{h,\delta}(x) \leq s_{h,\delta}(x) \), both \( b_{h,\delta}(x), s_{h,\delta}(x) \) 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 \( C_1 n_0^{-1/(2\alpha+d)} \leq \varphi_1(x) \leq C_2 n_0^{-1/(2\alpha+d)} \log n \) 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 \( \rho_{\tau-1}^* := \min_{z \in S_{\tau-1}} \vartheta_{\tau-1}(z) \). By inductive hypothesis, \( \rho_{\tau-1}^* \geq C_1[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} - (\tau-1)/n \). Note also that \( \tilde{\nu}_{\tau-1} \geq \tilde{\nu}_{\tau-2} \) because \( \tilde{S}_{\tau-1} \subseteq \tilde{S}_{\tau-2} \), which holds because \( S_{\tau-1} \subseteq S_{\tau-2} \) and \( \vartheta_{\tau-1}(z) \leq \vartheta_{\tau-2}(z) \) for all \( z \). Let \( h_t^* \) be the smallest number of the form \( j_t^*/n^2 \), \( j_t^* \in [n^2] \) such that \( h_t^* \geq C_1[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} - \tau/n \). We then have \( h_t^* \leq \rho_{\tau-1}^* \) and therefore query points in epoch \( \tau \) are uniformly distributed in \( B_{h_t^*}^\infty(x;G_n) \). Subsequently, applying Lemma 28 we have with probability \( 1 - O(n^{-1}) \) that
\[ b_{h_t^*,\delta}(x) \leq C'[h_t^*]^\alpha \quad \text{and} \quad s_{h_t^*,\delta}(x) \geq C'' \sqrt{\frac{\log n}{h_t^*} \tilde{\nu}_{\tau-1} n_0}, \] (3.83)

where \( C', C'' > 0 \) are constants that depend on \( d, \alpha, M, \rho_0, \tilde{\nu}_0 \) and \( C \), but not \( C_1, C_2, \tau \) or \( h_t^* \). By choosing \( C_1 \) appropriately (depending on \( C' \) and \( C'' \)) we can make \( b_{h_t^*,\delta}(x) \leq s_{h_t^*,\delta}(x) \) holds for all \( x \in S_{\tau-1} \), thus establishing \( \vartheta_{\tau}(x) \geq \min \{ \vartheta_{\tau-1}(x), h_t^* \} \geq C_1[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} - \tau/n \).

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
\[ b_{h,\delta}(x) \asymp \tilde{C}' h^\alpha \quad \text{and} \quad s_{h,\delta}(x) \leq \tilde{C}'' \sqrt{\frac{\log n}{\min \{ h, \rho_{\tau-1}^* \} \tilde{\nu}_{\tau-1} n_0}}, \] (3.84)

where \( \tilde{C}' \) and \( \tilde{C}'' \) are again constants depending on \( d, \alpha, M, \rho_0, \tilde{\nu}_0 \) and \( C \), but not \( C_1, C_2 \). Note also that \( \rho_{\tau-1}^* \geq C_1[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} - (\tau-1)/n \geq C_1[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} - \tau/n \), because \( \tilde{\nu}_{\tau-1} \geq \tilde{\nu}_{\tau-2} \). By selecting constant \( C_2 > 0 \) carefully (depending on \( \tilde{C}', \tilde{C}'' \) and \( C_1 \)), we can ensure \( b_{h,\delta}(x) > s_{h,\delta}(x) \) for all \( h \geq C_2[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} + \tau/n \). Therefore, \( \vartheta_{\tau}(x) \leq h_t(x) \leq C_2[\tilde{\nu}_{\tau-1} n_0]^{-1/(2\alpha+d)} + \tau/n \).
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 $\alpha, d, c_0$ and $C_0$ in $\lesssim, \gtrsim; \asymp, O(\cdot)$ and $\Omega(\cdot)$ notations. However, we do not suppress dependency on $C_\rho$ or $M$ in any of the above notations.

Let $\varphi_0 : [-2, 2]^d \rightarrow \mathbb{R}^*$ be a non-negative function defined on $\mathcal{X}$ such that $\varphi_0 \in \Sigma_0^{[\alpha]}(1)$ with $\kappa = \infty$, $\sup_{x \in \mathcal{X}} \varphi_0(x) = \Omega(1)$ and $\varphi_0(z) = 0$ for all $\|z\|_2 \geq 1$. Here $[\alpha]$ denotes the smallest integer that upper bounds $\alpha$. 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))

$$S_N(x) := \frac{1}{Z} x^{N+1} \sum_{n=0}^{N} \frac{(N+n)}{n} \left(\frac{2N+1}{N-n}\right) (-x)^n, \quad N = 0, 1, 2, \ldots$$

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 $\tilde{S}_{N,d} : [-2, 2]^d \rightarrow \mathbb{R}$ by considering $\tilde{S}_{N,d}(x) := 1/Z - S_N(a\|x\|_1)$ where $\|x\|_1 = |x_1| + \ldots + |x_d|$ and $a = 1/(2d)$. It is easy to verify that, with $Z$ chosen appropriately, $\tilde{S}_{N,d} \in \Sigma_0^{[\alpha]}(1)$, $\sup_{x \in \mathcal{X}} \tilde{S}_{N,d}(x) = 1/Z = \Omega(1)$ and $\tilde{S}_{N,d}(z) = 0$ for all $\|z\|_2 \geq 1$, where $M > 0$ is a constant.

For any $x \in \mathcal{X}$ and $h > 0$, define $\varphi_{x,h} : \mathcal{X} \rightarrow \mathbb{R}^*$ as

$$\varphi_{x,h}(z) := \|z \in B_h^\infty(x)\| \cdot \frac{Mh^\alpha}{2} \varphi_0 \left(\frac{z - x}{h}\right).$$

(3.85)

It is easy to verify that $\varphi_{x,h} \in \Sigma_0^{[\alpha]}(M/2)$, and furthermore $\sup_{x \in \mathcal{X}} \varphi_{x,h}(z) = Mh^\alpha$ and $\varphi_{x,h}(z) = 0$ for all $z \notin B_h^\infty(x)$.

Let $L_f_0(\varepsilon_n(f_0))$ be the level set of $f_0$ at $\varepsilon_n(f_0)$. Let $H_n \subseteq L_f_0(\varepsilon_n(f_0))$ be the largest packing set such that $B_h^\infty(x)$ are disjoint for all $x \in H_n$, and $\bigcup_{x \in H_n} B_h^\infty(x) \subseteq L_f_0(\varepsilon_n(f_0))$. By (A2') and the definition of $\varepsilon_n^d(f_0)$, we have that

$$|H_n| \geq M(L_f_0(\varepsilon_n(f_0)), 2\sqrt{dh}) \geq \mu_{f_0}(\varepsilon_n(f_0)) \cdot h^{-d} \geq [\varepsilon_n^d(f_0)]^{2+d/\alpha} \cdot nh^{-d}. \quad (3.86)$$

For any $x \in H_n$, construct $f_x : \mathcal{X} \rightarrow \mathbb{R}$ as

$$f_x(z) := f_0(z) - \varphi_{x,h}(z).$$

(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 $h = (\varepsilon_n^d(f_0)/M)^{1/\alpha}$ such that $\|\varphi_{x,h}\|_\infty = 2\varepsilon_n^d(f_0)$. We then have that $\|f_x - f_0\|_\infty \leq 2\varepsilon_n^d(f_0)$ and $f_x \in \Sigma_0^{[\alpha]}(M)$, because $f_0, \varphi_{x,h} \in \Sigma_0^{[\alpha]}(M/2)$.

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 $|\mathcal{F}_n| \geq 2$. Let $\mathcal{A}_n = (\chi_1, \ldots, \chi_n, \phi_n)$ be an active optimization algorithm operating with a sample budget $n$, which consists of samplers $\chi_t : \{(x_i, y_i)\}_{i=1}^{t-1} \mapsto x_t$ and an estimator $\phi_n : \{(x_i, y_i)\}_{i=1}^{n} \mapsto f_x \in \mathcal{F}_n$, both can be deterministic or randomized functions. Then

$$\inf_{\mathcal{A}_n} \sup_{f_x \in \mathcal{F}_n} \Pr \left[ f_x \neq f_x \right] \geq \frac{1}{2} - \sqrt{\frac{n \cdot \sup_{f_x \in \mathcal{F}_n} \|f_x - f_0\|_\infty^2}{2|\mathcal{F}_n|}}. \quad (3.88)$$
Lemma 31. There exists constant $M > 0$ depending on $\alpha, d, c_0, C_0$ 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 $z \in \mathcal{X}$, $\max\{\mathcal{L}(z; f_x), \mathcal{L}(z; f_{x'})\} \geq \varepsilon^*_n(f_0)$, 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

$$n_x := \mathbb{E}_{f_0} \left[ \sum_{i=1}^n \mathbb{I}[x \in B^x_h(x)] \right].$$

(3.89)

Because $B^x_h(x)$ are disjoint for $x \in H_n$, we have $\sum_{x \in H_n} n_x \leq n$. Also define, for every $x \in H_n$,

$$\varphi_x := \Pr_{f_0} \left[ \hat{f}_x = f_x \right].$$

(3.90)

Because $\sum_{x \in H_n} \varphi_x = 1$, by pigeonhole principle there is at most one $x \in H_n$ such that $\varphi_x > 1/2$. 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 $\varphi_x \leq 1/2$ and $n_x \leq 2n/|\mathcal{F}_n|$. By Le Cam’s and Pinsker’s inequality (see, for example, (Tsybakov, 2009)) we have that

$$\Pr_{f_x} \left[ \hat{f}_x = f_x \right] \leq \Pr_{f_0} \left[ \hat{f}_x = f_x \right] + d_{TV}(P^A_n \parallel P^A_n)$$

(3.91)

$$\leq \Pr_{f_0} \left[ \hat{f}_x = f_x \right] + \frac{1}{2} KL(P^A_n \parallel P^A_x)$$

(3.92)

$$= \varphi_x + \frac{1}{2} KL(P^A_n \parallel P^A_x)$$

(3.93)

$$\leq 1/2 + \frac{1}{2} KL(P^A_n \parallel P^A_x).$$

(3.94)

It remains to upper bound KL divergence of the active queries made by $A_n$. 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 $\mathcal{X} \setminus B^x_h(x)$, we have

$$KL(P^A_n \parallel P^A_x) = \mathbb{E}_{f_0, A_n} \left[ \frac{\log P_{f_0, A_n}(x_1:n, y_1:n)}{P_{f_x, A_n}(x_1:n, y_1:n)} \right]$$

(3.95)

$$= \mathbb{E}_{f_0, A_n} \left[ \frac{\prod_{i=1}^n P_{f_0}(y_i|x_i)P_{A_n}(x_i|x_{1:(i-1)}, y_{1:(i-1)})}{\prod_{i=1}^n P_{f_0}(y_i|x_i)P_{A_n}(x_i|x_{1:(i-1)}, y_{1:(i-1)})} \right]$$

(3.96)

$$= \mathbb{E}_{f_0, A_n} \left[ \frac{\prod_{i=1}^n P_{f_0}(y_i|x_i)}{\prod_{i=1}^n P_{f_0}(y_i|x_i)} \right]$$

(3.97)

$$= \mathbb{E}_{f_0, A_n} \left[ \sum_{x \in B^x_h(x)} \frac{\log P_{f_0}(y_i|x_i)}{P_{f_x}(y_i|x_i)} \right]$$

(3.98)
where

\[ P \]

By construction, \( \mathcal{H} \) to the largest packing set

\[ M \]

Then

\[ n \cdot \|f_0 - f_x\|_\infty^2. \]  

Therefore,

\[
\Pr_{f_x} \left[ \hat{f}_x = f_x \right] \leq \frac{1}{2} + \sqrt{\frac{1}{4} n_x \varepsilon_n^2 \leq \frac{1}{2} + \sqrt{\frac{n \|f_x - f_0\|^2_\infty}{2|\mathcal{F}_n|}}.}
\]

**Proof of Lemma 31** By construction, \( n \sup_{f_x \in \mathcal{F}_n} \|f_x - f_0\|^2_\infty \leq M^2 n h^{2\alpha} \) and \( |\mathcal{F}_n| = |H_n| \geq [C_\varepsilon \varepsilon_{n}^L(f_0)]^{2+d/\alpha} n h^{-d} \). Note also that \( h \approx (\varepsilon/M)^{1/\alpha} \approx (C_\varepsilon \varepsilon_{n}^L(f_0)/M)^{1/\alpha} \) because \( \|f_x - f_0\|_\infty = \varepsilon = C_\varepsilon \varepsilon_{n}^L(f_0). \) Subsequently,

\[
\frac{n \sup_{f_x \in \mathcal{F}_n} \|f_x - f_0\|^2_\infty}{2|\mathcal{F}_n|} \leq \frac{n [C_\varepsilon \varepsilon_{n}^L(f_0)]^2}{n [C_\varepsilon \varepsilon_{n}^L(f_0)]^2 \cdot M^{d/\alpha}} = M^{-d/\alpha}. \]

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 \( |\mathcal{F}_n| \geq 2 \), and \( \{(x_i, y_i)\}_{i=1}^n \) are i.i.d. random variables. Then

\[
\inf_{f_x} \sup_{f_x \in \mathcal{F}_n} \Pr_{f_x} \left[ \hat{f}_x \neq f_x \right] \geq 1 - \frac{\log 2 + n \cdot \sup_{f_x, f \neq f_x \in \mathcal{F}_n} \KL(P_{f_x} \| P_f)}{\log |\mathcal{F}_n|},
\]

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 \( L_{f_0}(\varepsilon_{n}^L) \) such that \( B^\psi_0(x) \) for all \( x \in H_n \) are disjoint, where \( h \approx (\varepsilon_{n}^L/M)^{1/\alpha} \) such that \( \|\varphi_{x,h}\|_{\infty} = 2\varepsilon_{n}^L \) for all \( x \in H_n \). Because \( f_0 \) satisfies (A2'), we have that \( |\mathcal{F}_n| = |H_n| \geq \mu_{f_0}(\varepsilon_{n}^L) h^{-d} \). Under the condition that \( \varepsilon_{n}^L(f_0) \ll \varepsilon_{n}^L \), it holds that \( \mu_{f_0}(\varepsilon_{n}^L) \geq [\varepsilon_{n}^L]^{2+d/\alpha} n \). Therefore,

\[
|\mathcal{F}_n| \geq [\varepsilon_{n}^L]^{2+d/\alpha} n h^{-d} \approx [\varepsilon_{n}^L]^{2} \cdot n M^{d/\alpha}.
\]

Because \( \log(n/\varepsilon_{n}^L) \approx \log n \) and \( M > 0 \) is a constant, we have that \( \log |\mathcal{F}_n| \geq c \log n \) for all \( n \geq N \), where \( c > 0 \) is a constant depending only on \( \alpha, d \) and \( N \in \mathbb{N} \) is a constant depending on \( M \).

Let \( U \) be the uniform distribution on \( \mathcal{X} \). Because \( x \sim U \) and \( f_x \equiv f_{x'} \) on \( \mathcal{X} \setminus B^\psi_0(x) \), we have that

\[
\KL(P_{f_x} \| P_{f_{x'}}) = \frac{1}{2} \int_{\mathcal{X}} |f_x(z) - f_{x'}(z)|^2 dU(z)
\]

76
\[ \leq \frac{1}{2} \Pr_{U}[ z \in B_{h}^{\alpha} (x)] \cdot \| f_x - f_{x'} \|_{\infty}^{2} \]  
\[ \leq \frac{1}{2} \lambda (B_{h}^{\alpha} (x)) \cdot [\varepsilon_{n}^{1}]^{2} \]  
\[ \leq h^{d}[\varepsilon_{n}^{1}]^{2} \leq [\varepsilon_{n}^{1}]^{2+d/\alpha} / M^{d/\alpha}. \]

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 \( \ell_{\infty} \) 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 \( x_{i} \) and \( z_{i} \))

\[ \left\| \sum_{i=1}^{n} \varepsilon_{i} z_{i} \right\|_{\infty} \leq \sigma \frac{\log d}{m} + H \delta. \]

**Proof.** Let \( \xi_{i} = \varepsilon_{i} / \delta \sim N(0, \sigma^{2}/\delta^{2}) \). Consider the following decomposition:

\[ \left\| \sum_{i=1}^{n} \varepsilon_{i} z_{i} \right\|_{\infty} \leq \frac{1}{m\delta} \left\| \sum_{i=1}^{m} \xi_{i} z_{i} \right\|_{\infty} + \delta \cdot \sup_{1 \leq i \leq m} \left| \xi_{i} \varepsilon_{i} H_{t}(\kappa_{i}, z_{i}) \right| \cdot \| z_{i} \|_{\infty}. \]

The second term on the right-hand side of the above inequality is upper bounded by \( O(H \delta) \) almost surely, because \( \| z_{i} \|_{\infty} \leq 1 \) and \( |z_{i}^{T} H_{t}(\kappa_{i}, z_{i})| \leq \| H_{t}(\kappa_{i}, z_{i}) \|_{1} \| z_{i} \|_{2}^{2} \leq H \). For the first term, because \( \xi_{i} \) are centered sub-Gaussian random variables independent of \( z_{i} \) and \( \| z_{i} \|_{\infty} \leq 1 \), we have that \( 1/m \cdot \left\| \sum_{i=1}^{m} \xi_{i} z_{i} \right\|_{\infty} \leq \sqrt{\sigma^{2} \log d/m} \) with probability \( 1 - O(d^{-3}) \), by invoking standard sub-Gaussian concentration inequalities. \( \square \)

Now define \( \hat{\theta} = (g_{t}, \hat{\mu}_{t}), \theta_{0} = (g_{t}, \delta^{-1} f(x_{t})) \) and \( \tilde{Z} = (z_{1}, \ldots, z_{m}) \) where \( z_{i} = (z_{i}, 1) \in \mathbb{R}^{d+1} \). Define also that \( Y = (\tilde{y}_{1}, \ldots, \tilde{y}_{m}) \). The estimator can then be written as \( \hat{\theta} = \arg \min_{\theta \in \mathbb{R}^{d+1}} \frac{1}{m} \| \tilde{Y} - \tilde{Z} \theta \|_{2}^{2} + \lambda \| \theta \|_{1} \). We first establish a “basic inequality” type results that are essential in performance analysis of Lasso type estimators. By optimality of \( \hat{\theta} \), we have that

\[ \frac{1}{m} \| Y - \tilde{Z} \theta \|_{2}^{2} + \lambda \| \theta \|_{1} \leq \frac{1}{m} \| Y - \tilde{Z} \theta_{0} \|_{2}^{2} + \lambda \| \theta_{0} \|_{1} = \frac{1}{m} \varepsilon \| z \|_{2}^{2} + \lambda \| \theta_{0} \|_{1}. \]

Re-organizing terms we obtain

\[ \lambda \| \hat{\theta} \|_{1} \leq \lambda \| \theta_{0} \|_{1} + \frac{2}{m} (\hat{\theta} - \theta_{0})^{T} \tilde{Z}^{T} \varepsilon. \]
On the other hand, by Hölder’s inequality and Lemma 33 we have, with probability $1 - O(d^{-2})$,

$$\frac{2}{m}(\hat{\theta} - \theta_0)^\top Z^\top \varepsilon \leq 2\|\hat{\theta} - \theta_0\|_1 \cdot \left\| \frac{1}{m} Z^\top \varepsilon \right\|_\infty \leq \|\hat{\theta} - \theta_0\|_1 \cdot \left( \frac{\sigma}{\delta} \sqrt{\frac{\log d}{m}} + H\delta \right).$$

Subsequently, if $\lambda \leq c_0(\sigma\delta^{-1}\sqrt{\log d/m} + H\delta)$ for some sufficiently small $c_0 > 0$, we have that $\|\hat{\theta}\|_1 \leq \|\theta_0\|_1 + 1/2\|\hat{\theta} - \theta_0\|_1$. Multiplying by 2 and adding $\|\hat{\theta} - \theta_0\|_1$ on both sides of the inequality we obtain $\|\hat{\theta} - \theta_0\|_1 \leq 2(\|\hat{\theta} - \theta_0\|_1 + \|\theta_0\|_1 - \|\hat{\theta}\|_1)$. Recall that $\theta_0$ is sparse and let $\mathcal{S} = S \cup \{d + 1\}$ be the support of $\theta_0$. We then have $\|\hat{\theta} - \theta_0\|_\mathcal{S} + \|\hat{\theta}_0\|_\mathcal{S} - \|\hat{\theta}_0\|_\mathcal{S} = 0$ and hence $\|\hat{\theta} - \theta_0\|_\mathcal{S} \leq 2\|\hat{\theta} - \theta_0\|_1$. Thus,

$$\|\hat{\theta} - \theta_0\|_\mathcal{S} \leq 3\|\hat{\theta} - \theta_0\|_1. \quad (3.109)$$

Now consider $\hat{\theta}$ that minimizes $\frac{1}{m}\|Y - \overline{Z}\theta\|_2^2 + \lambda\|\theta\|_1$. By KKT condition we have that

$$\left\| \frac{1}{m} Z^\top (Y - \overline{Z}\hat{\theta}) \right\|_\infty \leq \frac{\lambda}{2}.$$ 

Define $\hat{\Sigma} = \frac{1}{m}Z^\top Z$ and recall that $Y = \overline{Z}\theta_0 + \varepsilon$. Invoking Lemma 33 and the scaling of $\lambda$ we have that, with probability $1 - O(d^{-2})$

$$\|\hat{\Sigma}(\hat{\theta} - \theta_0)\|_\infty \leq \frac{\lambda}{2} + \left\| \frac{1}{m} Z^\top \varepsilon \right\|_\infty \leq \frac{\sigma}{\delta} \sqrt{\frac{\log d}{m}} + \delta H. \quad (3.110)$$

By definition of $\{\mathbb{E}_i\}_{i=1}^m$, we know that $\hat{\Sigma}_{i,j} = 1$ for all $j = 1, \ldots, d + 1$ and $\mathbb{E}[\hat{\Sigma}_{j,k}] = 0$ for $j \neq k$. By Hoeffding’s inequality (Hoeffding, 1963) and union bound we have that with probability $1 - O(d^{-2})$, $\|\hat{\Sigma} - I_{(d+1)\times(d+1)}\|_\infty \leq \sqrt{\log d/m}$, where $\|\cdot\|_\infty$ denotes the maximum absolute value of matrix entries. Also note that $\hat{\theta} - \theta_0$ satisfies $\|\hat{\theta} - \theta_0\|_\mathcal{S} \leq 3\|\hat{\theta} - \theta_0\|_1$ thanks to Eq. (3.109). Subsequently,

$$\|\hat{\theta} - \theta_0\|_\infty \leq \|\hat{\Sigma}(\hat{\theta} - \theta_0)\|_\infty + \|\hat{\Sigma} - I\|_\infty \|\hat{\theta} - \theta_0\|_1 \leq \|\hat{\Sigma}(\hat{\theta} - \theta_0)\|_\infty + \|\hat{\Sigma} - I\|_\infty \cdot 4\|\hat{\theta} - \theta_0\|_\mathcal{S} \leq \|\hat{\Sigma}(\hat{\theta} - \theta_0)\|_\infty + \|\hat{\Sigma} - I\|_\infty \cdot 4(s + 1)\|\hat{\theta} - \theta_0\|_1 \leq \frac{\sigma}{\delta} \sqrt{\frac{\log d}{m}} + \delta H + \left(\frac{s^2 \log d}{m}\right)\|\hat{\theta} - \theta_0\|_\infty. \quad (3.111)$$

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 $\ell_1$ error $\|\hat{\theta} - \theta_0\|_1$ 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 $\nu_1^2$ and $\nu_2^2$, respectively. Then $XY$ is a centered sub-exponential random variable with parameter $\nu = \sqrt{2v}$ and $\alpha = 2v$, where $v = 2e^{2/\nu_1^2}\nu_1\nu_2$.

**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 $k \geq 3$ (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 $\nu_1^2$, it holds that $(\mathbb{E}[X^k])^{1/k} \leq \nu_1 e^{1/k} \sqrt{k}$. Similarly, $(\mathbb{E}[X^k])^{1/k} \leq \nu_2 e^{1/k} \sqrt{k}$. Subsequently,

$$\mathbb{E}[|XY|^k] \leq (e^{2/\nu_1^2} \nu_1 \nu_2)^k \cdot k^k \leq \left( e^{2/\nu_1^2} \nu_1 \nu_2 + e^{\nu_2^2} \right)^k \cdot k^k \leq \frac{1}{2} k! \cdot \left( 2e^{2/\nu_1^2} \nu_1 \nu_2 \right)^k.$$

where in the second inequality we use the Stirling’s approximation inequality that $\sqrt{2\pi} k^ke^{-k} \leq k!$. The sub-exponential parameter of $XY$ can then be determined. \(\square\)

We use the “full-length” parameterization $\tilde{\theta}_t = \hat{\theta}_t + \frac{1}{m} Z_t^\top (\tilde{Y}_t - Z_t \hat{\theta}_t)$, where $Z_t$, $\tilde{Z}_t$ and $\tilde{Y}_t$ are notations defined in the proof of Lemma 13 (with subscripts $t$ added to emphasize that both $Z_t$ and $\tilde{Y}_t$ are specific to the $t$th epoch in Algorithm 5). Because $\tilde{Y}_t = Z_t \theta_0 + \varepsilon_t$ (where $\theta_0 \sim \mathcal{N}(0, \Sigma)$ defined in Eq. (3.22)), we have

$$\tilde{\theta}_t = \hat{\theta}_t + \frac{1}{m} Z_t^\top (\tilde{Z}_t \theta_0 + \varepsilon_t - \tilde{Z}_t \hat{\theta}_t) = \theta_0 + \frac{1}{m} Z_t^\top \varepsilon_t + (\tilde{\Sigma} - \Sigma)(\hat{\theta}_t - \theta_0),$$

where $\tilde{\Sigma} = \frac{1}{m} Z_t^\top Z_t$. 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 $\tilde{\theta}_t$ are identical to $\tilde{\theta}_t$ defined in Eq. (3.24). Subsequently,

$$\tilde{g}_t = g_t + \frac{1}{m} Z_i^\top \xi_i + \frac{\delta}{m} Z_i^\top b + \left( \tilde{\Sigma} - \Sigma \right)(\hat{\theta}_t - \theta_0) \mid_{i \leq d}. \quad (3.112)$$

In Eq. (3.112) we divide $\tilde{g}_t - g_t$ into two terms. We first consider the term $\zeta_t := \frac{1}{m} Z_i^\top \xi_i$. It is clear that $\mathbb{E}[\zeta_t | x_t] = 0$ because $\mathbb{E}[\xi_i | x_t, Z_t] = 0$. 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 $i \in [m]$, $z_i^\top a$ are i.i.d. sub-Gaussian random variables with common parameter $\nu^2 = \|a\|_2^2$. Also, $\tilde{\xi}_i$ is a sub-Gaussian random variable with parameter $\sigma^2$ and is independent of $z_i^\top a$. Thus, invoking Lemma 34 we have that $\xi_i z_i^\top a$ is a sub-exponential random variable with parameters $\nu = \alpha / \sqrt{2} \leq \sigma \|a\|_2$. Consequently, $\langle \zeta_t, a \rangle \sim \mathcal{N}(\gamma_t \alpha^2 / \sqrt{m})$, where $\gamma_t \alpha^2 = \sum_{i=1}^m \frac{\xi_i z_i^\top a}{\|a\|_2^2 \sqrt{m}}$. Also, $\gamma_t \alpha^2 = \delta \|a\|_2^2 \cdot \alpha \leq \mathbb{E}[\sum_{i=1}^m \xi_i z_i^\top a] / \|a\|_2^2 \sqrt{m}$. We next consider the term $\gamma_t = \frac{\delta}{m} Z_i^\top b + (\tilde{\Sigma} - \Sigma)(\hat{\theta}_t - \theta_0)$. By Assumption (A3) we know that $\|b\|_\infty \leq \delta H$. Consequently, by H"older’s inequality we have that

$$\|\gamma_t\|_\infty \leq \frac{\delta}{m} \|Z_t\|_{1,\infty} \|b\|_\infty + \|\tilde{\Sigma} - \Sigma\|_\infty \|\hat{\theta}_t - \theta_0\|_1$$

79
\[
H \delta + \sqrt{\frac{\log d}{m}} \left( \frac{\sigma s}{\delta} \sqrt{\frac{\log d}{m}} + s \delta H \right).
\]

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 \(\kappa\)-strongly convex with respect to \(\| \cdot \|_\psi\). Suppose \(f\) is \(\tilde{H}\)-smooth with respect to \(\| \cdot \|_\psi\), meaning that \(f(y) \leq f(x) + \nabla f(x)^T (y-x) + \frac{\tilde{H}}{2} \|y-x\|_\psi^2\) for all \(x, y \in \mathcal{X}\), and \(\eta < \kappa/\tilde{H}\). Define \(g_t = \nabla f(x_t)\), and let \(x_0, \ldots, x_{T'-1}\) be iterations in Algorithm 5. Then for every \(0 \leq t \leq T' - 1\) and any \(x^* \in \tilde{X}\),

\[
\eta [f(x_{t+1}) - f(x^*)] + \Delta_{\psi}(x_{t+1}, x^*) \leq \Delta_{\psi}(x_t, x^*) + \eta \langle \tilde{g}_t - g_t, x^* - x_t \rangle + \frac{\eta^2 \| \tilde{g}_t - g_t \|_{\psi^*}^2}{2(\kappa - \tilde{H} \eta)}. \tag{3.113}
\]

Adding both sides of Eq. (3.113) from \(t = 0\) to \(t = T' - 1\), telescoping and noting that \(\Delta_{\psi}(x_{T'}, x^*) \geq 0\), we obtain

\[
\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*) \leq \frac{\Delta_{\psi}(x_0, x^*)}{\eta T'} + \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \tilde{g}_t - g_t, x_t - x^* \rangle + \frac{\eta}{2(\kappa - \tilde{H} \eta)} \cdot \sup_{0 \leq t < T'} \| \tilde{g}_t - g_t \|_{\psi^*}^2. \tag{3.114}
\]

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

\[
\begin{align*}
    f(y) &\geq f(x) + \nabla f(x)^T (y-x) + \tilde{H} \|y-x\|_\psi^2 \\
    &\geq f(x) + \nabla f(x)^T (y-x) + \tilde{H} \|y-x\|_\psi^2
\end{align*}
\]

for all \(x, y \in \mathcal{X}\) with \(\tilde{H} \leq \epsilon H\), because \(\|x-y\|_\psi^2 \leq d^{1-1/\alpha} \|x-y\|_a^2 \leq d^{1/\log d} \|x-y\|_1^2 = \epsilon \|x-y\|_1^2\) by Hölder’s inequality. In addition, by definition of Bregman divergence we have that

\[
\Delta_{\psi}(x_0, x^*) \leq \frac{1}{2(a-1)} \|x^*\|_a^2 \leq \frac{1}{2(a-1)} \|x^*\|_1^2 \leq \|x^*\|_1^2 \log d \leq B^2 \log d, \tag{3.115}
\]

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 \(\frac{1}{T'} \sum_{t=0}^{T'-1} \langle \tilde{g}_t - g_t, x^* - x_t \rangle\) and \(\| \tilde{g}_t - g_t \|_{\psi^*}^2\) terms. By Lemma 14 and sub-exponential concentration inequalities (e.g., Lemma 91), we have that with probability \(1 - O(d^{-1})\)

\[
\| \tilde{g}_t - g_t \|_\infty \leq \| g_t \|_\infty + \| \gamma_t \|_\infty \lesssim \frac{\sigma}{\delta} \left( \sqrt{\frac{\log d}{m}} + \log d \right) + H \delta + \frac{\sigma s \log d}{\delta m} \lesssim \frac{\sigma}{\delta} \sqrt{\frac{\log d}{m}} + H \delta
\]

80
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

\[
\sup_{0 \leq t < T'} \|\tilde{g}_t - g_t\|_q^2 \leq d^{2(\alpha-1)/\alpha} \cdot \sup_{0 \leq t < T'} \|\tilde{g}_t - g_t\|_\infty^2 \lesssim \frac{\sigma^2 \log d}{\delta^2 m} + H^2 \delta^2.
\]  (3.116)

We now consider the first term \( \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \zeta_t, x^* - x_t \rangle \approx \frac{1}{T'} \sum_{t=0}^{T'-1} X_t \sup_{0 \leq t < T'-1} \|\gamma_t\|_\infty \|x^* - x_t\|_1 \), 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 \( \nu = \sqrt{m/2} \cdot \alpha \leq \sigma \|x^* - x_t\|_2/\delta \sqrt{n} \lesssim \sigma \|x^*\|_1/\delta \sqrt{m} \).

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 \( 1 - O(d^{-1}) \)

\[
\left| \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \zeta_t, x^* - x_t \rangle \right| \leq \frac{\sigma \|x^*\|_1}{\delta} \left( \sqrt{\frac{\log d}{n}} + \frac{\log d}{n} \right) \lesssim \frac{\sigma \|x^*\|_1}{\delta} \sqrt{\frac{\log d}{n}},
\]

where the last inequality holds because \( n \geq m = \Omega(s^2 \log d) \). Thus,

\[
\left| \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \tilde{g}_t - g_t, x^* - x_t \rangle \right| \lesssim \frac{\sigma \|x^*\|_1}{\delta} \sqrt{\frac{\log d}{n}} + \|x^*\|_1 \left( H \delta + \frac{\sigma s \log d}{\delta m} \right).
\]  (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 \( \|x^*\|_1 \leq B \), we obtain

\[
\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*) \lesssim \frac{1}{\eta} \frac{\|x^*\|_1^2 \log d m}{n} + \frac{\sigma \|x^*\|_1}{\delta} \sqrt{\frac{\log d}{n}} + \|x^*\|_1 \left( H \delta + \frac{\sigma s \log d}{\delta m} \right) + \eta \left( \frac{\sigma^2 \log d}{\delta^2 m} + H^2 \delta^2 \right)
\]

\[
\lesssim \frac{B^2 \log d m}{\eta} + \frac{\sigma B}{\delta} \sqrt{\frac{\log d}{n}} + B \left( H \delta + \frac{\sigma s \log d}{\delta m} \right) + \eta \left( \frac{\sigma^2 \log d}{\delta^2 m} + H^2 \delta^2 \right)
\]  (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 \( \eta \) and \( m \), it is easy to verify that \( \eta < 1/2H, m = \Omega(s^2 \log d) \) and \( m = O(n) \). Subsequently,

\[
\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*) \lesssim B \sqrt{\frac{m \log d}{n}} + \sigma B \sqrt{\frac{m}{sn}} + B(\sigma + H) \sqrt{\frac{s \log d}{m}} + B \sqrt{\frac{m \log d}{n}} \left( \frac{\sigma^2}{s} + \tilde{O}(m^{-1}) \right)
\]

\[
\lesssim B \left( \frac{(1 + H)^2 s \log^2 d}{n} \right)^{1/4} + \sigma B \sqrt{\frac{(1 + H)}{s^{1/4} n^{1/4}}} + B(\sigma + H) \left( \frac{s \log^2 d}{n} \right)^{1/4}
\]

81
+ B \left( \frac{(1 + H)^2 s \log d}{n} \right)^{1/4} \left( \frac{\sigma^2}{s} + \tilde{O}(n^{-1/2}) \right) \\
\leq \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} + \tilde{O}(n^{-1/2}) \\
\leq (1 + \sigma + \sigma^2/s) B \sqrt{\log d} \left[ \frac{(1 + H)^2 s}{n} \right]^{1/4} + \tilde{O}(n^{-1/2}).

### 3.6.4 Proof of Lemma 15

Using the model Eq. (3.22) we can decompose \( \bar{g}_t(\delta) - g_t \) as

\[
\bar{g}_t(\delta) - g_t = \frac{\delta}{2} \mathbb{E} \left[ (z^\top H_t z) z \right] + \frac{1}{n\delta} \sum_{i=1}^n (z_i^\top H_t z_i) z_i - \mathbb{E} \left[ (z^\top H_t z) z \right] := \tilde{\zeta}_t(\delta) + \beta_t(\delta) + \frac{\delta}{2n} \sum_{i=1}^n (z_i^\top (H_t(\delta z_i) - H_t) z_i) z_i + \left[ (\tilde{\Sigma} - I) (\tilde{\theta}_t - \theta_{0t}) \right] \mathbf{1}_d, \quad := \gamma_t(\delta),
\]

where \( \tilde{\Sigma}, \tilde{\theta}_t \) and \( \theta_{0t} \) are similarly defined as in the proof of Lemma 14. The sub-exponentiality of \( \langle \tilde{\zeta}_t(\delta), a \rangle \) for any \( a \in \mathbb{R}^d \) is established in Lemma 14. We next consider \( \bar{\beta}_t(\delta) \). For any \( a \in \mathbb{R}^d \) consider \( \langle \bar{\beta}_t(\delta), a \rangle = \frac{\delta}{2n} \sum_{i=1}^n X_i(a) \) where \( X_i(a) = (z_i^\top H_t z_i) (z_i^\top a) - \mathbb{E} \left[ (z_i^\top H_t z_i) (z_i^\top a) \right] \) are centered i.i.d. random variables conditioned on \( H_t \) and \( x_t \). In addition, \( |X_i(a)| \leq 2 \| H_t \|_1 \| z_i \|^2_\infty \cdot \| a \|_1 \| z_i \|_\infty \lesssim H \| a \|_1 \) almost surely. Therefore, \( X_i(a) \) is a sub-Gaussian random variable with parameter \( \nu = H \| a \|_1 \), and hence \( \langle \bar{\beta}_t(\delta), a \rangle \) is a sub-Gaussian random variable with parameter \( \nu = \delta H \| a \|_1 / \sqrt{n} \). Finally, for the deterministic term \( \gamma_t(\delta) \), we have that

\[
\| \gamma_t(\delta) \|_\infty \leq \frac{\delta}{2} \sup_{z \in \{ \pm 1 \}^d} \| H_t(\delta z) - H_t \|_1 \| z \|_\infty^2 + \| (\tilde{\Sigma} - I) (\tilde{\theta}_t - \theta_{0t}) \|_\infty \\
\leq \frac{\delta}{2} \sup_{z \in \{ \pm 1 \}^d} L \cdot \| \tilde{z} \|_\infty \| z \|_\infty^2 + \| \tilde{\Sigma} - I \|_{\max} \| \tilde{\theta}_t - \theta_{0t} \|_\infty \\
\leq L \delta^2 + \sqrt{\log d} \left( \frac{\sigma s \log d}{n} + s \delta H \right) \\
\lesssim L \delta^2 + \frac{\sigma s \log d}{n\delta} + s \delta H \sqrt{\log d}.
\]

### 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) \( \| \bar{g}_t - g_t \|_W^2 \) and \( \frac{1}{T} \sum_{t=0}^{T-1} \langle \bar{g}_t - g_t, x^* - x_t \rangle \).
For the first term, using sub-exponentiality of \( \zeta_t \) and sub-gaussianity of \( \beta_t \), we have with probability \( 1 - O(d^{-1}) \) uniformly over all \( t \in \{0, \ldots, T' - 1\} \),

\[
\|g_{t,w} - g_t\|_\infty \leq \|\zeta_t\|_\infty + \|\beta_t\|_\infty + \|\gamma_t\|_\infty \\
\lesssim \frac{\sigma}{\delta} \left( \frac{\log d}{m} + \frac{\log d}{m} \right) + \delta H \sqrt{\frac{\log d}{m}} + L\delta^2 + H\delta \sqrt{\frac{s^2 \log d}{m}} + \frac{\sigma s \log d}{\delta m} \\
\lesssim \left( \frac{\sigma}{\delta} + s\delta H \right) \sqrt{\frac{\log d}{m}} + L\delta^2,
\]

where the last inequality holds because \( m = \Omega(s^2 \log d) \). Subsequently, with probability \( 1 - O(d^{-1}) \)

\[
\sup_{0 \leq t \leq T' - 1} \|\zeta_{t,w} - g_t\|^2_{\psi} \lesssim \left( \frac{\sigma^2}{\delta^2} + s^2 \delta^2 H^2 \right) \frac{\log d}{m} + L^2 \delta^4.
\]

For the other term \( \frac{1}{T'} \sum_{t=0}^{T'-1} \langle g_{t,w} - g_t, x^* - x_t \rangle \), again using concentration inequalities of sub-exponential/sub-Gaussian martingales and noting that \( \|x^* - x_t\|_2 \leq \|x^* - x_t\|_1 \leq 2B \), we have

\[
\frac{1}{T'} \sum_{t=0}^{T'-1} \langle g_{t,w} - g_t, x^* - x_t \rangle = \frac{1}{T'} \sum_{t=0}^{T'-1} \langle \zeta_t + \beta_t + \gamma_t, x^* - x_t \rangle \\
\lesssim \left( \frac{\sigma}{\delta} + s\delta H \right) B \sqrt{\frac{\log d}{n}} + B \left( L\delta^2 + \frac{\sigma s \log d}{\delta m} + s\delta H \sqrt{\frac{\log d}{m}} \right).
\]

Subsequently, combining Eqs. (3.119,3.120) with Eq. (3.114) we have

\[
\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*) \\
\lesssim \frac{B^2 \log d}{\eta} \frac{m}{n} + \left( \frac{\sigma}{\delta} + s\delta H \right) B \sqrt{\frac{\log d}{n}} + (B + \eta) \left( L\delta^2 + \frac{\sigma s \log d}{\delta m} + s\delta H \sqrt{\frac{\log d}{m}} \right) \\
+ \eta \left( \frac{\sigma^2}{\delta^2} + s^2 \delta^2 H^2 \right) \frac{\log d}{m} + \eta L^2 \delta^4.
\]

We are now ready to prove Theorem 8. It is easy to verify that with the condition imposed on \( n \) and the selection of \( \eta \) and \( m \), it holds that \( \eta < 1/2H \), \( m = \Omega(s^2 \log d) \) and \( m \leq n/10 \). Subsequently,

\[
\frac{1}{T'} \sum_{t=0}^{T'-1} f(x_t) - f(x^*) \\
\lesssim Bm^{1/3} \sqrt{\frac{\log d}{n}} + \left[ \sigma \left( \frac{m}{s \log d} \right)^{1/3} + \tilde{O}(m^{-1/3}) \right] B \sqrt{\frac{\log d}{n}}
\]

83
\[ + \left( B + \tilde{O} \left( \frac{m^{2/3}}{\sqrt{n}} \right) \right) \left[ \left( L + \sigma \right) \left( \frac{s \log d}{m} \right)^{2/3} + \tilde{O}(m^{-5/6}) \right] \]

\[ + Bm^{2/3} \sqrt{\frac{\log d}{n}} \left( \sigma^2 \left( \frac{m}{s \log d} \right)^{2/3} + \tilde{O}(m^{-2/3}) \right) \frac{\log d}{m} \]

\[ + Bm^{2/3} \sqrt{\frac{\log d}{n}} L^2 \left( \frac{s \log d}{m} \right) 4/3 \]

\[ \preceq Bm^{1/3} \sqrt{\frac{\log d}{n}} + \sigma B \left( \frac{m}{s \log d} \right)^{1/3} \sqrt{\frac{\log d}{n}} + B(L + \sigma) \left( \frac{s \log d}{m} \right)^{2/3} \]

\[ + \sigma^2 B \left( \frac{m}{s^2 \log^2 d} \right)^{1/3} \sqrt{\frac{\log d}{n}} + \tilde{O}(n^{-5/12}) \]

\[ \preceq \left( B \sqrt{\frac{\log d}{n}} + \frac{\sigma B \sqrt{\log d}}{s^{1/3}} + \frac{\sigma^2 B \sqrt{\log d}}{s^{2/3}} \right) \left[ \frac{(1 + L)s^{2/3}}{n} \right]^{1/3} \]

\[ + \frac{B(L + \sigma)}{(1 + L)^{2/3}} \left( \frac{s^{2/3} \log d}{n} \right)^{1/3} + \tilde{O}(n^{-5/12}) \]

\[ \preceq \left( B \sqrt{\frac{\log d}{n}} + \frac{\sigma B \sqrt{\log d}}{s^{1/3}} + \frac{\sigma^2 B \sqrt{\log d}}{s^{2/3}} \right) \left[ \frac{(1 + L)s^{2/3}}{n} \right]^{1/3} \]

\[ + B \sigma \sqrt{\log d} \left( \frac{(1 + L)s^{2/3}}{n} \right)^{1/3} + \tilde{O}(n^{-5/12}) \]

\[ \preceq (1 + \sigma + \sigma^2 / s^{2/3}) B \sqrt{\frac{\log d}{n}} \left[ \frac{(1 + L)s^{2/3}}{n} \right]^{1/3} + \tilde{O}(n^{-5/12}). \]

### 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 \( \int p(u) du = 1 \), \( \mathbb{E}_u [\nabla \log p(u)] = 0 \) 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

\[ |f(x + u) - f(x)| = \langle \nabla f(x + \delta(u) u), u \rangle. \tag{3.122} \]

By definition of dual norms (also known as the Cauchy-Schwarz inequality), for any vectors \( a, b \) and pairs of dual norms \( \| \cdot \|_p, \| \cdot \|_q \) it holds that \( |\langle a, b \rangle| \leq \| a \|_p \| b \|_q \). Subsequently

\[ |f(x + u) - f(x)| \leq \| \nabla f(x + \delta(u) u) \|_q \| u \|_p. \tag{3.123} \]

84
Lemma 36. Suppose \( p \) using a union bound over all \( j \) where the last inequality holds because 

\[
\frac{\delta p^{1-1/p} f}{\Gamma(1/p)} \exp \{-|u|^p/p \delta^p\} \text{ for some } p \in (1, 2]. \]

Then for any \( \beta, r \geq 1 \), we have

\[
\mathbb{E} \|u\|_\beta^r \leq \delta^r d^{r/\beta} \log d \left[ 1 + \frac{\delta p^{1-1/p} r!}{\Gamma(1/p)} \right].
\]

Proof. For every \( t > 0 \),

\[
\Pr[|u_1| \geq t] = \frac{p^{1-1/p}}{\delta \Gamma(1/p)} \int_t^{\infty} e^{-z^p/p \delta^p} dx = \frac{p^{1-1/p}}{\Gamma(1/p)} \int_{t/\delta}^{\infty} e^{-z^p/p} dz
\]

\[
\leq \frac{p^{1-1/p}}{\Gamma(1/p)} \int_{t/\delta}^{\infty} e^{-z/2} dz \quad \forall t/\delta \geq 1,
\]

where the last inequality holds because \( e^{-z^p/p} \leq e^{-z/2} \) for all \( z \geq 1 \) and \( p \in (1, 2] \). Subsequently, using a union bound over all \( j \in \{1, \ldots, d\} \), we have

\[
\Pr[\|u\|_\infty \geq t] \leq \frac{2p^{1-1/p}}{\Gamma(1/p)} \cdot de^{-t/2}. \quad \forall t \geq \delta.
\]

For any \( \beta, r \geq 1 \), we have

\[
\mathbb{E}_{\mu} \|u\|_\beta^r = \mathbb{E}_{\mu} \left( \sum_{i=1}^{d} |u_i|^{r/\beta}\right)^{r/\beta} \leq \mathbb{E}_{\mu} d^{r/\beta} \|u\|_\infty^r
\]

\[
\leq d^{r/\beta} \left[ (2\delta \log d)^r + \frac{2dp^{1-1/p}}{\Gamma(1/p)} \int_{2\delta \log d}^{\infty} t^r e^{-t/2} dt \right]
\]

\[
= d^{r/\beta} \left[ (2\delta \log d)^r + \delta^{r+1} \cdot \frac{2dp^{1-1/p}}{\Gamma(1/p)} \int_{2\log d}^{\infty} z^r e^{-z/2} dz \right].
\]

Applying integration by parts for \( r \) times, we have for all \( r \in \mathbb{N} \) that

\[
\int_{a}^{\infty} z^r e^{-z/2} dz = 2a^r e^{-a} + 2r \int_{a}^{\infty} z^{r-1} e^{-z/2} dz = \ldots
\]

\[
= \sum_{\ell=0}^{r-1} \frac{r!}{(r-\ell-1)!} a^{\ell+1} e^{-a} + r! \int_{a}^{\infty} e^{-z/2} dz
\]

\[
\leq 2^r \cdot \sum_{\ell=0}^{r} a^\ell e^{-a} \leq r \cdot (2a)^r \cdot e^{-a}.
\]

85
Subsequently,
\[
\mathbb{E}_\mu \|u\|_\beta^p \lesssim d^{r/\beta} \left[ (\delta \log d)^r + \delta^{r+1} \cdot \frac{dp^{1-1/p}}{\Gamma(1/p)} \cdot r!(\log d)^r \cdot \frac{1}{d} \right]
\lesssim \delta^r d^{r/\beta} \log^r d + \delta^{r+1} d^{r/\beta} \log^r d \cdot \frac{p^{1-1/p}r!}{\Gamma(1/p)}
= \delta^r d^{r/\beta} \log^r d \left[ 1 + \frac{\delta p^{1-1/p}r!}{\Gamma(1/p)} \right].
\]
(3.131)

3.6.8 Proof of Lemma 18

By definition of \( \tilde{g}_t \) and the conditional independence between \( y_t, y'_t \) and \( u_t \), we have that
\[
\mathbb{E}_\mu \| \tilde{g}_t \|^2_q = \mathbb{E}_\mu (y_t - y'_t)^2 \| \nabla \log p(u_t) \|^2_q
= \mathbb{E}_\mu (f(x_t) - f(x_t + u_t) + \xi - \xi')^2 \| \nabla \log p(u_t) \|^2_q
= \mathbb{E}_\mu (f(x_t) - f(x_t + u_t))^2 + 2\sigma^2 \| \nabla \log p(u_t) \|^2_q.
\]
(3.132)
(3.133)
(3.134)

Here the last identity holds because \( \xi, \xi' \) are independent of \( f(x_t), f(x_t + u_t), \) and therefore
\[
\mathbb{E} f(x_t) - f(x_t + u_t) + \xi - \xi'|^2 = \mathbb{E} f(x_t) - f(x_t + u_t)|^2 + 2\mathbb{E}(\xi - \xi') (f(x_t) - f(x_t + u_t)) + \mathbb{E}|\xi - \xi'|^2 = \mathbb{E} f(x_t) - f(x_t + u_t)|^2 + 2\sigma^2, \text{ because } \xi, \xi' \text{ are independent } \mathcal{N}(0, \sigma^2) \text{ random variables.}
\]

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 \( \| \cdot \|_p, \| \cdot \|_q \) and the uniform boundedness of \( \| \nabla f(z) \|_q \), we have
\[
|f(x_t + u_t) - f(x_t)| \leq \| \nabla f(x_t + \delta(u_t)u_t) \|_q \| u_t \|_p \leq L \| u_t \|_p^2.
\]
(3.135)

We next prove the second inequality. Note that
\[
\mathbb{E}_\mu [\| u \|_p^2 \| \nabla \log p(u) \|_q^2] = \mathbb{E}_\mu [\| u \|_p^2 \| \tilde{u} \|_q^2] / \delta^{2/p},
\]
where \( |\tilde{u}| = |u|^{p-1} \). Using Cauchy-Schwarz inequality, we have that
\[
\mathbb{E}_\mu [\| u \|_p^2 \| \nabla \log p(u) \|_q^2] = \mathbb{E}_\mu [\| u \|_p^2 \| \tilde{u} \|_q^2] / \delta^{2/p} \leq \sqrt{\mathbb{E}[u]^4} \cdot \sqrt{\mathbb{E}[u_q^{4(p-1)}]} / \delta^p.
\]

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 \( P_x(y) := \arg \min_{x} \{ \langle y, z - x \rangle + D_\psi(z, x) \} \) as the prox-minimization. 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
\[
D_\psi(u, P_x(y)) \leq D_\psi(u, x) + \langle y, u - x \rangle + \frac{\|y\|_q^2}{2}.
\] (3.136)

The mirror descent update is given by \( x_{t+1} = \arg\min_{x \in \mathcal{X}} \{ \eta \langle x, \tilde{g}_t(x) \rangle + D_\psi(x, x_t) \} \). Using the language of prox-mapping, it can be re-written as \( x_{t+1} = P_{x_t}(\eta \tilde{g}_t(x_t)) \). Applying Lemma 37 with \( u = x^* \), \( x = x_t \) and \( y = \eta \tilde{g}_t(x_t) \), we have
\[
D_\psi(x^*, x_{t+1}) \leq D_\psi(x^*, x_t) + \eta \langle \tilde{g}_t(x_t), x^* - x_t \rangle + \frac{\eta^2 \|\tilde{g}_t(x_t)\|_q^2}{2}.
\] (3.137)

For simplicity denote \( \bar{f}_t(\cdot) \equiv \mathbb{E}_u[f(\cdot + u)] \). Because \( \bar{f}_t \) is convex, we have that \( \bar{f}_t(x^*) \geq \bar{f}_t(x_t) + \langle \nabla \bar{f}_t(x_t), x^* - x_t \rangle \) for any \( x^* \in \mathcal{X} \). Plugging this into the above inequality, we have
\[
D_\psi(x^*, x_{t+1}) \leq D_\psi(x^*, x_t) - \eta (\bar{f}_t(x_t) - \bar{f}_t(x^*)) + \eta \langle \tilde{g}_t(x_t) - \nabla \bar{f}_t(x_t), x^* - x_t \rangle + \frac{\eta^2 \|\tilde{g}_t(x_t)\|_q^2}{2}.
\]

Dividing both sides of the above inequality by \( \eta \), telescoping and using the fact that \( \mathbb{E}[\tilde{g}_t(x_t) - \nabla \bar{f}_t(x_t)] = 0 \), we have
\[
\mathbb{E} \left[ \sum_{t=1}^n \bar{f}_t(x_t) - \bar{f}_t(x^*) \right] \leq \frac{D_\psi(x^*, x_0)}{\eta} + \eta \sum_{t=1}^n \mathbb{E}[\|\tilde{g}_t(x_t)\|_q^2].
\] (3.138)

On the other hand, by Lemmas 17 and 18, \( \bar{f}_t(x_t) - f(x_t) \leq L \mathbb{E}_u \|u\|_p \) and \( \mathbb{E}[\tilde{g}_t(x_t)]_q^2 \leq \mathbb{E}_u [(2\sigma^2 + L^2 \|u\|)^2 \|\nabla \log p(u)\|_q^2] \). Subsequently, taking \( x^* \in \arg\min_{x \in \mathcal{X}} f(x) \), we have
\[
\mathbb{E} \left[ \sum_{t=1}^n f(x_t) - f^* \right] \leq \frac{D_\psi(x^*, x_0)}{\eta} + L \sum_{t=1}^n \mathbb{E}_u [\|u\|_p] + \eta \sum_{t=1}^n \mathbb{E}_u [\|\sigma^2 + L^2 \|u\|_p^2 \|\nabla \log p(u)\|_q^2].
\]

Note that \( D_\psi(x^*, x_0) \leq \|x^*\|_q^2/2 \leq B \) when \( x_0 = 0 \). Applying again Lemmas 17, 18 and the scalings of \( \eta, \delta \) 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 \( F : \mathbb{R}^d \rightarrow \mathbb{R}^d \), we have
\[
\int_{K} \sum_{i=1}^d \frac{\partial F_i}{\partial u_i} \, dV = \int_{\partial K} \langle F, \ell \rangle \, dS,
\] (3.139)
where \( dV \) and \( dS \) are the volume and surface integrals on \( K \) and \( \partial K \), and \( \ell \) denotes the outer normal vector on \( \partial K \). 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 \( i \)th standard basis vector, we have \( \frac{\partial F_i}{\partial u_i} = [\nabla f(x + \delta u)]_i \), \( \frac{\partial F_j}{\partial u_i} = 0 \) for \( j \neq i \) and therefore
\[
\int_{K} \delta \nabla f(x + \delta u) \, dV = \int_{\partial K} f(x + \delta u) \ell(v) \, dS.
\]
Subsequently,
\[ \frac{\delta}{\mu_d(K)} \mathbb{E}_{u \sim \nu_K} [\nabla f(x + \delta u)] = \frac{1}{\mu_{d-1}(\partial K)} \mathbb{E}_{u \sim \sigma_{d-1}} [f(x + \delta v) \ell(v)]. \]

The lemma is then proved by using the definition that \( \tilde{f}(x) = \mathbb{E}_{u \sim \nu_K} [f(x + \delta u)] \) and interchanging the expectation and integration.

### 3.6.11 Proof of Lemma 20

The differentiability and convexity of \( \tilde{f} \) can be easily verified by definition. Because \( \tilde{f} \) is a local average of \( f \), \( \sup_{x \in \mathcal{X}} |\tilde{f}(x)| \leq \sup_{x \in \mathcal{X}} |f(x)| \leq C \). Using the convexity of \( \| \cdot \|_q \) we have \( \| \nabla \tilde{f}(x) \|_q = \| \mathbb{E} \nabla f(x + \delta u) \|_q \leq \mathbb{E} \| \nabla f(x + \delta u) \|_q \leq L \). 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 \( |f(x)| \leq C \) and \( \| \nabla f(x) \|_q \leq L \) for all \( x \in \mathcal{X} \), we know that \( \| \nabla f(x + \delta' u) \|_q \leq L \). Applying Hölder’s inequality we have
\[ |f(x + \delta u) - f(x)| \leq \delta L \| u \|_p. \]

Consequently,
\[ |\tilde{f}(x) - f(x)| \leq \mathbb{E}_{u \sim \nu_K} |f(x + \delta u) - f(x)| \leq \delta L \cdot \mathbb{E}_{u \sim \nu_K} \| u \|_p. \]

### 3.6.12 Proof of Lemma 21

For the sake of readability, in this section we shall use \( p, q \) instead of \( p, \ell \) 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
\[ \ell(v) = \frac{(\text{sgn}(v_1)|v_1|^{p-1}, \text{sgn}(v_2)|v_2|^{p-1}, \ldots, \text{sgn}(v_d)|v_d|^{p-1})}{\sqrt{|v_1|^{2(p-1)} + |v_2|^{2(p-1)} + \cdots + |v_d|^{2(p-1)}}}. \]

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
\[ \frac{d \times \mu_d(\mathbb{B}_p^d)}{\partial} = \int_{\mathbb{B}_p^d} dV(u) = \int_{\mathbb{B}_p^d} \sum_{i=1}^d \frac{\partial F_i(u)}{\partial u_i} dV(u) = \int_{\partial \mathbb{B}_p^d} \langle v, \ell(v) \rangle dS(v) \quad (3.140) \]
\[ = \int_{\partial \mathbb{B}_p^d} \langle v, (\text{sgn}(v_1)|v_1|^{p-1}, \ldots, \text{sgn}(v_d)|v_d|^{p-1}) \rangle \frac{dS(v)}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-2)}}} \quad (3.141) \]
\[ = \int_{\partial \mathbb{B}_p^d} \frac{dS(v)}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-2)}}}. \quad (3.142) \]
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 $\sigma_p^{d-1}$ on $\partial \mathbb{B}_p^d$, we obtain

$$d \times \mu_d(\mathbb{B}_p^d) = \mu_d(\partial \mathbb{B}_p^d) \times \int_{\partial \mathbb{B}_p^d} \frac{d\sigma_p^{d-1}(v)}{\sqrt{\sum_{i=1}^d |v_i|^p}}. \quad (3.143)$$

It remains to estimate the integration term in Eq. (3.143). However, the uniform surface measure $\sigma_p^{d-1}$ is complicated. To simplify the problem, we consider a closely related measure $\gamma_p^{d-1}$ 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 $\gamma_p^{d-1}$ satisfies

$$\gamma_p^{d-1}(A) = \mu_d(\{ta : a \in A, 0 \leq t \leq 1\})/\mu_d(\mathbb{B}_p^d). \quad (3.144)$$

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 $\sigma_p^{d-1}$, the cone measure $\gamma_p^{d-1}$ 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 $\gamma_p^{d-1}$ has several analytical advantages of the uniform surface measure $\sigma_p^{d-1}$. 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 rise 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

$$\mathbb{E}|X|^r = \frac{1}{\Gamma(1 + 1/p)} \int_0^\infty t^re^{-t}dt = \frac{1}{\Gamma(1 + 1/p)} \frac{1}{p} \int_0^\infty t^{(r+1)/p}e^{-t}dt = \frac{1}{\Gamma(1 + 1/p)} \frac{1}{p} \Gamma\left(\frac{r+1}{p}\right) = \frac{\Gamma((1 + r)/p)}{\Gamma(1/p)}.$$

Here in the second identity we apply the change-of-variable $t \rightarrow t^{1/p}$ and the last equality holds because $\Gamma(1 + 1/p) = \Gamma(1/p)/p$. The proposition is therefore proved. \qed
Lemma 39. For any \( p, r \in [1, \infty) \) let \( X_1, \ldots, X_d \) be i.i.d. random variables distributed according to \( f_p \) in Lemma 38. Then for any \( \omega \in [0, 1] \) it holds that

\[
\Pr \left[ (1 - d^{-1/6}) \kappa(p, r) \times d \leq \sum_{i=1}^{d} |X_i|^r \leq (1 + d^{-1/6}) \kappa(p, r) \times d \right] \geq 1 - \frac{\kappa(p, 2r)}{\kappa(p, r)^2} d^{-\omega}. \tag{3.145}
\]

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

\[
\text{Var}(S) = d \times \text{Var}(|X_i|^r) \leq d \times \mathbb{E}|X_i|^{2r} = \kappa(p, 2r) \times d. \tag{3.146}
\]

Applying the Chebyshev’s inequality, we have that for all \( \epsilon > 0 \),

\[
\Pr \left[ |S - \mathbb{E}S| > \epsilon \right] \leq \frac{\kappa(p, 2r) \times d}{\epsilon^2}. \tag{3.147}
\]

Setting \( \epsilon = \kappa(p, r) \times d^{(1+\omega)/2} \) we complete the proof of Lemma 39. \( \square \)

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 \( 1/\sum_{i=1}^{d} |v_i|^{2(p-1)} \) with respect to the cone measure \( \gamma_p^{d-1} \) on \( \partial \mathbb{B}_p^d \).

Lemma 40. For any \( p \in (1, \infty) \), we have

\[
\int_{\partial \mathbb{B}_p^d} \frac{d\gamma_p^{d-1}(v)}{\sqrt{\sum_{i=1}^{d} |v_i|^{2(p-1)}}} = (1 + o(1)) \sqrt{\frac{\kappa(p, p)^2(p-1)/p}{\kappa(p, 2(p-1))}} d^{1/2-1/p} \quad \text{as} \quad d \to \infty. \tag{3.148}
\]

Proof. Using Lemma 38, we can equivalently write \( u_i = x_i / \|x\|_p \) where \( x_1, \ldots, x_d \) 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 \( \mathcal{A}_p^r \) denote the event that Eq. (3.145) occurs with parameter \( \omega \). We shall set in the rest of the proof that \( \omega = 2/3 \). Conditioned on the event \( \mathcal{A}_p^{2/3} \cap \mathcal{A}_p^{2/3} \), 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 \( \mathcal{A}_p^{2/3} \cap \mathcal{A}_p^{2/3} \) we have that

\[
\sqrt{\frac{\sum_{i=1}^{d} |v_i|^{2(p-1)}}{(\sum_{i=1}^{d} |x_i|^p)^2(p-1)/p}} = \sqrt{\frac{(1 + o(1)) \kappa(p, 2(p-1)) \times d}{(1 + o(1)) \kappa(p, p) \times d}^{2(p-1)/p}} \tag{3.151}
\]
\[
(1 + o(1)) \sqrt{\frac{\kappa(p, 2(p - 1))}{\kappa(p, p)^2(p-1)/p}} \times d^{1/p-1/2}.
\] (3.152)

On the other hand, for all \( v \in \partial B^d_p \) satisfying \( \|v\|_p = 1 \), applying Hölder’s inequality the term \( \sum_i |v_i|^{2(p-1)} = \|v\|^{p-1}_{2(p-1)} \) can be lower bounded as
\[
\|v\|^{p-1}_{2(p-1)} \geq \left( d^{\frac{p-1}{2(p-1)} - \frac{1}{p}} \cdot \|v\|_p \right)^{p-1} = d^{1/p-1/2}, \quad p \in [2, \infty); 
\] (3.153)
\[
\|v\|^{p-1}_{2(p-1)} \geq (\|v\|_p)^{p-1} = 1, \quad p \in [1, 2]. 
\] (3.154)

Combining Eqs. (3.152,3.153,3.154) and using the total expectation formula we have
\[
\int_{\partial B^d_p} \frac{d\gamma^{d-1}_p(v)}{\sum_{i=1}^d |v_i|^{2(p-1)}} = (1 + o(1)) \sqrt{\frac{\kappa(p, p)^2(p-1)/p}{\kappa(p, 2(p-1))}} \times d^{1/2-1/p} \times \gamma^{p-1}_d (A^{2/3}_p \cap A^{2/3}_{2(p-1)})
\]
\[
\pm \max\{d^{1/2-1/p} \cdot 1, 1\} \times \gamma^{p-1}_d (- (A^{2/3}_p \cap A^{2/3}_{2(p-1)})). 
\] (3.155)

Invoking Lemma 39 with \( \omega = 2/3 \) and the union bound we know that \( \gamma^{d-1}_p (A^{2/3}_p \cap A^{2/3}_{2(p-1)}) = 1 - O(d^{-2/3}) \). Also note that both \( \kappa(p, p) \) and \( \kappa(p, 2(p-1)) \) are positive constants depending only on \( p \). Subsequently,
\[
\int_{\partial B^d_p} \frac{d\gamma^{d-1}_p(u)}{\sum_{i=1}^d |u_i|^{2(p-1)}} = (1 + o(1)) \sqrt{\frac{\kappa(p, p)^2(p-1)/p}{\kappa(p, 2(p-1))}} \times d^{1/2-1/p} + O(d^{-2/3}) 
\] (3.156)
and the lemma is proved, because \( 1/2 - 1/p \approx -2/3 \) for all \( p \in [1, \infty) \). \( \qed \)

Lemma 40 constructs an estimate of the integration term in Eq. (3.143) by replacing the uniform surface measure \( \sigma^{d-1}_p \) with the cone measure \( \gamma^{d-1}_p \) 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 \( \sigma^{d-1}_d \) and \( \gamma^{p-1}_d \) 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 \( TV(\sigma^{d-1}_d, \gamma^{d-1}_p) \leq c_p/\sqrt{d} \).

Combining Lemmas 40 and 41, we have that for all \( p \in (1, \infty) \),
\[
\int_{\partial B^d_p} \frac{d|\gamma^{d-1}_p(v) - \sigma^{d-1}_p(v)|}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-1)}}} \leq TV(\gamma^{d-1}_p, \sigma^{d-1}_p) \times \sup_{v \in \partial B^d_p} \frac{1}{\sqrt{\sum_{i=1}^d |v_i|^{2(p-1)}}} \leq \frac{c_p}{\sqrt{d}} \times \max\{d^{1/2-1/p} \cdot 1\} = o(1) \times d^{1/2-1/p}. 
\] (3.158)

Combining Eqs. (3.143,3.145,3.158) we obtain
\[
\rho(B^d_p) = \frac{\mu_{d-1}(\partial B^d_p)}{\mu_d(B^d_p)} = d \times (1 + o(1)) \sqrt{\frac{\kappa(p, 2(p-1))}{\kappa(p, p)^2(p-1)/p}} \times d^{1/p-1/2} 
\]
\[(1 + o(1)) \sqrt{\frac{\kappa(p, 2(p-1))}{\kappa(p, p)^{2(p-1)/p}}} \times d^{1/2+1/p}.\]

Lemma 22 is then proved.

### 3.6.13 Proof of Lemma 23

In this section we use the notation of \(p, q\) instead of \(p, q\) 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 \(\mathbb{E} \|\ell(v)\|^2_q = \mathbb{E} \|\ell(v)\|_1^2 = 1\). For \(p \in [2, \infty)\), because \(q \leq 2\), we can apply the Hölder’s inequality to obtain

\[
\mathbb{E} \|\ell(v)\|^2_q \leq \mathbb{E} \left[ (d^{1/q-1/2} \|\ell(v)\|_2)^2 \right] = d^{2/q-1} = d^{1/q-1/p},
\]

where the second to last equality holds because \(\ell(v)\) is an outer normal vector and therefore always has unit \(\ell_2\) norm.

We next consider the relatively more difficult case of \(p \in (1, 2)\). We shall again use the cone measure \(\gamma_{p}^{d-1}\) defined in Eq. (3.144) to approximate \(\sigma_{p}^{d-1}\), 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 \(\|\ell(v)\|_q\) can be written as

\[
\|\ell(v)\|_q^2 = \frac{\sum_{i=1}^{d} |v_i|^{2(p-1)/q}}{\|v\|_p^{2(p-1)}} = \|v\|_{q}^{2(p-1)/q} = 1/\|v\|_{2(p-1)},
\]

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 \(\gamma_{p}^{d-1}\), the random variable \(v\) can be equivalently written as \(v = x_i/\|x\|_p\), where \(x = (x_1, \ldots, x_d)\) 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 \(\ell(v) = x/\|x\|_p\), Eq. (3.160) can be re-formulated as

\[
\|\ell(v)\|_q^2 = \|x\|_p^{2(p-1)/q} = \|x\|_{2(p-1)}^{2(p-1)}. \tag{3.161}
\]

Let \(A_p^\omega \cap A_{2(p-1)}^\omega\) 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 \(A_p^\omega \cap A_{2(p-1)}^\omega\), 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 \(\sigma_{p}^{d-1}\), we have

\[
\mathbb{E}_{v \sim \sigma_{p}^{d-1}}[\|\ell(v)\|_q^2] \leq \frac{[(1 + o(1)) \kappa(p, q(p-1)) \times d]^{2(p-1)/p}}{(1 + o(1)) \kappa(p, 2(p-1)) \times d} \times \sigma_{p}^{d-1}(A_p^\omega \cap A_{2(p-1)}^\omega) + 1 \times \sigma_{p}^{d-1}(-A_p^\omega \cap A_{2(p-1)}^\omega)) \tag{3.164}
\]

92
\[ \leq O(d^{1-2/p}) + \sigma_p^{d-1}(-A_p^\omega) + \sigma_p^{d-1}(-A_{2(p-1)}^\omega) \quad (3.165) \]
\[ = O(d^{1/q-1/p}) + \sigma_p^{d-1}(-A_p^\omega) + \sigma_p^{d-1}(-A_{2(p-1)}^\omega). \quad (3.166) \]

It remains to show that \( \sigma_p^{d-1}(-A_p^\omega) \) and \( \sigma_p^{d-1}(-A_{2(p-1)}^\omega) \) 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
\[ \gamma_p^{d-1}(-A_p^\omega) + \gamma_p^{d-1}(-A_{2(p-1)}^\omega) = O(d^{2/q-1}) = O(d^{1/q-1/p}). \quad (3.167) \]

We then need to bound the discrepancy between \( \gamma_p^{d-1} \) and \( \sigma_p^{d-1} \) on \( -A_p^\omega \) and \( -A_{2(p-1)}^\omega \). Unfortunately, the TV(\( \sigma_p^{d-1}, \gamma_p^{d-1} \)) \( \leq c_p/\sqrt{d} \) 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 \( A \subseteq \mathcal{O} \)
\[ \left| \frac{\sigma_p^{d-1}(A)}{\gamma_p^{d-1}(A)} - 1 \right| \leq \frac{C}{\sqrt{d}} \cdot \sqrt{\log \left( \frac{100}{\gamma_p^{d-1}(A)} \right)}. \quad (3.168) \]

Define \( \alpha := \max\{ (2\omega - 1)/2\omega, 0 \} \) and abbreviate \( \sigma(A) = \sigma_p^{d-1}(-A_p^\omega) \), \( \gamma(A) = \gamma_p^{d-1}(-A_p^\omega) \). Because \( \omega \in (0, 1) \), we have \( \alpha \in [0, 1/2) \). Invoking Lemma 42 and Eq. (3.167), we have
\[ \sigma(A) \leq \gamma(A) + \frac{C \gamma(A)}{\sqrt{d}} \sqrt{\log \left( \frac{100}{\gamma(A)} \right)} \quad (3.169) \]
\[ = \gamma(A) + \frac{C \gamma^\alpha(A)}{\sqrt{d}} \sqrt{\gamma^{2(1-\alpha)}(A) \log \left( \frac{100}{\gamma(A)} \right)} \quad (3.170) \]
\[ = O(d^{-\omega}) + O(d^{-\omega}) \cdot \sqrt{\gamma^{2(1-\alpha)}(A) \log \left( \frac{100}{\gamma(A)} \right)}. \quad (3.171) \]

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 \( \alpha \geq (2\omega - 1)/2\omega \). 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 \( \sigma_p^{d-1}(-A_{2(p-1)}^\omega) \) and \( \gamma_p^{d-1}(-A_{2(p-1)}^\omega) \) 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 \( x \in \mathcal{X} \) and \( z \in \mathbb{R}^d \) define \( \|z\|_x := \sqrt{\langle z, \nabla^2 \phi(x) z \rangle} \). The Dikin ellipsoid \( W_1(x) \) is defined as \( W_1(x) := \{ z \in \mathbb{R}^d : \|z - x\|_x \leq 1 \} \) for all \( x \in \mathcal{X} \). It is a well-known fact that \( W_1(x) \subseteq \mathcal{X} \)
for all $x \in \mathcal{X}$ (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

$$\|z - x\|^2_x = u^T (\nabla^2 \varphi(x) + \delta I_d)^{-1/2} \nabla \varphi(x) (\nabla \varphi(x) + \delta I_d)^{-1/2} u$$

$$= \|u\|^2_2 - \delta\| (\nabla^2 \varphi(x) + \delta I_d)^{-1/2} u \|^2_2 \leq \|u\|^2_2 = 1.$$ 

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 $L_q$-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 $\pi$, 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 $R^\pi$ defined in Eq. (4.28), in $S^\pi$ the benchmark solution $x^*$ is forced to be stationary among all $T'$ epochs, resulting in smaller regret. In fact, it always holds that $S^\pi(f; x^*) \leq R^\pi(f)$ for any $f$ and $x^* \in \mathcal{X}$. In the remainder of this section, we shall refer to $S^\pi$ as the “weak regret” and $R^\pi$ 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 $1 \leq T' \leq T$. Let $f = (f_1, \cdots, f_T)$ be an arbitrary sequence of smooth and strongly convex functions satisfying (A1) through (A5). Suppose $\varphi$ is a strictly convex $\kappa$-self-concordant barrier of $\mathcal{X}$, with $\kappa = O(d)$, and $\eta = d(H + 10\sigma\sqrt{\log T})/\sqrt{2T}$. Then

$$S^\pi(f; x^*) = O(\sqrt{\frac{T' \log T}{\nu T}}), \quad \text{for all } x^* \in \mathcal{X}^0_{\nu/T}. \tag{3.173}$$

Recall the definition that $\mathcal{X}^0_{\nu/T} := \{x \in \mathcal{X}^0 : \forall z \in \mathcal{B}_d(\nu/T), x + z \in \mathcal{X}\}$ is the strict interior of $\mathcal{X}$ that is at least $\nu/T$ apart from $\partial \mathcal{X}$. Also, in both results we omit dependency on $\sigma, d, D, \nu, H, L$ 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 is the minimizer of $f_t$. By Assumption (A3) and the definition of $\mathcal{X}^0_{\nu/T}$, we have $x^*_t \in \mathcal{X}^0_{\nu/T}$.
functions the difference of the two functions. We give an illustrative example in Figure 3.5, where two functions differ by an amount that can be arbitrarily large. This difference is upper bounded by the two-point difference of the functions evaluated at two points.

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 $x^*_r$ of $f_r$, for arbitrary $r \in \{1, \cdots, T\}$. Such differences, however, can be large as $x^*_t$ could be far away from $x^*_r$ as the functions drift. In the special case of $p = \infty$, Besbes et al. (2015) observes

$$f_t(x^*_r) - f_t(x^*_t) = f_t(x^*_r) - f_r(x^*_r) + f_r(x^*_r) - f_t(x^*_t) \leq f_t(x^*_r) - f_r(x^*_r) + f_r(x^*_r) - f_t(x^*_t)$$

and further bounds both $|f_t(x^*_r) - f_r(x^*_r)|$ and $|f_t(x^*_r) - f_t(x^*_t)|$ with $\|f_t - f_r\|_\infty$. Such arguments, however, meet significant challenges in the more general setting when $1 \leq p < \infty$, 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), $f_t(x^*_r) - f_t(x^*_t)$ (i.e., the difference of the same function $f_t$ at two points) using $\|f_t - f_r\|_p$. The following is our key affinity lemma:
Lemma 44. Suppose $\mathcal{X} \subseteq \mathbb{R}^d$. Fix $1 \leq p < \infty$, $t \neq \tau$ and let $x_t^*, x_\tau^*$ be the minimizers of $f_t$ and $f_\tau$, respectively. Then under (A1) through (A5) we have that

$$\max \{ \| f_t(x_t^*) - f_t(x_\tau^*) \|, \| f_\tau(x_t^*) - f_\tau(x_\tau^*) \| \} = O(\| f_t - f_\tau \|^p_p) \quad \text{where} \quad r = \frac{2p}{2p + d} \in (0, 1).$$

Proof. Without loss of generality we assume $f_t(x_t^*) \leq f_t(x_\tau^*)$ throughout this proof. Define $\delta = \| f_t - f_\tau \|^{p/2}_p$. We first prove that $\| x_t^* - x_\tau^* \|_2 \leq 2C\delta$, 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 $x \in \mathcal{X}$ and $0 < \delta < \varepsilon \in \partial \mathcal{X}$. It is easy to verify that $\mathcal{X}_\alpha(x) \subseteq \mathcal{X}$ and $\sup_{x' \in \mathcal{X}_\alpha(x)} \| x' - x \|_2 \leq \delta x$. (recall that $D = \sup_{y \in \mathcal{X}} \| y - y' \|_2$ is the diameter of $\mathcal{X}$). In addition, $\text{vol}(\mathcal{X}_\alpha(x)) \geq \alpha^d \cdot \text{vol}(\mathcal{X})$, 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 $D \geq \| x_t^* - x_\tau^* \|_2 > 2C\delta \geq 2\delta$. By strong convexity of $f_\tau$, we have $\forall x \in \mathcal{X}_\alpha(x_t^*)$,

$$f_\tau(x) \geq f_\tau(x_t^*) + \frac{M}{2} \| x_t^* - x \|^2_2 \geq f_t(x_t^*) + \frac{M}{2} \| x_t^* - x \|^2_2 \quad (3.176)$$

Here Eq. (3.176) holds because $f_\tau(x_t^*) \geq f_t(x_t^*)$, and Eq. (3.177) is true because $\| x_t^* - x \|^2_2 > 2C\delta$ and $\| x - x_t^* \|^2_2 \leq \alpha D = \delta \leq C\delta$ for all $x \in \mathcal{X}_\alpha(x_t^*)$. On the other hand, by smoothness of $f_t$, we have that

$$f_t(x) \leq f_t(x_t^*) + \frac{L}{2} \| x - x_t^* \|^2_2 \leq f_t(x_t^*) + L\delta^2 \quad \forall x \in \mathcal{X}_\alpha(x_t^*). \quad (3.178)$$

Combining Eqs. (3.177,3.178) we have that, for arbitrary $1 \leq p < \infty$ and $x \in \mathcal{X}_\alpha(x_t^*)$

$$\left| f_\tau(x) - f_t(x) \right|^p \geq \left| \left( f_t(x_t^*) + \frac{MC^2}{2} \delta^2 \right) - \left( f_t(x_t^*) + L\delta^2 \right) \right|^p \geq (MC^2/2 - L)^p \delta^{2p}, \quad (3.179)$$

provided that $L \leq MC^2/2$, which holds true because $C \geq \sqrt{2L/M}$ by definition. Integrating both sides of Eq. (3.179) on $\mathcal{X}_\alpha(x_t^*)$ and recalling the definition of $\| f_t - f_\tau \|^p_p$, we have that

$$\| f_t - f_\tau \|^p_p = \frac{1}{\text{vol}(\mathcal{X})} \int_{\mathcal{X}} \| f_t(x) - f_\tau(x) \|^p \, dx \geq \frac{1}{\text{vol}(\mathcal{X})} \int_{\mathcal{X}_\alpha(x_t^*)} \| f_t(x) - f_\tau(x) \|^p \, dx \geq \frac{\text{vol}(\mathcal{X}_\alpha(x_t^*))}{\text{vol}(\mathcal{X})} \cdot (MC^2/2 - L)^p \delta^{2p} \geq \frac{\delta^d}{D^d} \cdot (MC^2/2 - L)^p \delta^{2p} \geq \frac{\left( MC^2/2 - L \right)^p}{D^d} \delta^{2p + d} \| f_t - f_\tau \|^p_p,$$

where the last equality holds because $\delta = \| f_t - f_\tau \|^{p/2}_p$ and $(2p + d) \cdot r/2 = p$. With $C \geq \sqrt{(4D^d/p + 2L)/M}$, we have that $(MC^2/2 - L)^p/D^d \geq 2^d > 1$ and hence the contradiction.
We have now established that \( \| x_t^* - x_r^* \|_2 \leq 2C\delta \leq O(\delta) \). By smoothness of \( f_t \) and \( f_r \),

\[
\begin{align*}
    f_t(x_t^*) &\leq f_t(x_r^*) \leq f_t(x_t^*) + \frac{L}{2} \| x_t^* - x_r^* \|_2^2 \leq f_t(x_t^*) + O(\delta^2); \\
    f_r(x_r^*) &\leq f_r(x_t^*) \leq f_r(x_r^*) + \frac{L}{2} \| x_t^* - x_r^* \|_2^2 \leq f_r(x_r^*) + O(\delta^2).
\end{align*}
\]

The proof of Lemma 44 is then completed by plugging in \( \delta = \| f_t - f_r \|_p^{\ell/2} \).

**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 \( \Delta_T \) or \( \Delta_T + 1 \) epochs. Applying Lemmas 43, 44 together with Eq. (3.174) we have

\[
R^x(f) \leq \sum_{t=1}^{J} \inf_{\tau \in B_t} \left\{ S^x(f_{2t}, \cdots, f_{ht}; x_t^*) + \sum_{t=1}^{\bar{b}_t} f_t(x_t^*) - f_t(x_t^*) \right\}
\]

\[
\leq \sum_{t=1}^{J} O(\sqrt{|B_t| \log |B_t|}) + |B_t| \cdot \sup_{t, \tau \in B_t} \| f_t(x_t^*) - f_t(x_t^*) \|
\]

\[
\leq O \left( \frac{T}{\sqrt{\Delta_T}} \cdot \sqrt{\Delta_T \log \Delta_T} \right) + O(\Delta_T) \cdot \sum_{t=1}^{J} \sup_{t, \tau \in B_t} \| f_t - f_r \|_p^r
\]

\[
\leq O \left( \frac{T \log T}{\sqrt{\Delta_T}} \right) + O(\Delta_T) \cdot \sum_{t=1}^{J} \left( \sum_{t=1}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p \right)^r.
\]

(3.180)

Here the last inequality holds because (assuming without loss of generality that \( \bar{b}_t \leq t \leq \tau \leq \bar{b}_t \))

\[
\| f_t - f_r \|_p \leq \sum_{k=t}^{\bar{b}_t-1} \| f_{k+1} - f_k \|_p \leq \sum_{k=t}^{\bar{b}_t-1} \| f_{k+1} - f_k \|_p.
\]

We next present another key lemma that upper bounds the critical summation term in Eq. (3.180) using \( J, \Delta_T \) and \( \text{Var}_{p,q}(f) \). The proof is based on consecutively applying the Hölder’s inequality.

**Lemma 45.** Suppose \( \max_{1 \leq \ell \leq J} |B_{\ell}| \leq \Delta_T + 1 \), \( 1 \leq q \leq \infty \) and \( \text{Var}_{p,q}(f) \leq V_T \). Then

\[
\sum_{t=1}^{J} \left( \sum_{t=1}^{\bar{b}_{t\ell}-1} \| f_{t+1} - f_t \|_p \right)^r \leq \Delta_T^{r-1/r \cdot J^{1-r/q}} \cdot T^{r/q} \cdot V_T^r.
\]

**Proof.** By Hölder’s inequality, for any \( d \)-dimensional vector \( x \) we have that

\[
\| x \|_\alpha \leq \| x \|_\beta \leq d^{1/\beta - 1/\alpha} \| x \|_\alpha \quad \forall \ 0 < \beta \leq \alpha \leq \infty.
\]

Apply Eq. (3.181) with \( \alpha = q \) and \( \beta = 1 \) on \( x = (\| f_{2t+1} - f_{2t} \|_p, \cdots, \| f_{\bar{b}_t} - f_{\bar{b}_t-1} \|_p) \in \mathbb{R}^{\bar{b}_t-1} \):

\[
\sum_{t=\bar{b}_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p = \| x \|_1 \leq |B_{\ell} - 1|^{1-1/q} \| x \|_q \leq \Delta_T^{1-1/q} \cdot \left( \sum_{t=\bar{b}_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p^q \right)^{1/q}.
\]

97
Subsequently,

\[
\sum_{t=1}^{T} \left( \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p \right)^{r/q} \leq \sum_{t=1}^{T} \Delta_T^{r-r/q} \cdot \left( \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p \right)^{r/q}.
\]  
(3.182)

We next consider \( \tilde{x} = (\tilde{x}_1, \cdots, \tilde{x}_d) \in \mathbb{R}^d \), where \( \tilde{x}_i = \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p^q \). Apply Eq. (3.181) with \( \alpha = 1 \) and \( \beta = r/q \) on \( \tilde{x} \) (\( \beta < 1 \) because \( r \in (0, 1) \) and \( q \geq 1 \)):

\[
\left[ \sum_{t=1}^{\bar{b}_t-1} \left( \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p \right)^{r/q} \right]^{q/r} = \| \tilde{x} \|_q^q \leq J^{1/\beta-1/\alpha} \cdot \| \tilde{x} \|_1 = J^{q/r-1} \cdot \sum_{t=1}^{T} \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p^q.
\]

Raise both sides of the inequality to the power of \( r/q \) and note that \( \sum_{t=1}^{T} \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p^q = \sum_{t=1}^{T-1} \| f_{t+1} - f_t \|_p^q \leq T \cdot V_T^q \). We then have

\[
\sum_{t=1}^{T} \left( \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p \right)^{r/q} \leq J^{1-r/q} \cdot \left( \sum_{t=1}^{T} \sum_{t=b_t}^{\bar{b}_t-1} \| f_{t+1} - f_t \|_p \right)^{r/q} \leq J^{1-r/q} T^{r/q} V_T^r.
\]  
(3.183)

Combining Eqs. (3.182,3.183) we proved the desired lemma.

We now prove Theorem 11 by combining Lemmas 43, 44 and 45 with Eq. (3.180) and setting \( \Delta_T \) appropriately. By Lemma 43, \( S^x(f_{\bar{b}_t}, \cdots, f_{\bar{b}_t}; x_T^x) \leq O(\sqrt{b_{\bar{b}_t}} \log T) \leq O(\sqrt{\Delta_T} \log T) \) for \( \phi = \phi_t(x_t, f_t) \). Subsequently,

\[
R^x_{\phi}(f) \leq O(J \sqrt{\Delta_T} \log T) + O(\Delta_T^{1+r-r/q} J^{1-r/q} T^{r/q} V_T^r).
\]

If \( V_T = O(T^{-(6p+d)/4p}) \), then we set \( \Delta_T = T \), \( J = 1 \) and obtain regret \( \tilde{O}(\sqrt{T}) + O(T^{1+r} V_T^{r}) = \tilde{O}(\sqrt{T}) \). 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 \( \tilde{O}(T \cdot V_T^{2p/(6p+d)}) \).

### 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 $f, \tilde{f} : \mathcal{X} \to \mathbb{R}$:

$$\chi(f, \tilde{f}) := \inf_{x \in \mathcal{X}} \max \left\{ f(x) - f^* , \tilde{f}(x) - \tilde{f}^* \right\} \quad \text{where} \quad f^* = \inf_{x \in \mathcal{X}} f(x), \tilde{f}^* = \inf_{x \in \mathcal{X}} \tilde{f}(x).$$

Intuitively, $\chi(f, \tilde{f})$ characterizes the best regret $f(x) - f^*$ one could achieve without knowing whether $f$ or $\tilde{f}$ 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 $\pi$ 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 $\Theta$ with large probability:

**Lemma 46.** Fix $1 \leq p < \infty$, $1 \leq q \leq \infty$ 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 $\pi$ such that

$$\sup_{f \in \mathcal{F}_{p,q}(V_T)} R_\pi^T(f) \leq \frac{1}{9} \cdot \inf_{f, f, \Theta} \sum_{t=1}^{T} \chi(f_t, \tilde{f}_t),$$

then there exists an estimator $\hat{f}$ such that

$$\sup_{f \in \Theta} \Pr_{f} \left[ f \neq \hat{f} \right] \leq 1/3,$$

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 $\pi$ 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 $\Theta$ 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 $\text{KL}(P||Q) = \int \log \frac{dP}{dQ} dP$ 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 $P_\theta$ be the distribution of observations parameterized by $\theta$. Suppose there exists $0 < \beta < \infty$ such that $\text{KL}(P_\theta||P_{\theta'}) \leq \beta$ for all $\theta, \theta' \in \Theta$. Then

$$\inf_{\theta} \sup_{\theta'} \Pr_{f} \left[ \hat{\theta} \neq \theta \right] \geq \left( \frac{1}{M} + \frac{2}{\beta} \right)^{-\beta},$$

where $\Pr_f$ denotes the probability distribution parameterized by the underlying true function sequence $f \in \Theta$.
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 $F_\lambda$ with intermediate values $\lambda = 0.25$ and $\lambda = 0.75$. For the sake of better visualization, the regions $0 \leq x \leq \sqrt{h}$ ($I_1$) and $\sqrt{h} \leq x \leq 2\sqrt{h}$ ($I_2$) are greatly exaggerated. In the actual construction both regions are very small compared to the entire domain $X = [0, 1]$. 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 $\lambda$ first increase and then decrease, or vice versa, depending on the coding $i_j \in \{0, 1\}$ for the particular batch. Also note that $\lambda$ 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 $\Theta \subseteq F_{p,q}(V_T)$ such that the minimum discrepancy $\inf_{f,\tilde{f} \in \Theta} \sum_{t=1}^{T} \chi(f_t, \tilde{f}_t)$ is lower bounded and the maximum KL divergence $\sup_{f,\tilde{f} \in \Theta} \text{KL}(P_f \| \tilde{P}_f)$ 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” $\Theta$ 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 $1/8T^2 \leq h \leq 1/8$. Define $F_0, F_1 : \mathcal{X} \to \mathbb{R}$ as follows:

$$F_0(x) := \begin{cases} x^2, & 0 \leq x < \sqrt{h}; \\ \frac{4}{h}x^3 - 11x^2 + 12\sqrt{h}x - 4h, & \sqrt{h} \leq x < 2\sqrt{h}; \\ 8(x - \sqrt{h})^2, & 2\sqrt{h} \leq x \leq 1. \end{cases}$$

(3.187)

$$F_1(x) := \begin{cases} (x - \sqrt{h})^2, & 0 \leq x < \sqrt{h}; \\ 8(x - \sqrt{h})^2, & \sqrt{h} \leq x \leq 1. \end{cases}$$

(3.188)

Further define

$$F_\lambda := F_0 + \lambda(F_1 - F_0), \quad \lambda \in [0, 1]$$

(3.189)
as a convex combination of $F_0$ and $F_1$. Figure 3.6 gives a graphical sketch of $F_0$, $F_1$ and $F_\lambda$. 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 $F_\lambda$ 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 $F_\lambda$.

**Lemma 48.** The following statements are true for all $\lambda, \mu \in [1/4, 3/4]$.

1. $F_\lambda$ satisfies (A1) through (A5) with $D = 2$, $\nu = 1/64$, $H = 16$, $L = 26$ and $M = 2$.
2. $\|F_\lambda - F_\mu\|_\infty \leq |\lambda - \mu| \cdot O(h)$ and $\|F'_\lambda - F'_\mu\|_\infty \leq |\lambda - \mu| \cdot O(\sqrt{h})$.
3. $\|F_\lambda - F_\mu\|_p \leq |\lambda - \mu| \cdot O(h^{(2p+1)/2p})$ for all $1 \leq p < \infty$.
4. $\chi(F_\lambda, F_{\lambda-\lambda}) = |1/2 - \lambda|^2 \cdot h/4$.

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 $1 \leq q \leq \infty$ due to the monotonicity of $\text{Var}_{p,q}(f)$. Therefore we shall focus solely on the $q = \infty$ case, whose construction is automatically valid for all $1 \leq q < \infty$.

Let $1 \leq J \leq T$ 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 $B_1, \ldots, B_J$, where each batch consists of either $\Delta_T$ 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 $f_i = (f_{i,1}, \ldots, f_{i,T})$ defined as

$$f_{i,(j-1)\Delta_T+\ell} = \begin{cases} F_{0.5+0.5\ell/|B_j|}, & i_j = 0, 1 \leq \ell \leq |B_j|/2; \\ F_{0.75-0.5\ell/|B_j|}, & i_j = 0, |B_j|/2 < \ell \leq |B_j|; \\ F_{0.5-0.5\ell/|B_j|}, & i_j = 1, 1 \leq \ell \leq |B_j|/2; \\ F_{0.25+0.5\ell/|B_j|}, & i_j = 1, |B_j|/2 < \ell \leq |B_j|. \end{cases}$$ (3.190)

Figure 3.6 gives a visual illustration of the change pattern of $f_i$ and $f_\ell$ by plotting the values of $\lambda$ for each function in the constructed sequences. For a particular batch $B_j$, when $i_j = i'_j$ then $f_i$ and $f_{\ell'}$ 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 $f_{\ell'}$ will drift towards $F_1$, creating gaps between $f_i$ and $f_{\ell'}$ within batch $B_j$. For regularity reasons, we constrain the $\lambda$ value to be within the range of $(0.25, 0.75)$ regardless of $i_j$ values. We also note that $f_i$ and $f_\ell$ 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 $f_\ell$:

**Lemma 49.** Suppose $\xi_{i,d} \sim \mathcal{N}(0, 1)$ For any $i, \ell \in \{0, 1\}^J$ consider $f_i$ and $f_\ell$ as defined in Eq. (3.190). Then the following statements are true:

1. (Variation). $\text{Var}_{p,q}(f) \leq \text{Var}_{p,\infty}(f) \leq O(h^{(2p+1)/2p}/\Delta_T)$, for all $1 \leq p < \infty$ and $1 \leq q \leq \infty$.
2. (Discrepancy). $\sum_{t=1}^T \chi(f_{i,t}, f_{\ell',t}) \geq \Delta_H(i, \ell') \cdot \Omega(h\Delta_T)$, where $\Delta_H(i, \ell') = \sum_{j=1}^J [i_j \neq \ell'_j]$ is the Hamming distance between $i$ and $\ell'$. 

101
3. (KL divergence). Let \( P_{f_i}^T \) 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 \( \pi \). Then for any such policy \( \pi \) we have that

\[
\text{KL}(P_{f_i}^T \| P_{f_j}^T) \leq \Delta_H(i, i') \cdot O(h^2 \Delta_T).
\]

The proof of Lemma 49 is deferred later.

Finally, we describe the construction of \( \mathcal{I} \subseteq \{0, 1\}^J \) and the choices of \( J, \Delta_T \) 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 \( \Delta_H(i, i') \geq J/16 \). 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 \( J \geq 2 \) 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 \( \forall i, i' \in \mathcal{I}, \Delta_H(i, i') \geq J/16 \). Furthermore, \( \log |\mathcal{I}| \geq 0.0625J \).

The univariate case of Theorem 12 can then be proved by appropriately setting the scalings of \( h, \Delta_T \) and invoking Lemmas 48, 49 and 50. Because an \( O(\sqrt{T}) \) 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 \( V_T = \Omega(T^{-6p/(4p+1)}) \). More specifically, we set \( h \approx V_T^{2p/(6p+1)} \) and \( \Delta_T \approx V_T^{-4p/(6p+1)} \).

It is easy to verify that with the additional lower bound on \( V_T, \Delta_T = o(T) \) and \( h \approx 1/T^2 \), 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 \( \mathcal{X} = \{ x \in \mathbb{R}^d : 0 \leq x, 1^\top x \leq 1 \} \). Define \( \overline{F}_\lambda : \mathcal{X} \to \mathbb{R} \) as follows:

\[
\overline{F}_\lambda(x) := F_\lambda(1^\top x) + \|x\|_2^2, \quad \lambda \in [0, 1], x \in \mathcal{X}.
\]

Here \( F_\lambda \) 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 \( F_\lambda \). The following lemma lists the properties of \( \overline{F} \), which are rigorously verified later this section.

**Lemma 51.** Suppose \( 1/8T^2 \leq h \leq 1/8 \). 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. \( \| \overline{F}_\lambda - \overline{F}_\mu \|_\infty \leq |\lambda - \mu| \cdot O(h) \) and \( \sup_{x \in \mathcal{X}} \| \nabla \overline{F}_\lambda(x) - \nabla \overline{F}_\mu(x) \|_2 \leq |\lambda - \mu| \cdot O(\sqrt{h}) \).
3. $\|F_\lambda - F_\mu\|_p \leq |\lambda - \mu| \cdot O(h^{(2p+d)/2p})$ for all $1 \leq p < \infty$.

4. $\chi(F_\lambda, F_{1-\lambda}) = \frac{d}{d+1} (\frac{1}{2} - \lambda)^2 \cdot h$.

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 $\|F_\lambda - F_\mu\|_p$ 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 $\Theta \subseteq F_{p,q}(V_T)$ in the univariate case and invoking Lemma 51:

**Corollary 3.** Suppose $1 \leq J \leq T$ is even, $\Delta_T = |T/J|$ and $1/8T^2 \leq h \leq 1/8$. 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 $F_\lambda$ is replaced with its high-dimensional version $F_\lambda$ defined in Eq. (3.191). Then the following holds:

1. (Variation). $\sup_{f \in \Theta} \text{Var}_{p,q}(f) \leq O\left(h^{(2p+d)/2p}/\Delta_T\right)$ for all $1 \leq p < \infty$, $1 \leq q \leq \infty$.

2. (Discrepancy). $\inf_{f, \tilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \tilde{f}_t) \geq \Omega(hT)$.

3. (KL-divergence). For all admissible policy $\pi$, $\sup_{f, \tilde{f} \in \Theta} \text{KL}(P_f^\pi \| P_{\tilde{f}}^\pi) \leq O(h^2T)$.

We now prove the multi-dimensional case based on Corollary 3. Set $h = V_T^{2p/(6p+d)}$ and $\Delta_T$ accordingly such that $\text{Var}_{p,q}(f) \leq O\left(h^{(2p+d)/2p}/\Delta_T\right) = V_T$. This yields $\Delta_T = V_T^{-4p/(6p+d)}$ and $J = T/\Delta_T = TV_T^{4p/(6p+d)}$. The KL divergence is then upper bounded by $O(h^2T) = O(TV_T^{4p/(6p+d)})$ 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 $\pi$ such that $\sup_{f \in F_{p,q}(V_T)} R^\pi(f) \leq 1/9 \cdot \inf_{f, \tilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \tilde{f}_t)$. The lower bound proof is then completed by the discrepancy claim in Corollary 3 that $\inf_{f, \tilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \tilde{f}_t) \geq \Omega(h^2T) = \Omega(TV_T^{2p/(6p+d)})$.

**Proof of Lemma 67** Let $\pi$ be a policy that attains the minimax rate. By Markov’s inequality, with probability $2/3$ it holds that

$$\frac{1}{T} \sum_{t=1}^T f_t(x_t) - f_t(x^*_t) \leq \frac{1}{3} \cdot \inf_{f, \tilde{f} \in \Theta} \sum_{t=1}^T \chi(f_t, \tilde{f}_t), \quad \forall f, \tilde{f} \in \Theta. \quad (3.192)$$

Define $\hat{f} := \arg\min_{f \in \Theta} \sum_{t=1}^T \hat{f}_t(x_t) - \hat{f}_t(x^*_t)$, where $x^*_t$ is the (unique) minimizer of $\hat{f}_t$. Let $\hat{f}^* = \inf_{x \in X} \hat{f}_t(x)$ and $f^*_t = \inf_{x \in X} f_t(x)$. Because $\hat{f}$ minimizes the “empirical” regret on $\{x_t\}_{t=1}^T$, it holds that

$$\sum_{t=1}^T \hat{f}_t(x_t) - \hat{f}^* \leq \sum_{t=1}^T f_t(x_t) - f^*_t.$$

Subsequently,

$$\sum_{t=1}^T \chi(\hat{f}_t, f_t) = \sum_{t=1}^T \inf_{x \in X} \max\left\{\hat{f}_t(x) - \hat{f}^*, f_t(x) - f^*_t\right\}.$$
\begin{align*}
\leq & \sum_{t=1}^{T} \max \left\{ \hat{f}_t(x_t) - \hat{f}_t^*, f_t(x_t) - f_t^* \right\} \\
\leq & \sum_{t=1}^{T} \hat{f}_t(x_t) - \hat{f}_t^* + f_t(x_t) - f_t^* \\
\leq & 2 \left( \sum_{t=1}^{T} f_t(x_t) - f_t^* \right) \leq \frac{2}{3} \cdot \inf_{f_t \in \Theta} \sum_{t=1}^{T} \chi(f_t, \hat{f}_t).
\end{align*}

Therefore, we must have $\hat{f} = f$ 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 differentiability 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 $F(x^+) = \lim_{t \to 0^+} F(x + t)$, $F(x^-) = \lim_{t \to 0^-} F(x + t)$, $F'(x^+) = \lim_{t \to 0^+} \frac{F(x+h)-F(x)}{h}$ and $F'(x^-) = \lim_{t \to 0^-} \frac{F(x+h)-F(x)}{h}$.

We then have that $F_1(\sqrt{h}^-) = F_0(\sqrt{h}^-) = h$, $F_0(2\sqrt{h}^+) = F_0(2\sqrt{h}^-) = 8h$, $F_0'(\sqrt{h}^+) = F_0'(\sqrt{h}^-) = 2\sqrt{h}$, $F_0'(2\sqrt{h}^+) = F_0'(2\sqrt{h}^-) = 16\sqrt{h}$, $F_2(\sqrt{h}^-) = F_1(\sqrt{h}^-) = F_1'(\sqrt{h}^+) = F_1'(\sqrt{h}^-) = 0$. Therefore, both $F_0$ and $F_1$ are differentiable on $[0, 1]$. It is then easy to check that $\sup_{0 \leq x \leq 1} \max\{|F_0(x)|, |F_1(x)|\} \leq 8$ and $\sup_{0 \leq x \leq 1} \max\{|F_0'(x)|, |F_1'(x)|\} \leq 16$. 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 $2 \leq F''_0(x) \leq 26$ 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 $x^*_\lambda$ be the unique minimizer of $F_\lambda = F_0 + \lambda(F_1 - F_0)$. Elementary algebra yields that $x^*_\lambda = \lambda\sqrt{h}$. Because $h \geq 1/8T^2$, we know that $F_\lambda$ 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 $\|F_\lambda' - F_\mu'\|_p \leq |\lambda - \mu| \cdot \|F_0' - F_1'\|_p$ for all $1 \leq p \leq \infty$. 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

$$
\|F_0 - F_1\|_p = \left( \int_{0}^{2\sqrt{h}} |F_0(x) - F_1(x)|^p \, dx \right)^{1/p} = O(h^{1/2p}) \cdot \|F_0 - F_1\|_\infty = O(h^{(2p+1)/2p}).
$$

**Verification of property 4:** We have that $x^*_\lambda = \lambda\sqrt{h}$ and $F^*_\lambda = F_\lambda(x^*_\lambda) = \lambda(1 - \lambda)h$. Subsequently, $\chi(F_\lambda, F_{1-\lambda}) = F_\lambda(\sqrt{h}/2) - F^*_\lambda = (1/2 - \lambda)^2 \cdot h/4$. 

104
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 $\Delta_T$ 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,

$$\text{Var}_{p,\infty}(f) = \sup_{1 \leq t \leq T-1} \|f_{t+1} - f_t\|_p = O(h^{(2p+1)/2p}/\Delta_T).$$

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,

$$\sum_{t \in I_j} \chi(f_{i,t}, f_{i',t}) \geq 2 \sum_{t=0}^{[\Delta_T/2]} \chi(F_{0.5+t/\Delta_T}, F_{0.5-t/\Delta_T}) \geq 2 \sum_{t=0}^{[\Delta_T/2]} \left(\frac{t}{\Delta_T}\right)^2 \cdot \Omega(h) = \Omega(h\Delta_T).$$

Subsequently, summing over all intervals with $i_j \neq i'_j$ we have that $\sum_{t=1}^T \chi(f_{i,t}, f_{i',t}) \geq \Delta_I(i, i') \cdot \Omega(h\Delta_T)$.

Finally we compute the KL divergence $\text{KL}(P^\pi_{f_i} || P^\pi_{f_{i'}})$. Let $y_t = f_{i}(x_t) + \xi_t$ 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 $\pi$, we have that

$$\text{KL}(P^\pi_{f_i} || P^\pi_{f_{i'}}) = \mathbb{E}_{f_i, \pi} \left[ \log \frac{P^\pi_{f_{i}}(x^T, y^T)}{P^\pi_{f_{i'}}(x^T, y^T)} \right]$$

$$= \mathbb{E}_{f_i, \pi} \left[ \log \frac{P_i(y^T|x^T) \cdot \prod_{t=1}^{T} P_{\pi}(x_{t+1}|y_{t-1}, x_{t-1})}{P_{i'}(y^T|x^T) \cdot \prod_{t=1}^{T} P_{\pi}(x_{t+1}|y_{t-1}, x_{t-1})} \right]$$

$$= \sum_{t=1}^{T} \mathbb{E}_{f_i, \pi} \left[ \log \frac{P_{i,t}(y_t|x_t)}{P_{i',t}(y_t|x_t)} \right]$$

$$\leq \sum_{t=1}^{T} \sup_{x \in X} \text{KL}(P_{i,t}(\cdot|x) || P_{i',t}(\cdot|x)).$$

Here the third identity holds because $\varepsilon_t$ are independent. For $y_t = f_{i}(x_t) + \xi_t \sim \mathcal{N}(f_{i}(x_t), 1)$, it holds that

$$\sup_{x \in X} \text{KL}(P^\phi_{f_{i,t}}(\cdot|x) || P^\phi_{f_{i',t}}(\cdot|x)) = \sup_{x \in X} \left| f_{i,t}(x) - f_{i',t}(x) \right|^2 = \|f_{i,t} - f_{i',t}\|_\infty^2 = O(h^2).$$

where in the last inequality we invoke Lemma 48. Summing over $t = 1$ to $T$ we have that

$$\text{KL}(P^\pi_{f_i} || P^\pi_{f_{i'}}) = O(h^2T).$$

Proof of Lemma 51  We verify the properties separately. Verification of property 1: Because $\forall x \in \mathcal{X}$, $\|x\|_1 \leq 1$, we have that $\|x - y\|_2 \leq \|x - y\|_1 \leq 2$ for all $x, y \in \mathcal{X}$ and therefore $\mathcal{X}$ satisfies (A1) with $D = 2$.  

105
We next verify (A3). Because $F_\lambda$ is convex differentiable, it holds that $\overline{F}_\lambda(x) = F_\lambda(1^T x) + \|x\|_2^2$ is also convex differentiable because convexity is preserved with affine transform. In particular, $\sup_{x \in \mathcal{X}} \overline{F}_\lambda(x) \leq \|F_\lambda\|_\infty + 1 \leq 9$ and $\sup_{x \in \mathcal{X}} \|\overline{F}_\lambda(x)\|_2 \leq \sup_{x \in \mathcal{X}} |F_\lambda(1^T x)| \cdot \|1\|_2 + 2 \|x\|_2 \leq 16\sqrt{d} + 2$. 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 $\{x : 1^T x = \sqrt{\lambda h}\} \cup \{x : 1^T x = 2\sqrt{\lambda h}\}$. Furthermore, $\nabla^2\overline{F}_\lambda(x) = F_\lambda(1^T x) \cdot 11^T + 2I_d$. 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 $x^*_\lambda$ be the unique minimizer of $\overline{F}_\lambda$ on $\mathcal{X}$. It is clear that $x^*_\lambda$ 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 $F_\lambda(1^T x)$. Completing the squares in $\overline{F}_\lambda$ we have that

$$\overline{F}_\lambda(\overline{x}_\lambda) = d(d + 1) \left[ \overline{x} - \frac{\lambda \sqrt{\lambda}}{d + 1} \right]^2 + \lambda h \left[ 1 - \frac{\lambda d}{d + 1} \right].$$

Subsequently, $\overline{x}_\lambda = \frac{\lambda \sqrt{\lambda}}{d + 1}$. It is easy to verify that for $\lambda \leq 1/8$, $\inf \{ t \geq 0 : \overline{x}_\lambda + tu \in \mathcal{X} \forall u \in B_d(1) \} \geq \|\overline{x}_\lambda\|_2 \geq \lambda \sqrt{\lambda h} / (d + 1)$. Therefore, for all $\lambda \in [1/4, 3/4]$ and $1/8 T^2 \leq h_1 \leq 1/8$ 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, $\sup_{x \in \mathcal{X}} \|\nabla \overline{F}_\lambda(x) - \nabla \overline{F}_\mu(x)\|_2 = \|F_\lambda' - F_\mu'\|_\infty \cdot \|1\|_2 = O(\sqrt{h d})$. Omitting the dependency on $d$ we obtain property 2.

**Verification of property 3:** Define $\overline{B}_d(r) := \{x \in \mathbb{R}^d : x \geq 0, \|x\|_1 \leq r \}$. It is easy to verify that $\text{vol}(\overline{B}_d(r_1)) / \text{vol}(\overline{B}_d(r_2)) = (r_1/r_2)^d$. Subsequently, for any $1 \leq p < \infty$ we have that

$$\|\overline{F}_0 - \overline{F}_1\|_p \leq \left[ \frac{\text{vol}(\overline{B}_d(2\sqrt{h}))}{\text{vol}(\overline{B}_d(1))} \cdot \|\overline{F}_0 - \overline{F}_1\|_\infty^p \right]^{1/p} = O(h (2p + d)/2p).$$

**Verification of property 4:** From previous derivations we know that $x^*_\lambda = (\overline{x}^*_\lambda, \cdots, \overline{x}^*_\lambda)$ with $\overline{x}_\lambda = \frac{\lambda \sqrt{\lambda}}{d + 1}$ and $\overline{F}_\lambda = \inf_{x \in \mathcal{X}} \overline{F}_\lambda(x) = \lambda h (1 - \frac{\lambda}{d + 1})$. Subsequently,

$$\chi(F_\lambda, F_{1-\lambda}) = \overline{F}_\lambda \left( \frac{1}{2} \frac{\sqrt{\lambda}}{d + 1} \right) = \frac{d}{d + 1} \left[ \frac{1}{2} - \lambda \right]^2 \cdot h.$$
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 \( r_i \in [0, 1] \) 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 probabilistic model

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

where \( \theta_0 \) 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 \( v_i = \exp\{x_i^T \theta_0\} \). 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

\footnote{The constraint \( r_i \leq 1 \) 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 \( v_i \). Given assortment \( S \subseteq [N] \), the logit choice model posits that (define \( v_0 := 0 \) for notational simplicity)

\[
\Pr[i | S] = \frac{v_i}{1 + \sum_{i \in S} v_i}, \quad i \in S \cup \{0\}.
\]  

(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

\[
R(S) = \mathbb{E}_{i \sim \mathcal{P}(S)}[r_i] = \frac{\sum_{i \in S} r_i v_i}{1 + \sum_{i \in S} v_i}.
\]  

(4.2)

Suppose a policy \( \pi \) 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

\[
\text{Regret} \left( \{S_t\}_{t=1}^T \right) := \sum_{t=1}^T R(S^*) - \mathbb{E}^\pi \left[ R(S_t) \right] \quad \text{where} \quad S^* \in \arg\max_S R(S).
\]  

(4.3)

We also impose the condition \( \max_i r_i \leq 1 \) 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 \( \mathcal{L}_\theta := \{i \in [N] : r_i \geq \theta\} \) satisfies \( R(\mathcal{L}_\theta) = R(S^*) \).

Proposition 14 suggests that it suffices to consider “level-set” type assortments \( \mathcal{L}_\theta = \{i \in [N] : r_i \geq \theta\} \) 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 \( \theta \). 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...
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 $S = \{s_1, \ldots, s_m\}$ 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 $c_0, \ldots, c_m \geq 0$ satisfying $c_i \neq c_{i+1}$ for all $i = 0, \ldots, m - 1$, and $S = \{s_1, \ldots, s_m\} \subseteq \{r_i\}_{i=1}^N$, such that

$$F(\theta) = c_0 \cdot \mathbb{I}[\theta \leq s_1] + \sum_{i=1}^{m-1} c_i \cdot \mathbb{I}[s_i < \theta \leq s_{i+1}] + c_m \cdot \mathbb{I}[\theta > s_m],$$

where $c_m = 0$.

Define $F^* := \max_{0 \leq i \leq m} c_i = \sup_{\theta \geq 0} F(\theta)$ 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 $\theta \geq \theta^*$, $F(\theta) \leq F(\theta^*)$. 

**Lemma 54.** For any $\theta \leq \theta^*$, $F(\theta) \geq F(\theta^*)$.

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 $\theta^*$ that attains the maximum function value $F^*$.
and monotonically decreases as one moves away from $\theta^*$, 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 $\theta^*$ and above the $y = x$ line to the left of $\theta^*$. This helps us judge the positioning of a particular revenue level $\theta$ by simply comparing the expected revenue of $R(L_\theta)$ with $\theta$ 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 $\theta^*$ 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:
- $a_\tau$ and $b_\tau$: left and right boundaries that contain $\theta^*$; it is guaranteed that $a_\tau \leq \theta^* \leq b_\tau$ with high probability, and the regret incurred on failure events is strictly controlled;
- $x_\tau$ and $y_\tau$: trisection points; $x_\tau$ is closer to $a_\tau$ and $y_\tau$ is closer to $b_\tau$;
- $\ell_t(y_\tau)$ and $u_t(y_\tau)$: lower and upper confidence bands for $F(y_\tau)$ established at iteration $t$; it is guaranteed that $\ell_t(y_\tau) \leq F(y_\tau) \leq u_t(y_\tau)$ with high probability, and the regret incurred on failure events is strictly controlled;
- $\rho_t(y_\tau)$: accumulated reward by exploring level set $L_{y_\tau}$ 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 $\tau$ is to find the relative position between trisection points $(x_\tau, y_\tau)$ and the “reference” location $\theta^*$, after which the algorithm either moves $a_\tau$ to $x_\tau$ or $b_\tau$ to $y_\tau$, effectively shrinking the length of the interval $[a_\tau, b_\tau]$ that contains $\theta^*$ to its two thirds. Furthermore, to avoid a large cumulative regret, level set corresponding to the left endpoint $a_\tau$ is exploited in each time period within the epoch $\tau$ to offset potentially large regret incurred by exploring $y_\tau$.

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 $y_\tau$ and $F(y_\tau)$ is clear, or a pre-specified number of inner iterations for outer iteration $\tau$ 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 $L_{y_\tau}$.

After sufficiently many explorations of $L_{y_\tau}$, 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 $y_\tau$ must be to the right of $\theta^*$ (i.e., $y_\tau \geq \theta^*$) due to Lemma 53. Therefore, we will shrink the value of right boundary by setting $b_{\tau+1} \leftarrow y_\tau$. 

110
2. On the other hand, when \( u_t(y_r) \geq y_r \), we can conclude that \( x_r \text{ must be to the left of } \theta^* \) (i.e., \( x_r < \theta^* \)). We show this by contradiction. Assuming that \( x_r > \theta^* \), since \( y_r \) is always greater than \( x_r \) (and thus \( y_r > \theta^* \)) and the gap between \( y_r \) and \( F(y_r) \) is at least \( y_r - x_r \), the gap will be detected by the confidence bands and thus we will have \( u_t(y_r) < y_r \) with high probability. This leads to a contradiction.

Therefore, since \( x_r \) is to the left of \( \theta^* \), we should increase the value of the left boundary by setting \( a_{r+1} \leftarrow x_r \).

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 \( y_r - F(y_r) \geq y_r - F(x_r) \geq y_r - x_r \)

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
\[
\text{Regret}(\{S_i\}_{i=1}^T) = \mathbb{E} \sum_{i=1}^T R(S^*_i) - R(S_i) \leq C_1 \sqrt{T \log T}.
\] (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 \( X_1, \ldots, X_L \) be i.i.d. random variables with mean \( \mu \) and satisfy \( a \leq X_i \leq b \) almost surely for all \( \ell \in [L] \). For any \( \delta \in (0, 1] \), it holds that
\[
\Pr \left[ \forall \ell \in [L], \left| \frac{1}{\ell} \sum_{i=1}^{\ell} X_i - \mu \right| \leq \sqrt{\frac{2(b - a)^2 \ln(\delta/\ell))}{\ell}} \right] \geq 1 - L\delta.
\] (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 \( \ell \) increases (effectively \( \ell \delta \) in \( \sqrt{2 \ln(\delta/\ell) (b-a)^2} \) instead of \( \delta \)). This allows the confidence intervals to be much shorter for large \( \ell \).

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
  \[
  [\ell_t(\theta), u_t(\theta)] = \frac{\rho_t(\theta)}{t} \pm \sqrt{\frac{\ln(8/(\delta t))}{t}}.
  \] (4.7)

- Step 9 in Algorithm 8 is replaced with \( \text{EXPLORE}(y_T, t, 1/T) \); correspondingly, the number of inner iterations is changed to \( n_T = 8[(y_T - x_T)^{-2} \ln(8T(y_T - x_T)^2)] \).

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
\[
\text{Regret}([S_t]_{t=1}^T) = \mathbb{E} \sum_{t=1}^T R(S^*) - R(S_t) \leq C_2 \sqrt{T}.
\] (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 $\pi$, under the “uncapacitated setting” in which there is no capacity constraint imposed on the provided assortments.

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 $\pi$,
\[
\sup_{v_1, \cdots, v_N \geq 0} \text{Regret}([S_t]_{t=1}^T) \geq \frac{\sqrt{T}}{384}.
\] (4.9)

Theorem 15 shows that our regret upper bounds in Theorems 13 and 14 are tight up to an $O(\log T)$ 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 $|S_t| \leq K$ for all $t$, for some pre-specified capacity limit $K \leq N$. 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 $O(\sqrt{NT})$, 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 $K \leq N/4$. There exists an absolute constant $C \geq 10^{-3}$ independent of $N$, $T$ and $K$ such that for all policy $\pi$,
\[
\sup_{v_1, \cdots, v_N \geq 0} \text{Regret}([S_t]_{t=1}^T) \geq C \cdot \min\{\sqrt{NT}, T\}.
\] (4.10)

Remark 26. When the revenue parameters $\{r_i\}_{i=1}^N$ are uniformly bounded (i.e., $r_i \leq 1$ 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” $K \leq N/4$ 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 $\gamma \geq 1/2$ and we leave it as an interesting technical open problem. We remark that for capacitated problems the $K \leq N/4$ 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 $[0.4, 0.5]$. 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 $S = \{i \in \mathcal{N} : r_i \geq 0.42\}$. For the single assortment $S = \{i\}$, because the preference parameter for each item is rather small ($v_i \leq 20/N$), 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 $S = \{i \in \mathcal{N} : r_i \geq 0.42\}$, 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 $\{i \in \mathcal{N} : r_i \geq r^*\}$ 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.

<table>
<thead>
<tr>
<th>$(N,T)$</th>
<th>UCB mean</th>
<th>UCB max</th>
<th>THOMPSON mean</th>
<th>THOMPSON max</th>
<th>GRS mean</th>
<th>GRS max</th>
<th>TRISEC. mean</th>
<th>TRISEC. max</th>
<th>ADAP-TRISEC. mean</th>
<th>ADAP-TRISEC. max</th>
</tr>
</thead>
<tbody>
<tr>
<td>(100,500)</td>
<td>34.9</td>
<td>38.1</td>
<td>1.28</td>
<td>2.97</td>
<td>10.9</td>
<td>22.4</td>
<td>7.68</td>
<td>7.68</td>
<td>1.99</td>
<td>1.99</td>
</tr>
<tr>
<td>(250,500)</td>
<td>54.3</td>
<td>56.2</td>
<td>2.81</td>
<td>4.95</td>
<td>7.93</td>
<td>34.2</td>
<td>7.57</td>
<td>7.57</td>
<td>2.23</td>
<td>2.23</td>
</tr>
<tr>
<td>(500,500)</td>
<td>73.4</td>
<td>75.5</td>
<td>4.90</td>
<td>4.95</td>
<td>7.02</td>
<td>43.4</td>
<td>7.43</td>
<td>7.43</td>
<td>2.23</td>
<td>2.23</td>
</tr>
<tr>
<td>(1000,500)</td>
<td>90.3</td>
<td>93.5</td>
<td>8.17</td>
<td>10.7</td>
<td>5.34</td>
<td>45.1</td>
<td>7.44</td>
<td>7.44</td>
<td>2.25</td>
<td>2.25</td>
</tr>
<tr>
<td>(100,1000)</td>
<td>73.1</td>
<td>78.2</td>
<td>1.36</td>
<td>2.79</td>
<td>139.9</td>
<td>175.0</td>
<td>8.69</td>
<td>8.69</td>
<td>3.90</td>
<td>3.90</td>
</tr>
<tr>
<td>(250,1000)</td>
<td>113.7</td>
<td>119.3</td>
<td>3.36</td>
<td>5.17</td>
<td>90.1</td>
<td>110.1</td>
<td>8.69</td>
<td>8.69</td>
<td>4.13</td>
<td>4.14</td>
</tr>
<tr>
<td>(500,1000)</td>
<td>136.8</td>
<td>140.3</td>
<td>5.65</td>
<td>7.64</td>
<td>65.7</td>
<td>113.9</td>
<td>9.38</td>
<td>9.38</td>
<td>3.80</td>
<td>3.80</td>
</tr>
<tr>
<td>(1000,1000)</td>
<td>160.8</td>
<td>165.4</td>
<td>9.31</td>
<td>12.4</td>
<td>8.43</td>
<td>22.8</td>
<td>9.77</td>
<td>9.77</td>
<td>3.97</td>
<td>3.97</td>
</tr>
</tbody>
</table>

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 $\sqrt{\frac{2\ln(8/\delta)}{T}}$ confidence interval configuration with $\sqrt{\frac{0.1\ln(8/\delta)}{T}}$. 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 $\beta = \min_{j \neq p} |R(L_{r_j}) - R(L_{r_p})|$, 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.
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 \(S_i = 2^N\) 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 \(\gamma_i\) is a measure of the degree of independence among the items in nest \(i\): a larger value of \(\gamma_i\) 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 \(S_i(t) \in S_i\) 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 \(t\).
Probabilistic model for the purchasing option \((i_t, j_t)\) can be formulated below:

\[
\Pr[i_t = i | S^{(t)}] = \frac{V_i(S^{(t)})^{\gamma_i}}{V_0 + \sum_{i' = 1}^{M} V_{i'}(S^{(t)})^{\gamma_{i'}}}, \quad \text{where } V_0 \equiv 1, \quad V_i(S^{(t)}) = \sum_{j \in S^{(t)}_i} v_{ij} \quad \text{for } i \in [M];
\]

\[
\Pr[j_t = j | i_t = i, S^{(t)}] = \frac{v_{ij}}{\sum_{j' \in S^{(t)}_i} v_{ij'}} \quad \text{for } i \in [M], \quad j \in S^{(t)}_i.
\]

(4.11)

(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

\[
R(S^{(t)}) = \sum_{i=1}^{M} \Pr[i_t = i | S^{(t)}] \sum_{j \in S^{(t)}_i} r_{ij} \Pr[j_t = j | i_t = i, S^{(t)}]
\]

\[
= \sum_{i=1}^{M} R_i(S^{(t)}) V_i(S^{(t)})^{\gamma_i} \frac{1}{1 + \sum_{i=1}^{M} V_i(S^{(t)})^{\gamma_i}}, \quad \text{where } R_i(S^{(t)}) = \frac{\sum_{j \in S^{(t)}_i} r_{ij} v_{ij}}{\sum_{j \in S^{(t)}_i} v_{ij}}.
\]

(4.13)

The objective of the seller to minimize expected (accumulated) regret, which is defined as follows:

\[
\text{Regret}([S^{(t)}]_{t=1}^T) := \sum_{t=1}^{T} R^* - \mathbb{E}^\pi [R(S^{(t)})], \quad \text{where } R^* = \max_{S \in \mathcal{S}_1 \times \cdots \times \mathcal{S}_M} R(S).
\]

(4.14)

Throughout this section, we make the following boundedness assumptions on revenue and utility parameters:

(A1) \(0 \leq r_{ij} \leq 1\) for all \(i \in [M]\) and \(j \in [N]\).

(A2) \(0 < u_{ij} \leq C_V\) for all \(i \in [M]\) and \(j \in [N]\) with some constant \(C_V \geq 1\).

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) \(\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 \times \cdots \times \mathcal{S}_M\) 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 \(\mathcal{S}\) independently would easily incur a regret also exponentially large. It is thus mandatory to reduce the number of candidate assortment sets in \(\mathcal{S}\).

Fortunately, existing results on the structure of optimal \(S\) show that it suffices to consider level sets \(\mathcal{L}_i(\theta_i) := \{j \in [N] : r_{ij} \geq \theta_i\}\) 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 $\theta_i \geq 0$. Define $\mathbb{P}_i := \{\mathcal{L}_i(\theta_i) : \theta_i \geq 0\} \subseteq \mathbb{S}_i$ to be all the possible level sets of $\mathbb{S}_i$ and let

$$\mathbb{P} := \mathbb{P}_1 \times \mathbb{P}_2 \times \cdots \times \mathbb{P}_M \subseteq \mathbb{S}. \quad (4.15)$$

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 $S^* = (\mathcal{L}_1(\theta_1^*), \ldots, \mathcal{L}_M(\theta_M^*)) \in \mathbb{P}$ such that the following hold:

1. $R(S^*) = \max_{S \in \mathbb{S}} R(S) = R^*$;
2. $\theta_i^* \geq \gamma_i R^* + (1 - \gamma_i) R_i(S_i^*)$ 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 $\mathcal{L}_i(\infty) = \emptyset$ 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 $\theta$ 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 $S = (S_1, \ldots, S_M) \in \mathbb{P}$ can be equivalently written as

$$\theta = (\theta_1, \ldots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M. \quad (4.16)$$

Here $\theta_i \in \mathcal{K}_i$ corresponds to level set $S_i = \mathcal{L}_i(\theta_i) = \{j \in [N] : r_{ij} \geq \theta_i\}$ 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 $\theta_i$.

After presenting the customer with assortment combination $S \in \mathbb{P}$ (or equivalently parameterized by $\theta$), 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 \( r_{i_s,j_t} \) of item \( j_t \) in nest \( i_s \) (if \( i_s = 0 \), then \( r_t = 0 \) almost surely). In the nested singleton model, we discard the item choice \( j_t \) within nest \( i_s \), and only record the nest-level selection \( i_s \) 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 \( i_s \in [M] \cup \{0\} \) and corresponding random revenue \( r_t \in [0, 1] \) satisfy the following:

\[
\Pr[i_s = i|\theta^{(t)}] = \frac{u_{i,\theta_i}}{1 + \sum_{i' = 1}^{M} u_{i',\theta_{i'}}}; \quad \mathbb{E}[r_t|i_s = i] = \phi_{i,\theta_i}; \quad r_t = 0 \text{ a.s. if } i_s = 0. \tag{4.17}
\]

In fact, Eq. (4.17) resembles the classical plain MNL model, with two important differences. First, the revenues collected on each purchased singleton \( i_s \) 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,

\[
R'(\theta^{(t)}) := \sum_{i = 1}^{M} \Pr[i_s = i|\theta^{(t)}] \cdot \mathbb{E}[r_t|i_s = i] = \frac{\sum_{i = 1}^{M} \phi_{i,\theta_i} u_{i,\theta_i}}{1 + \sum_{i = 1}^{M} u_{i,\theta_i}}.
\]

The objective is to minimize the regret:

\[
\text{Regret}(\{\theta^{(t)}\}_{t=1}^{T}) := \mathbb{E}\sum_{t=1}^{T} R'(\theta^*) - R'(\theta^{(t)}) \quad \text{where} \quad R'(\theta^*) = \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} R'(\theta). \tag{4.18}
\]

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 \( (NC_V)^{\gamma_i} \leq NC_V \), since \( \gamma_i \in [0, 1] \) and \( C_V \geq 1 \).

For a given dynamic assortment selection policy \( \pi' \) under this nested singleton model, it is easy to construct a policy \( \pi \) 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 \( \pi' \) that attains a regret of at most \( \Delta \) 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 \( U \leq NC_V \) for all \( i \in [M] \) and \( \theta_i \in \mathcal{K}_i \). Then there is a meta-policy \( \pi \) (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 \( \Delta \).

Proposition 16 is a simple consequence of Lemma 56 and we omit its proof.
Algorithm 10 The meta-policy $\pi$ built upon policy $\pi'$ under the nested singleton model.

1: **Input:** a dynamic assortment planning policy $\pi'$ under the nested singleton model.
2: **Output:** a dynamic assortment planning policy $\pi$ 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 $\pi'$;
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 $\pi'$;
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 $\tau$ where the same assortment combination $\theta$ 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 $i$; $\mathcal{T}(i, \theta) = |\mathcal{T}(i, \theta)|$ denotes the cardinality of $\mathcal{T}(i, \theta)$;
- $\mathcal{n}_{i, \tau}$: the number of iterations in the epoch $\tau$ (i.e., $\mathcal{E}_\tau$) in which an item in nest $i$ is purchased;
- $\mathcal{r}_{i, \tau}$: the total revenue collected for all iterations in $\mathcal{E}_\tau$ in which an item in nest $i$ is purchased;
- $\tilde{u}_{i, \theta}, \tilde{\phi}_{i, \theta}, \overline{u}_{i, \theta}, \overline{\phi}_{i, \theta}$: 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 $\theta$ 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 $\tilde{\theta}_i \in \mathcal{K}_i$ be the singleton provided in nest $i$. The expectations of the number of iterations and total revenues collected in which nest
Simple calculations show that (see for example Corollary A.1 of Agrawal et al. (2017a))

\[ \Pr [\hat{n}_{i,\tau} = k] = \left( \frac{u_{i,\hat{\theta}_i}}{1 + u_{i,\hat{\theta}_i}} \right)^k \left( \frac{1}{1 + u_{i,\hat{\theta}_i}} \right) \quad \text{for} \ k = 0, 1, 2, \cdots \quad (4.19) \]

That is, \( \hat{n}_{i,\tau} \) is a geometric random variable with parameter \( 1/(1 + u_{i,\hat{\theta}_i}) \). Hence, \( \hat{n}_{i,\tau} \) is an unbiased estimator of \( u_{i,\hat{\theta}_i} \), meaning that \( \mathbb{E} [\hat{n}_{i,\tau}] = u_{i,\tau} \).
The distribution and expectation of $\hat{r}_{i,\tau}$ can be similarly derived, using the property that $\mathbb{E}[\hat{r}_{i,\tau}|i_t = i] = \phi_{i,\hat{\theta}_i}$.

The above properties motivate intuitive parameter estimators $\hat{u}_{i,\theta}, \hat{\phi}_{i,\theta}$ of $u_{i,\theta}$ and $\phi_{i,\theta}$ for $\theta = \hat{\theta}_i$, which are taken to be the sample averages of $\hat{u}_{i,\tau}$ and $\hat{r}_{i,\tau}$ over all epochs $\mathcal{E}_\tau$ in which the item $\hat{\theta}_i$ 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 $\hat{u}_{i,\tau}$ and $\hat{r}_{i,\tau}$ are independent of $\hat{\theta}_i$ for $i' \neq i$. This key independence property enables us to combine purchasing information of vastly different assortment combinations (provided that $\hat{\theta}_i$ remains the same), which forms an important aggregation strategy that avoids exponentially large regret if assortment combinations are treated independently.

**Efficient computation of $\hat{\theta}$**

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 $\theta$ takes $\mathcal{O}(K^M)$ time and is computationally intractable even for small $M$ values. In this section we introduce a computationally efficient procedure to compute $\hat{\theta}$ 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 $\theta = (\theta_1, \ldots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ define potential function

$$\psi_\lambda(\theta) := \sum_{i=1}^M (\bar{\phi}_{i,\theta_i} - \lambda) \bar{u}_{i,\theta_i}.$$  \hfill (4.20)

Recall the definition of $\mathcal{R}(\theta) = \frac{\sum_{i=1}^M \bar{u}_{i,\theta_i} \bar{\pi}_{i,\theta_i}}{1 + \sum_{i=1}^M \bar{u}_{i,\theta_i}}$ in Step 5 of Algorithm 11. The following lemma characterizes the properties of $\psi_\lambda(\theta)$ and its relationship with $\mathcal{R}^* = \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} \mathcal{R}(\theta)$:

**Lemma 58.** The following holds for all $\lambda \in [0, 1]$:

1. If $\mathcal{R}^* \geq \lambda$, then there exists a $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ such that $\psi_\lambda(\theta) \geq \lambda$; furthermore if $\mathcal{R}^* > \lambda$, then the inequality is strict;

2. If $\mathcal{R}^* \leq \lambda$, then for all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$, $\psi_\lambda(\theta) \leq \lambda$; furthermore if $\mathcal{R}^* < \lambda$, then the inequalities are strict.

**Proof.** Let $\theta^* = (\theta_1^*, \ldots, \theta_M^*) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ be a maximizer of $\mathcal{R}$ (i.e., $\mathcal{R}^* = \mathcal{R}(\theta^*)$). By definition, $\sum_{i=1}^M (\bar{\phi}_{i,\theta_i} - \mathcal{R}^*) \bar{u}_{i,\theta_i} = \mathcal{R}^*$. If $\mathcal{R}^* \geq \lambda$, then $\sum_{i=1}^M (\bar{\phi}_{i,\theta_i} - \lambda) \bar{u}_{i,\theta_i} \geq \sum_{i=1}^M (\bar{\phi}_{i,\theta_i} - \mathcal{R}^*) \bar{u}_{i,\theta_i} = \mathcal{R}^* \geq \lambda$. Therefore $\psi_\lambda(\theta^*) \geq \lambda$. Furthermore, if $\mathcal{R}^* > \lambda$ 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 $\theta = (\theta_1, \ldots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$ such that $\psi_\lambda(\theta) > \lambda$, meaning that $\sum_{i=1}^M (\bar{\phi}_{i,\theta_i} - \lambda) \bar{u}_{i,\theta_i} > \lambda$. Rearranging terms and dividing both sides by $(1 + \sum_{i=1}^M \bar{u}_{i,\theta_i})$ we have $\mathcal{R}(\theta) = \frac{\sum_{i=1}^M \bar{\phi}_{i,\theta_i} \bar{u}_{i,\theta_i}}{1 + \sum_{i=1}^M \bar{u}_{i,\theta_i}} > \lambda$. This contradicts the assumption that $\mathcal{R}^* = \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} \mathcal{R}(\theta) \leq \lambda$. To prove the second half of the second property, simply replace all occurrences of $>$ by $\geq$. \hfill $\square$
Based on Lemma 58, an efficient optimization algorithm computing the maximizer \( \hat{\theta}(\epsilon) \) can be designed by a binary search over \( \lambda \in [0, 1] \). In particular, for each fixed value of \( \lambda \), the \( \theta^*(\lambda) = (\theta^*_1(\lambda), \ldots, \theta^*_M(\lambda)) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M \) that maximizes \( \psi_\lambda(\theta) \) can be found by setting \( \theta^*_i(\lambda) \in \arg \max_{\theta_i \in \mathcal{K}_i} (\varphi_{i, \theta_i} - \lambda \pi_{i, \theta_i}) \). If \( \psi_\lambda(\theta^*(\lambda)) > \lambda \), then \( \overline{R}^* > \lambda \), because otherwise it violates the second property in Lemma 58. Similarly, if \( \psi_\lambda(\theta^*(\lambda)) \leq \lambda \), then \( \overline{R}^* \leq \lambda \), because otherwise it violates the second part of the first property in Lemma 58 (note that since \( \theta^*(\lambda) \) is the maximizer of \( \psi_\lambda(\theta) \), \( \psi_\lambda(\theta^*(\lambda)) \leq \lambda \) implies that \( \psi_\lambda(\theta) \leq \lambda \) for all \( \theta \)). Thus, whether \( \overline{R}^* > \lambda \) or \( \overline{R}^* \leq \lambda \) can be determined by solely comparing \( \psi_\lambda(\theta^*(\lambda)) \) with \( \lambda \).

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 \( \epsilon \). 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

\[
\text{Regret}(\{\theta(t)\}_{t=1}^T) \leq \sqrt{MKT \log(MKT)} + MKU \log^2(MKT) + O(1),
\]

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

\[
\text{Regret}(\{S(t)\}_{t=1}^T) \leq \sqrt{MNT \log(MNT)} + MN^2 C_V \log^2(MNT) + O(1) = \tilde{O}(\sqrt{MNT} + MN^2)
\]

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 \( \tilde{O}(\sqrt{MNT}) \). 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 \( \tilde{O}(MN^2) \). 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 $\mathcal{L}_i(\theta) = \{j \in [N] : r_{ij} \geq \theta\}$ and we only consider level set threshold parameters $\theta$ that are multiples of $1/\delta$. Let $\mathbb{N}$ be the set of non-negative integers and define

$$\tilde{\mathcal{K}}^\delta_i := \{\theta : 0 \leq \theta \leq 1, \ \theta/\delta \in \mathbb{N}, \ \mathcal{L}_i(\theta)'s \text{ are distinct} \} \cup \{\infty\}, \quad \text{for } i \in [M]$$

where each $\theta \in \tilde{\mathcal{K}}^\delta_i$ corresponds to a unique level set $\mathcal{L}_i(\theta)$. When there are multiple $\theta$’s leading to the same level set, we keep any one of these $\theta$’s in $\tilde{\mathcal{K}}^\delta_i$ and thus the level sets induced by $\tilde{\mathcal{K}}^\delta_i$ (i.e., $\mathcal{L}_i(\theta) : \theta \in \tilde{\mathcal{K}}^\delta_i$) are unique. Since duplicate assortment sets are removed in $\tilde{\mathcal{K}}^\delta_i$, we have $\tilde{\mathcal{K}}^\delta_i \subseteq \mathcal{K}_i$, and thus $|\tilde{\mathcal{K}}^\delta_i| \leq |\mathcal{K}_i| = K = N + 1$. On the other hand, we also have $|\tilde{\mathcal{K}}^\delta_i| \leq \lceil 1/\delta \rceil + 2$ because level set thresholds in $\tilde{\mathcal{K}}^\delta_i$ must be an integer multiple of $\delta$. On one hand, when $\delta$ is not too small, the size of $\tilde{\mathcal{K}}^\delta_i$ could be significantly smaller than $\mathcal{N}$ 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 $\tilde{\mathcal{K}}^\delta_i := \mathcal{K}_i$ when $\delta = 0$.

The following key discretized reduction lemma shows that by restricting ourselves to $\tilde{\mathcal{K}}^\delta_i$ instead of $\mathcal{K}_i$, the approximation error in terms of expected revenue can be upper bounded by $\delta$, which goes to zero as we take $\delta \to 0$.

**Lemma 59 (Discretized reduction lemma).** Fix an arbitrary $\delta \in (0, 1)$. Then

$$\max_{\theta \in \tilde{\mathcal{K}}^\delta_i \times \cdots \times \tilde{\mathcal{K}}^\delta_M} R'(\theta) - \max_{\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M} R'(\theta) \leq \delta,$$

where $R'(\theta) := \left[\sum_{i=1}^M \phi_i, \theta, u_i, \theta_i\right]/\left[1 + \sum_{i=1}^M u_{i, \theta_i}\right]$.

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 $|\tilde{\mathcal{K}}^\delta_i| \leq \min\{N, \lceil 1/\delta \rceil + 2\}$ for all $i \in [M]$ and $\delta \in [0, 1]$.

**Proposition 17.** Suppose there exists a policy $\pi'$ that attains a regret of at most $\Delta$ on any instance of the nested singleton model with $|\tilde{\mathcal{K}}^\delta_i| \leq \min\{N, \lceil 1/\delta \rceil + 2\}$ for all $i \in [M]$ and $u_{i, \theta_i} \in [0, U]$ with $U \leq NC_V$ for all $i \in [M]$ and $\theta_i \in \tilde{\mathcal{K}}^\delta_i$. Then there is a meta-policy $\pi$ (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 $\delta T$ comes from the loss of the discretization in Lemma 59.

Now for the nested singleton model, we invoke Algorithm 11 with the discretized singletons $\tilde{\mathcal{K}}^\delta_1, \cdots, \tilde{\mathcal{K}}^\delta_M$ as input to construct the policy $\pi'$. Then we obtain a class of policies parameterized
Algorithm 12. The meta-policy $\pi$ built upon policy $\pi'$ and the discretization argument under the nested singleton model.

1: **Input:** a dynamic assortment planning policy $\pi'$ under the nested singleton model, discretization granularity $\delta$.
2: **Output:** a dynamic assortment planning policy $\pi$ for the original nested logit model.
3: Construct discretized singleton sets $\tilde{\mathcal{K}}_1^\delta, \cdots, \tilde{\mathcal{K}}_M^\delta$ in Eq. (4.23);
4: for $t = 1, 2, \cdots, T$ do
5: Let $\theta(t) = (\theta_1(t), \cdots, \theta_M(t)) \in \tilde{\mathcal{K}}_1^\delta \times \cdots \times \tilde{\mathcal{K}}_M^\delta$ be the output of $\pi'$;
6: Produce assortment combination $S(t) = (S_1(t), \cdots, S_M(t))$, where $S_i(t) = L_i(\theta_i(t))$;
7: Observe the nest-selection $i_t \in [M] \cup \{0\}$ and revenue $r_t$, and pass $i_t, r_t$ to policy $\pi'$.
8: end for

by $\delta$. By replacing $K = |\mathcal{K}|$ in Corollary 5 with $K = |\tilde{\mathcal{K}}_i^\delta| \leq \min\{N, [1/\delta] + 2\}$, 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 $\delta$ is used to convert Algorithm 11 into a dynamic assortment planning algorithm for the original nested model, then

$$\text{Regret}\{S(t)\}_{t=1}^T \lesssim \sqrt{\min\{N, \delta^{-1}\}MT \log(MNT)}$$

$$+ \min\{N, \delta^{-1}\} MNC \log^2(MNT) + \delta T + O(1). \quad (4.24)$$

For example, by choosing $\delta \approx T^{-1/3}$, we have $\text{Regret}\{S(t)\}_{t=1}^T = \tilde{O}(\sqrt{MT^{2/3} + MNT^{1/3}})$.

**Remark 28.** In cases with many items per nest (i.e., $N$ is large), a small value of $\delta$ 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 $\delta$ 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 $\tilde{\mathcal{K}}_i^0 = \mathcal{K}_i$ 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 $\delta$ to be a non-zero value. In particular, by setting $\delta \approx T^{-1/3}$, we obtain a regret upper bound of $\tilde{O}(\sqrt{MT^{2/3} + MNT^{1/3}})$, 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 $\tilde{O}(M N^2)$.)

### 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
for any dynamic assortment planning policy $\pi$,

$$\sup_{\{r_{ij}, v_{ij}\}} \sum_{t=1}^{T} R^* - \mathbb{E}[R(S^{(t)})] \geq C_0 \sqrt{MT} \quad \text{where} \quad R^* = \max_{S \in \mathbb{S}} R(S). \quad (4.25)$$

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 $\delta$ (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.

![Figure 4.3: Accumulated regret of our policy with $M = 5$ nests, varying the number of items per nest $N$ and the granularity parameter $\delta$.](image)
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.

<table>
<thead>
<tr>
<th>$(M, N)$</th>
<th>$\delta = 0$</th>
<th>$\delta = 10^{-3}$</th>
<th>$\delta = 5 \times 10^{-3}$</th>
<th>$\delta = 10^{-2}$</th>
<th>$\delta = 5 \times 10^{-2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MED</td>
<td>MAX</td>
<td>MED</td>
<td>MAX</td>
<td>MED</td>
</tr>
<tr>
<td>$T = 100$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5,100)</td>
<td>5.5</td>
<td>6.4</td>
<td>5.5</td>
<td>6.0</td>
<td>3.8</td>
</tr>
<tr>
<td>(10,100)</td>
<td>4.8</td>
<td>6.2</td>
<td>5.4</td>
<td>5.5</td>
<td>4.7</td>
</tr>
<tr>
<td>(5,250)</td>
<td>10.4</td>
<td>14.1</td>
<td>9.8</td>
<td>12.0</td>
<td>5.7</td>
</tr>
<tr>
<td>(10,250)</td>
<td>10.8</td>
<td>12.0</td>
<td>9.7</td>
<td>12.3</td>
<td>5.5</td>
</tr>
<tr>
<td>(5,1000)</td>
<td>22.0</td>
<td>25.3</td>
<td>16.0</td>
<td>18.2</td>
<td>6.2</td>
</tr>
<tr>
<td>(10,1000)</td>
<td>21.5</td>
<td>24.1</td>
<td>15.1</td>
<td>17.7</td>
<td>5.1</td>
</tr>
<tr>
<td>$T = 500$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5,100)</td>
<td><strong>14.3</strong></td>
<td><strong>18.5</strong></td>
<td>18.3</td>
<td>22.6</td>
<td>26.8</td>
</tr>
<tr>
<td>(10,100)</td>
<td><strong>15.7</strong></td>
<td>23.0</td>
<td>16.5</td>
<td><strong>22.1</strong></td>
<td>28.4</td>
</tr>
<tr>
<td>(5,250)</td>
<td>14.2</td>
<td>17.3</td>
<td><strong>12.7</strong></td>
<td><strong>14.9</strong></td>
<td>16.4</td>
</tr>
<tr>
<td>(10,250)</td>
<td>13.8</td>
<td><strong>15.9</strong></td>
<td><strong>13.0</strong></td>
<td>17.4</td>
<td>16.6</td>
</tr>
<tr>
<td>(5,1000)</td>
<td>41.1</td>
<td>46.1</td>
<td>22.7</td>
<td>25.7</td>
<td><strong>14.1</strong></td>
</tr>
<tr>
<td>(10,1000)</td>
<td>39.3</td>
<td>44.2</td>
<td>21.0</td>
<td>27.2</td>
<td><strong>13.7</strong></td>
</tr>
<tr>
<td>$T = 10000$:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5,100)</td>
<td>491.5</td>
<td>505.5</td>
<td><strong>489.4</strong></td>
<td><strong>496.5</strong></td>
<td>494.5</td>
</tr>
<tr>
<td>(10,100)</td>
<td>548.4</td>
<td>558.0</td>
<td>548.6</td>
<td>552.9</td>
<td><strong>529.3</strong></td>
</tr>
<tr>
<td>(5,250)</td>
<td>534.4</td>
<td>543.7</td>
<td>529.7</td>
<td>543.9</td>
<td>523.4</td>
</tr>
<tr>
<td>(10,250)</td>
<td>551.0</td>
<td>560.5</td>
<td>554.5</td>
<td>563.3</td>
<td><strong>547.4</strong></td>
</tr>
<tr>
<td>(5,1000)</td>
<td>669.0</td>
<td>704.4</td>
<td>570.5</td>
<td>584.8</td>
<td>538.8</td>
</tr>
<tr>
<td>(10,1000)</td>
<td>703.5</td>
<td>738.2</td>
<td>613.1</td>
<td>633.6</td>
<td>555.7</td>
</tr>
</tbody>
</table>
In Table 4.2, we compare the accumulated regret of our proposed policies with different granularity parameters $\delta$, 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 $v_t$, 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 $S_t \subseteq [N]$ 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 $r_{it}$ 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 $v_{t0} = 0$)

$$p_{\theta_0,t}(j|S) = \frac{\exp\{v_{tj}^T \theta_0\}}{1 + \sum_{k \in S} \exp\{v_{tk}^T \theta_0\}} \quad \forall j \in S \cup \{0\}. \quad (4.26)$$

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

$$R_t(S) := \mathbb{E}_{\theta,t}[r_{tj}|S] = \frac{\sum_{j \in S} r_{tj} \exp\{v_{tj}^T \theta_0\}}{1 + \sum_{j \in S} \exp\{v_{tj}^T \theta_0\}}. \quad (4.27)$$

Note that throughout the paper, we use $\mathbb{E}_{\theta,t}[\cdot | S]$ 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 $\pi$ such that the regret

$$\text{Regret}(\{S_t\}_{t=1}^T) = \mathbb{E}^\pi \sum_{t=1}^T R_t(S^*_t) - R_t(S_t) \quad \text{where } S^*_t = \arg\max_{S \subseteq [N], |S| \leq K} R_t(S) \quad (4.28)$$

is minimized. Here, $S^*_t$ is an optimal assortment chosen when the full knowledge of choice probabilities is available (i.e., $\theta_0$ is known).
in this phase, which facilitates us to adapt existing analysis in the works of Filippi et al. (2010); only one item is present in the assortment. In the second phase, we use a UCB-type approach. We propose an MLE-UCB policy, described in Algorithm 13.

Algorithm 13: The MLE-UCB policy for dynamic assortment optimization

1: Input: Number of pure explorations $T_0$, constraint radius $\tau$.
2: Output: Assortment selections $\{S_t\}_{t=1}^{T_0} \subseteq [N]$ satisfying $|S_t| \leq K$.
3: Pure exploration: for $t = 1, \ldots, T_0$, pick $S_t = \{\ell_t\}$ for a single product $\ell_t$ sampled uniformly at random from $\{1, \ldots, N\}$ and record purchasing actions $(i_1, \ldots, i_{T_0})$;
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 $r_{tj}$ and preference features $v_{tj}$ at time $t$;
7: Compute local MLE $\hat{\theta}_{t-1} \in \arg \max_{\theta | S_t} \sum_{t'=1}^{t-1} \log p_{\theta,t}(i_{t'} | S_{t'})$;
8: For every assortment $S \subseteq [N], |S| \leq K$, compute its upper confidence bound
   \[
   \hat{\mathcal{R}}_t(S) := \mathbb{E}_{\hat{\theta}_{t-1,t}}[r_{tj}|S] + \min \left\{1, \omega \sqrt{\frac{1}{L_{t-1}^2} (\hat{\theta}_{t-1}) M_t(\hat{\theta}_{t-1}|S) \hat{L}_{t-1}^{1/2} (\hat{\theta}_{t-1}) \|_{op}} \right\};
   \]
   \[
   \hat{L}_{t-1}(\theta) := \sum_{t'=1}^{t-1} \hat{M}_{t'}(\theta|S_t'); \quad \hat{M}_t(\theta|S) := \mathbb{E}_{\theta,t}[v_{tj}v_{tj}^T|S] - \{\mathbb{E}_{\theta,t}[v_{tj}|S]\}^T; \quad \omega = \sqrt{d \log (\rho \nu TK)};
   \]
9: Pick $S_t \in \arg \max_{S \subseteq [N], |S| \leq K} \hat{\mathcal{R}}_t(S)$ and observe purchasing action $i_t \in S_t \cup \{0\}$;
10: end for

Remark: the expectations admit the following closed-form expressions:
\[
\mathbb{E}_{\theta,t}[r_{tj}|S] = \sum_{j \in S} p_{\theta,t}(j|S)r_{tj} = \frac{\sum_{j \in S} r_{tj} \exp[v_{tj}^\top \theta]}{1 + \sum_{j \in S} \exp[v_{tj}^\top \theta]};
\]
\[
\mathbb{E}_{\theta,t}[v_{tj}|S] = \sum_{j \in S} p_{\theta,t}(j|S)v_{tj} = \frac{\sum_{j \in S} v_{tj} \exp[v_{tj}^\top \theta]}{1 + \sum_{j \in S} \exp[v_{tj}^\top \theta]};
\]
\[
\mathbb{E}_{\theta,t}[v_{tj}v_{tj}^\top|S] = \sum_{j \in S} p_{\theta,t}(j|S)v_{tj}v_{tj}^\top = \frac{\sum_{j \in S} v_{tj}^\top v_{tj} \exp[v_{tj}^\top \theta]}{1 + \sum_{j \in S} \exp[v_{tj}^\top \theta]}.
\]

### 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 $\theta_0$, i.e., a good initial estimator for $\theta_0$. 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 $\hat{\mathcal{R}}_t(S_t)$ of the expected revenue $R_t(S_t)$. Such upper bounds are built using a local Maximum Likelihood Estimation (MLE) of $\theta_0$. In particular, in Step 5, instead of computing an MLE, we compute a local MLE, where the estimator $\hat{\theta}_{t-1}$ lies in a ball centered at the pilot estimator $\theta^*$ with a radius $\tau$. This localization also simplifies the technical analysis based on Taylor expansion, which benefits from the constraint
that $\hat{\theta}_{t-1}$ is not too far away from $\theta^*$. To construct the confidence bound, we introduce the matrices $\hat{M}_t(\hat{\theta}_{t-1}|S)$ and $\hat{I}_{t-1}(\hat{\theta}_{t-1})$ in Step 6 of Algorithm 13, which are empirical estimates of the Fisher’s information matrices $-\mathbb{E}[\nabla^2_\theta \log p(\cdot|\theta)]$ 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 $\tau_0$ in the local MLE formulation. While theoretical values of $\omega, T_0$ and $\tau$ 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 $\nu$ such that $\|v_{tj}\|_2 \leq \nu$ for all $t$ and $j$. Moreover, for all $t \leq T_0$ and $j \in [N]$, $v_{tj}$ are i.i.d. generated from an unknown distribution with the density $\mu$ satisfying that $\lambda_{\min}(\mathbb{E}_\mu v v^\top) \geq \lambda_0$ 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 $|S| \leq K$, $\frac{p_{\theta_0}(S|\mathcal{F}_t)}{p_{\theta_0}(\emptyset|\mathcal{F}_t)} \leq \rho$ 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 $t \leq T_0$ 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 $\|\theta_0\|_2 \leq C$, then the boundedness assumption in (A1) implies (A2) with $\rho \leq e^{2 \max\{1, C\}}$.

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 $T_0 = \max\{b^2 d \log T/\lambda_0^2, \rho^2 (d+\log T)/(\tau^2 \lambda_0)\}$ and $\tau = 1/\sqrt{\rho^2 \nu^2 K^2}$ in Algorithm 13, then the regret of the MLE-UCB policy is upper bounded by

$$C_1 \left[ d\sqrt{T} \cdot \log(\lambda_0^{-1} \rho \nu TK) + d^2 \lambda_0^{-2} \rho^4 \nu^2 K^2 \log T \right] + C_2,$$

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, $d\sqrt{T} \cdot \log(\lambda_0^{-1} \rho \nu TK)$, is the main regret term that scales as $\tilde{O}(d\sqrt{T})$ 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., 2020).
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 $\pi$, there is a worst-case problem instance with $N = K \cdot 2^d$ items and uniformly bounded feature and coefficient vector (i.e., $\|v_{ti}\|_2 \leq 1$ and $\|\theta_0\|_2 \leq 1$ for all $i \in [N], t \in [T]$) such that the regret of $\pi$ is lower bounded by $C_2 \cdot d \sqrt{T}$.

Theorem 20 essentially implies that the $\tilde{O}(d \sqrt{T})$ 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 $\{r_{ij}\}$ for $j \in [N]$ are independently and identically generated from the uniform distribution $[0.5, 0.8]$. For the feature vectors $\{v_{ij}\}$, each of them is independently generated as a uniform random vector $v$ such that $\|v\|_2 = 2$ and $v^T \theta_0 < -0.6$. Here we set an upper bound of $-0.6$ for the inner product so that the utility parameters $u_{ij} = \exp\{v_{ij}^T \theta_0\}$ 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 $T_0 = \lceil \sqrt{\bar{T}} \rceil$ 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.
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.
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 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 $d \geq 2$. 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., $|\mu(i) - \mu(j)| \geq L|i-j|$ 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 $\tilde{O}(T^{2/3})$ 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 $s < F^* < s'$, which then implies Lemma 52.

We first prove $s < F^*$. Assume by contradiction that $F^* > s$. Clearly $s \neq 0$ because $F^* > 0$. By definition of $F$ and $F^*$, we have

$$F^* = F(s') = \frac{\sum_{r_i \geq s} r_i v_i}{1 + \sum_{r_i \geq s} v_i} \implies \sum_{r_i \geq s} (r_i - F^*) v_i = F^*. \tag{4.30}$$

Because $F^* < s$, adding we have that

$$\sum_{r_i \geq s} (r_i - F^*) v_i \geq F^* \implies F(s) \geq F^*. \tag{4.31}$$

This contradicts with the fact that $F(s) \neq F(s^+)$ and that $F^*$ is the maximum value of $F$.

We next prove $F^* < s'$. Assume by contradiction that $F^* > s'$. Removing all items corresponding to $r_i = s'$ in Eq. (4.30), we have

$$\sum_{r_i > s} (r_i - F^*) v_i \geq F^* \implies F(s^+) \geq F^*. \tag{4.32}$$

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 $F(\theta) \leq \theta$ for all $\theta \geq \theta^*$. In addition, for any $\theta \geq \theta^*$, by definition of $F$ we have

$$F(\theta) - F(\theta^+) = R(\{i \in N : r_i \geq \theta\}) - R(\{i \in N : r_i > \theta\}) \tag{4.33}$$

$$= \frac{\sum_{r_i \geq \theta} r_i v_i}{1 + \sum_{r_i \geq \theta} v_i} - \frac{\sum_{r_i > \theta} r_i v_i}{1 + \sum_{r_i > \theta} v_i} \tag{4.34}$$

135
By Eq. (4.37), we have

\[
(1 + \sum_{r_i > \theta} v_i) \left( \sum_{r_i > \theta} r_i v_i \right) - \left(1 + \sum_{r_i > \theta} v_i \right) \left( \sum_{r_i > \theta} r_i v_i \right) = \frac{\sum_{r_i = \theta} v_i}{1 + \sum_{r_i > \theta} v_i} \left[ \theta - F(\theta^+) \right].
\]  

(4.35)

\[
= \frac{1 + \sum_{r_i > \theta} v_i} {1 + \sum_{r_i > \theta} v_i} \left( \sum_{r_i > \theta} r_i v_i \right) - \left( \sum_{r_i = \theta} v_i \right) \left( \sum_{r_i > \theta} r_i v_i \right) = \frac{\sum_{r_i = \theta} v_i}{1 + \sum_{r_i > \theta} v_i} \left[ \theta - F(\theta^+) \right].
\]  

(4.36)

\[
= \frac{\sum_{r_i = \theta} v_i}{1 + \sum_{r_i > \theta} v_i} \left[ \theta - F(\theta^+) \right].
\]  

(4.37)

Because \( \theta \geq F(\theta) \) holds for all \( \theta \geq \theta^* \), we conclude that \( \theta \geq F(\theta^+) \) also holds for all \( \theta \geq \theta^* \). Subsequently, the right-hand side of Eq. (4.37) is non-negative and therefore \( F(\theta) \geq F(\theta^+) \).

### 4.5.3 Proof of Lemma 54

If \( F(\theta) \equiv F^* \) for all \( \theta \leq \theta^* \) 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 \( \theta^* \). Formally, we let \( 0 < s_1 < s_2 < \cdots < s_t < \theta^* \) be all jumping points that are strictly smaller than \( \theta^* \). To prove Lemma 54, it suffices to show that \( F(s_j) \geq s_j \) and \( F(s_j) \geq F(s_j^+) \) 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 \( F(s_t) \leq F_s = F(t) \). In addition, because \( s_t \leq \theta^* = F^* = F(s_t^+) \), invoking Eq. (4.37) we have that \( F(s_t) \leq F(s_t^+) \). 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, \( F(s_{j+1}) \geq s_{j+1} \geq s_j \). Also, \( F(s_j^+) = F(s_{j+1}) \) because there is no jump points between \( s_j \) and \( s_{j+1} \), and subsequently \( F(s_j^+) \geq s_j \). Invoking Eq. (4.37) we proved \( F(s_j) \leq F(s_j^+) \).

To prove \( F(s_j) \geq s_j \), define \( \gamma_j := (\sum_{r_i = s_j} v_i)/(1 + \sum_{r_i > s_j} v_i) \). It is clear that \( 0 \leq \gamma_j \leq 1 \). By Eq. (4.37), we have

\[
F(s_j) - s_j = F(s_j) - F(s_j^+) + F(s_j^+) - s_j = \gamma_j [s_j - F(s_j^+)] + F(s_j^+) - s_j = (1 - \gamma_j) [F(s_j^+) - s_j].
\]  

(4.38)

As we have already proved \( F(s_j^+) \geq s_j \), the right-hand side of the above inequality is non-negative and therefore \( F(s_j) \geq s_j \).

### 4.5.4 Proof of Theorem 13

We first state a simple lemma showing that the confidence bound \( \ell_t(y_t) \) and \( u_t(y_t) \) constructed in Algorithm 8 contains \( F(y_t) \) with high probability.

**Lemma 60.** With probability \( 1 - O(T^{-1}) \), \( \ell_t(\theta) \leq F(\theta) \leq u_t(\theta) \) 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 \( 0 \leq F(\theta) \leq 1 \) for all \( \theta \), we have

\[
\Pr \left[ F(\theta) \notin [\ell_t(\theta), u_t(\theta)] \right] = \Pr \left[ \left| \frac{\rho_t(\theta)}{t} - F(\theta) \right| > \sqrt{\frac{\ln(1/\delta)}{2t}} \right].
\]  

(4.41)
\[ \leq 2 \exp \left\{ -2t \cdot \frac{\ln(1/\delta)}{2t} \right\} \leq 2\delta = 2/T^2. \] (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 \( O(T^{-1}) \).

The following lemma, based on properties of the potential function \( F \) and Lemma 60, establishes that (with high probability) the shrinkage of \( a_{\tau} \) or \( b_{\tau} \) are “consistent”; i.e., \( \theta^* \) 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 \( 1 - O(T^{-1}) \), \( a_{\tau} \leq \theta^* \leq b_{\tau} \) for all \( \tau = 1, 2, \cdots, \tau_0 \), where \( \tau_0 \) 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 \( \tau \) using the distance between the trisection points \( x_{\tau} \) and \( y_{\tau} \).

**Lemma 62.** For \( \tau = 0, 1, \cdots \) let \( T(\tau) \) denote the set of all indices of inner iterations at outer iteration \( \tau \). Conditioned on the success events in Lemmas 60 and 61, it holds that

\[ \mathbb{E} \sum_{t \in T(\tau)} R(S_t^*) - R(S_t) \leq \varepsilon_{\tau}^{-1} \log T. \] (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

\[ \mathbb{E} \sum_{t=1}^{T} R(S_t^*) - R(S_t) \leq \sum_{\tau=0}^{\tau_0} \varepsilon_{\tau}^{-1} \log T \leq \varepsilon_{\tau_0}^{-1} \log T, \] (4.44)

where \( \tau_0 \) is the total number of outer iterations executed by Algorithm 8. On the other hand, because at each outer iteration \( \tau \) the revenue level \( a_{\tau} \) is exploited for exactly \( n_{\tau} = 16\lceil y_{\tau} - x_{\tau} \rceil^2 \ln(T^2) \rceil \) times, we have

\[ T \geq n_{\tau_0} \geq \varepsilon_{\tau_0}^{-2} \log T. \] (4.45)

Combining Eqs. (4.44) and (4.45) we conclude that \( \text{Regret}(\{S_t\}_{t=1}^{T}) \leq \sqrt{T \log T}. \)

**Proof of Lemma 61** We use induction to prove this lemma. We also conditioned on the fact that \( \ell_t(x_{\tau}) \leq F(x_{\tau}) \leq u_t(x_{\tau}) \) and \( \ell_t(y_{\tau}) \leq F(y_{\tau}) \leq u_t(y_{\tau}) \) for all \( t \) and \( \tau \), 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 \( \theta \geq 0 \) it holds that \( 0 \leq F(\theta) \leq F^* \leq 1 \). Therefore, \( 0 \leq \theta^* \leq 1 \) and hence \( a_{\tau} \leq \theta^* \leq b_{\tau} \) for \( \tau = 0 \).

We next prove the lemma for outer iteration \( \tau \), assuming the lemma holds for outer iteration \( \tau - 1 \) (i.e., \( a_{\tau-1} \leq r^* \leq b_{\tau-1} \)). According to the trisection parameter update step in Algorithm 8, the proof can be divided into two cases:
Lemma 53: iterations can be upper bounded by $\tau$ of $u_1$. Regret from exploring $a$ can then be upper bounded by $\tau$ of $u_1$. Invoking Lemma 60 we then have iteration in Algorithm 8 at outer iteration $\tau$. Case 1: $u_1(\tau)$. This lemma upper bounds the expected regret incurred at each outer iteration. We now establish that $u_1(\tau)$ at outer iteration $\tau$. We analyze the regret incurred at outer iteration $\tau$ from exploration of $y_\tau$ and exploitation of $a_\tau$ separately.

1. Regret from exploring $y_\tau$: suppose the level set $L_{y_\tau} (\mathcal{N})$ is explored for $m_\tau \leq n_\tau$ times at outer iteration $\tau$. Then we have $u_{m_\tau} (y_\tau) \geq y_\tau$. 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 $\ell_{m_\tau} (y_\tau) \leq F(y_\tau) \leq u_{m_\tau} (y_\tau)$ and $|u_{m_\tau} (y_\tau) - \ell_{m_\tau} (y_\tau)| \leq 2 \sqrt{\frac{\ln(T^2)}{2m_\tau}}$. Subsequently,

$$F(y_\tau) \geq \ell_{m_\tau} (y_\tau) \geq u_{m_\tau} (y_\tau) - 2 \sqrt{\frac{\ln(T^2)}{2m_\tau}} \geq y_\tau - 2 \sqrt{\frac{\ln(T^2)}{2m_\tau}}. \quad (4.49)$$

Note also that $y_\tau \geq a_\tau \geq \theta^* - 3\varepsilon_\tau = F^* - 3\varepsilon_\tau$; we have

$$F^* - F(y_\tau) \leq 3\varepsilon_\tau + 2 \sqrt{\frac{\ln(T)}{m_\tau}}. \quad (4.50)$$

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 $y_\tau$ (corresponding to the
assortment selection $L_{y_r}$) once. As the exploration is carried out for $m_\tau$ times, the total regret for all exploration steps at revenue level $x_\tau$ can be upper bounded by

$$m_\tau \left[ 3\varepsilon_\tau + 2\sqrt{\ln T \over m_\tau} \right] \leq 3m_\tau\varepsilon_\tau + \sqrt{4m_\tau \ln T} \leq 3n_\tau\varepsilon_\tau + \sqrt{4n_\tau \ln T} \leq \varepsilon_\tau^{-1} \log T. \quad (4.51)$$

Here the last inequality holds because $n_\tau \leq 16\varepsilon_\tau^{-2} \ln(T^2)$.

2. Regret from exploiting $a_\tau$: 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 $\varepsilon_\tau$. 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 $a_\tau$ (corresponding to the assortment selection $L_{a_\tau}$) can be upper bounded by $\varepsilon_\tau$. Because the revenue level $a_\tau$ is exploited for $n_\tau$ times and $n_\tau \leq 16\varepsilon_\tau^{-2} \ln(T^2)$, the total regret of exploitation of $a_\tau$ at outer iteration $\tau$ can be upper bounded by

$$n_\tau \cdot 3\varepsilon_\tau \leq \varepsilon_\tau^{-1} \log T. \quad (4.54)$$

### 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 $k := \lceil \log_2 L \rceil$. For each $\ell \in \{1, 2, 4, \cdots, 2^k\}$, by standard Hoeffding's inequality (Lemma 89), we have

$$\Pr \left[ \left| {1 \over \ell} \sum_{i=1}^{\ell} X_i - \mu \right| \leq \sqrt{\ln[8/(\delta \ell)] \over 2\ell} \right] \geq 1 - {\delta \ell \over 4}. \quad (4.55)$$

Subsequently, by union bound and the fact that $1 + 2 + 4 + \cdots + 2^k \leq 2^{k+1} \leq 2L$, we have

$$\Pr \left[ \forall \ell = 1, 2, 4, \cdots, 2^k, \left| {1 \over \ell} \sum_{i=1}^{\ell} X_i - \mu \right| \leq \sqrt{\ln[8/(\delta \ell)] \over 2\ell} \right] \geq 1 - {\delta L \over 2}. \quad (4.56)$$

Next consider any $\ell \in \{1, 2, 4, \cdots, 2^k\}$. By Hoeffding’s maximal inequality (Lemma 93), we have

$$\Pr \left[ \forall i \leq \min\{\ell, n-\ell\}, |X_{\ell+1} + \cdots + X_{\ell+i} - i \cdot \mu| \leq \sqrt{\ell \ln[8/(\delta \ell)] \over 2} \right] \geq 1 - {\delta \ell \over 4}. \quad (4.57)$$

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

$$\Pr \left[ \forall \ell = 1, 2, \cdots, 2^k, i \leq \min\{\ell, n-\ell\}, |X_{\ell+1} + \cdots + X_{\ell+i} - i \cdot \mu| \leq \sqrt{\ell \ln[8/(\delta \ell)] \over 2} \right] \geq 1 - {\delta L \over 2}. \quad (4.58)$$
Combining Eqs. (4.55,4.56), we have with probability $1-\delta L$ uniformly over all $\ell = 1,2,4,\cdots,2^k$ and $i \leq \min\{\ell,n-\ell\}$ that
\[
|X_1 + \cdots + X_\ell + X_{\ell+1} + \cdots + X_{\ell+i} - (\ell+i)\mu| \leq \sqrt{2\ell \ln[8/(\delta \ell)]}.
\]
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 $\tau$, and $n_\tau = 8[\varepsilon_\tau^2 \ln(8T\varepsilon_\tau^2)]$ 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 $E_1(\tau)$: $\theta^* < a_\tau < b_\tau$;
- Event $E_2(\tau)$: $a_\tau \leq \theta^* \leq b_\tau$;
- Event $E_3(\tau)$: $a_\tau < b_\tau < \theta^*$.

Let $\tau_0 \in \mathbb{N}$ be the last outer iteration in Algorithm 8. Let also $T(\tau) \subseteq [T]$ be the indices of inner iterations in outer iteration $\tau$, satisfying $|T(\tau)| \leq 2n_\tau$ almost surely. For $\omega \in \{1,2,3\}$, $\tau \in \mathbb{N}$ and $\alpha,\beta \in \mathbb{R}^+$, define
\[
\psi_O^\omega(\alpha,\beta) := \mathbb{E} \left[ \sum_{\tau'=\tau}^{\tau_0} \sum_{\tau \in T(\tau')} R(S') - R(S_t) \left| E_\omega(\tau), |a_\tau - \theta^*| = \alpha, |F(a_\tau) - a_\tau| = \beta \right. \right]. \quad (4.57)
\]

Intuitively, $\psi_O^\omega(\alpha,\beta)$ is the expected regret Algorithm 8 incurs for outer iterations $\tau,\tau+1,\cdots,\tau_0$, conditioned on the event $E_\omega(\tau)$ and other boundary conditions at the left margin $a_\tau$.

The following three lemmas are the central steps in our proof, which establish recurrence relationships among $\psi_O^\omega(\alpha,\beta)$, for $\omega \in \{1,2,3\}$. The proofs are technically involved and given later. To simplify notations, we write $a_n \leq b_n$ or $b_n \geq a_n$ if there exists a universal constant $C > 0$ such that $|a_n| \leq C|b_n|$ for all $n \in \mathbb{N}$.

**Lemma 64 (Regret in Case 1).** $\psi_1^0(\alpha,\beta) \leq \beta T + \sum_{\tau'=\tau}^{\tau_0} \sup_{\Delta > \varepsilon_\tau} \Delta T \exp\{-n_\tau \Delta^2\} + O(\varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2))$.

**Lemma 64 (Regret in Case 2).** $\psi_2^0(\alpha,\beta) \leq O(\varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2)) + \psi_{\tau+1}^2(\alpha',\beta') + \psi_{\tau+1}^3(\alpha',\beta') \cdot O(\log(T\varepsilon_\tau^2)/(T\varepsilon_\tau^2)) + \sup_{\Delta > \varepsilon_\tau} \psi_{\tau+1}^1(\alpha',\beta'(\Delta)) \exp\{-n_\tau \Delta^2\}$ for parameters $\alpha'_1,\beta'_1(\Delta),\alpha'_2,\beta'_2,\alpha'_3,\beta'_3$ that satisfy $\beta'_1(\Delta) \leq \Delta$ and $\alpha'_3 \leq 3\varepsilon_\tau$.

**Lemma 65 (Regret in Case 3).** $\psi_3^0(\alpha,\beta) \leq \alpha T$.

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 $\psi_1^0(\alpha,\beta)$ using Lemma 63. First note that $\Delta \rightarrow \Delta \exp\{-n_\tau \Delta^2\}$ attains its maximum on $\Delta > 0$ at $\Delta = \sqrt{1/2n_\tau}$. Also note that $n_\tau = [8\varepsilon_\tau^{-2} \ln(8T\varepsilon_\tau^2)]$ and therefore $\sqrt{1/2n_\tau} \leq \varepsilon_\tau$. Subsequently,
\[
\sum_{\tau'=\tau}^{\tau_0} \sup_{\Delta > \varepsilon_\tau} \Delta T \exp\{-n_\tau \Delta^2\} \leq \sum_{\tau'=\tau}^{\tau_0} \varepsilon_\tau T \exp\{-n_\tau \varepsilon_\tau^2\} \leq \sum_{\tau'=\tau}^{\tau_0} \varepsilon_\tau T \exp\{-\ln(T\varepsilon_\tau^2)\}
\]
\[ \sum_{t'=\tau}^{T_{\epsilon\tau}} \varepsilon_{t'}^{-1} = O(\varepsilon_{\tau_0}^{-1}), \]  
\hspace{1cm} (4.58)

where the last asymptotic holds because \( \{\varepsilon_t\} \) forms a geometric series. Subsequently,

\[ \psi_1(\alpha, \beta) \leq \beta T + \sum_{t'=\tau}^{T_{\epsilon\tau}} O(\varepsilon_{t'}^{-1} \log(T\varepsilon_{t'}^2)). \]  
\hspace{1cm} (4.59)

It remains the bound the summation term on the right-hand side of the above inequality. Denote \( s_{t'} = \varepsilon_{t'}^{-1} \ln(T\varepsilon_{t'}^2) = \rho^{-t'} \ln(T\rho^{t'}), \) where \( \rho = 2/3 \). We then have \( s_{t'} = \rho^{\tau_0-t'} [1 + \ln \rho^{-2(\tau_0-t')} ] s_{\tau_0} \leq 2(\tau_0 - t' + 1) \rho^{\tau_0-t'} \ln(1/\rho) \) for all \( \tau' \leq \tau_0 \). Subsequently,

\[ \sum_{t'=\tau}^{T_{\epsilon\tau}} s_{t'} \leq \sum_{t'=0}^{\tau_0} 2(\tau_0 - t' + 1) \rho^{\tau_0-t'} \ln(1/\rho) \cdot s_{\tau_0} \leq O(1) \cdot s_{\tau_0}. \]  
\hspace{1cm} (4.60)

Therefore,

\[ \psi_1(\alpha, \beta) \leq \beta T + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)). \]  
\hspace{1cm} (4.61)

We are now ready to derive the final regret upper bound by analyzing \( \psi_2(\alpha, \beta) \), because the event \( \mathcal{E}_2(0) \) always holds since \( 0 \leq \theta^* \leq 1 \). Applying Lemma 64 with Lemma 65 and Eq. (4.61), we have for all \( \tau \in \{0, 1, \cdots, \tau_0\} \) that

\[
\psi_2(\alpha, \beta) \leq \psi_2(\alpha_2', \beta_2') + O(\varepsilon_{\tau}^{-1} \log(T\varepsilon_{\tau}^2)) + O(\varepsilon_{\tau} T) \cdot \frac{\ln(T\varepsilon_{\tau}^2)}{T\varepsilon_{\tau}^2}
\]

\[ + \sup_{\Delta \geq \varepsilon_{\tau}} \left( \Delta T + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)) \right) \exp\{-n_{\tau} \Delta^2\} \]

\[ \leq \psi_2(\alpha_2', \beta_2') + O(\varepsilon_{\tau}^{-1} \log(T\varepsilon_{\tau}^2)) + \sup_{\Delta \geq \varepsilon_{\tau}} \Delta T \exp\{-n_{\tau} \Delta^2\}
\]

\[ + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)) \cdot \exp\{-n_{\tau} \varepsilon_{\tau}^2\}. \]  
\hspace{1cm} (4.62)

Using the same analysis as in Eq. (4.58), we know \( \sup_{\Delta \geq \varepsilon_{\tau}} \Delta T \exp\{-n_{\tau} \Delta^2\} \leq O(\varepsilon_{\tau}^{-1}) \) and \( \exp\{-n_{\tau} \varepsilon_{\tau}^2\} \leq 1/(T\varepsilon_{\tau}^2) \). Subsequently, summing all terms \( \tau = 0, 1, \cdots, \tau_0 \) together we have

\[ \psi_2(\alpha, \beta) \leq \sum_{\tau=0}^{\tau_0} O(\varepsilon_{\tau}^{-1} \log(T\varepsilon_{\tau}^2)) + O(\varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2)) \cdot \frac{1}{T\varepsilon_{\tau}^2}
\]

\[ \leq \varepsilon_{\tau_0}^{-1} \log(T\varepsilon_{\tau_0}^2) \cdot (1 + 1/(T\varepsilon_{\tau_0}^2)). \]  
\hspace{1cm} (4.63)

Finally, note that \( n_{\tau_0} \geq \varepsilon_{\tau_0}^{-2} \) and \( n_{\tau_0} \leq T \), implying that \( \varepsilon_{\tau_0} \geq \sqrt{1/T} \). Plugging the lower bound on \( \varepsilon_{\tau_0} \) into the above inequality we have \( \psi_2(\alpha, \beta) \leq \sqrt{T} \), which completes the proof of Theorem 14. \( \Box \)

**Proof of Lemma 64** First analyze the expected regret incurred at outer iteration \( \tau \). by exploiting the left end-point \( a_{\tau} \) (corresponding to assortment \( \mathcal{L}_{a_{\tau}} \)) for \( n_{\tau} \) iterations. Also, because \( a_{\tau} \leq \theta^* \leq b_{\tau} \) conditioned on \( \mathcal{E}_2(\tau) \), by Lemmas 52 and 54 we have \( F(a_{\tau}) \geq a_{\tau} \geq \theta^* - |b_{\tau} - a_{\tau}| = F(\theta^*) - |b_{\tau} - a_{\tau}| \geq R(S^*) - 3\varepsilon_{\tau} \). Subsequently,

\[ \text{Regret by exploiting } \mathcal{L}_{a_{\tau}} : \leq 3\varepsilon_{\tau} \cdot n_{\tau} \leq \varepsilon_{\tau}^{-1} \log(T\varepsilon_{\tau}). \]  
\hspace{1cm} (4.64)
Next we analyze the expected regret incurred at outer iteration $\tau$ by exploring the right trisection point $y_\tau$ (corresponding to assortment $L_{y_\tau}$). This is done by a case analysis. If $y_\tau \leq \theta^*$, then the regret incurred by exploiting $L_{y_\tau}$ at outer iteration $\tau$ is again upper bounded (up to numerical constants) by $\epsilon^{-1} \log(T\epsilon^2)$, 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 $\Delta_\tau \geq 0$, and also by Lemma 52, each exploration of $L_{y_\tau}$ incurs a regret of no more than $\Delta_\tau$. Let $m_\tau$ be the number of times $L_{y_\tau}$ is explored at outer iteration $\tau$. By definition of the stopping rule in Algorithm 8, we have

$$\Pr [m_\tau \geq \ell] \leq \Pr \left[ \frac{\rho_\ell}{\ell} + \sqrt{\frac{2\ln(8T/\ell)}{\ell}} \geq y_\tau \right]$$

$$= \Pr \left[ \frac{\rho_\ell}{\ell} - F(y_\tau) \geq \Delta_\tau - \sqrt{\frac{2\ln(8T/\ell)}{\ell}} \right]. \quad (4.65)$$

Because $\rho_\ell$ is a sum of $\ell$ 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

$$\Pr [m_\tau \geq \ell] \leq \exp \left\{ -2 \left( \sqrt{\ell} \Delta_\tau - \sqrt{2\ln(8T/\ell)} \right)^2 \right\}$$

$$\lesssim \begin{cases} 1, & \text{if } \Delta_\tau \leq \sqrt{8\ln(8T/\ell)/\ell}; \\ \exp\{-\ell \Delta_\tau^2/2\}, & \text{otherwise.} \end{cases}$$

Subsequently,

**Regret by exploring $L_{y_\tau}$:**

$$\leq \sum_{\ell=1}^{n_\tau} \ell \Delta_\tau \Pr [m_\tau = \ell] \leq \sum_{\ell=1}^{n_\tau} \Delta_\tau \Pr [m_\tau \geq \ell]$$

$$\lesssim \sum_{\ell=1}^{\ell_0-1} \sqrt{\frac{\ln(T/\ell)}{\ell}} + \sum_{\ell=\ell_0}^{n_\tau} \Delta_\tau \exp\{-\ell \Delta_\tau^2/2\} \quad (4.66)$$

$$\lesssim \sqrt{\ell_0 \ln(T/\ell_0)} + \sup_{\Delta > \sqrt{8\ln(8T/\ell_0)/\ell_0}} \Delta \cdot \sum_{\ell=\ell_0}^{\infty} \exp\{-\ell \Delta^2/2\} \quad (4.67)$$

$$\lesssim \sqrt{\ell_0 \ln(T/\ell_0)} + \frac{\Delta}{1 - \exp\{-\Delta^2/2\}} \frac{\Delta \exp\{-\ell_0 \Delta^2/2\}}{1 - \exp\{-\Delta^2/2\}}$$

$$\lesssim \sqrt{\ell_0 \ln(T/\ell_0)} + \frac{\Delta}{1 - \exp\{-\Delta^2/2\}} \frac{\Delta \exp\{-\Delta^2/2\}}{1 - \exp\{-4\ln(8T/\ell_0)\}} \quad (4.68)$$

$$\lesssim \sqrt{\ell_0 \ln(T/\ell_0)}. \quad (4.69)$$

Here in Eq. (4.66), $\ell_0$ is the smallest positive integer not exceeding $n_\tau$ such that $\Delta_\tau > \sqrt{8\ln(8T/\ell_0)/\ell_0}$. (If $\Delta_\tau \leq \sqrt{8\ln(8T/\ell_0)/\ell_0}$ holds for all $1 \leq \ell_0 \leq n_\tau$, 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

\[
\sum_{\ell=1}^{\ell_0} \sqrt{\frac{\ln(T/\ell)}{\ell}} \leq \sum_{j=1}^{[\log_2 \ell_0]} 2^j \sqrt{\frac{\ln(T/2^j)}{2^j}} = \sum_{j=1}^{[\log_2 \ell_0]} \sqrt{2^j \ln(T/2^j)} \leq \ell_0 \ln(T/\ell_0);
\]

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 \( \ell_0 \leq n_\tau \) and \( n_\tau \leq \varepsilon_\tau^{-2} \log(T\varepsilon_\tau^2) \geq \varepsilon_\tau^{-2} \), we have

\[
\text{Regret by exploring } \mathcal{L}_{y_\tau} \leq \sqrt{n_\tau \ln(T/n_\tau)} \leq \varepsilon_\tau^{-1} \log(T\varepsilon_\tau^2). \quad (4.70)
\]

Finally, we consider regret incurred at later outer iterations \( \tau' = \tau + 1, \ldots, \tau_0 \). This is done by another case analysis on the relative location of \( \theta^* \) with respect to \( a_{\tau+1} \) and \( b_{\tau+1} \):

- \( \mathcal{E}_2(\tau + 1) \): \( a_{\tau+1} \leq \theta^* \leq b_{\tau+1} \): the additional regret is upper bounded by \( \psi_{\tau+1}^2(\alpha_1', \beta_1') \) for some values of \( \alpha_1', \beta_1' \) 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 \( \beta_2 \leq \Delta_\tau = y_\tau - F(y_\tau) \) and the value of \( \alpha_2' \) 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 \( \alpha_3' \leq \Delta_\tau \) and the value of \( \beta_3' \) 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 \( \ell \in T(\tau) \), the exploration step fails to detect \( F(y_\tau) \) below \( y_\tau \), meaning that \( \frac{\rho_\tau}{\ell} + \sqrt{\frac{2\ln(8T/\ell)}{\ell}} > y_\tau \) for all \( \ell \in \{1, \ldots, n_\tau\} \). Also note that because \( \theta^* < a_{\tau+1} = x_\tau = y_\tau - \varepsilon_\tau \), by Lemma 53 we know that \( \Delta_\tau = y_\tau - F(y_\tau) \geq \varepsilon_\tau \). Using Hoeffding’s inequality, we have

\[
\Pr[\mathcal{E}_1(\tau + 1)] \leq \Pr \left[ \forall \ell, \frac{\rho_\tau}{\ell} - F(y_\tau) > \Delta_\tau - \sqrt{\frac{2\ln(8T/\ell)}{\ell}} \right] \\
\leq \Pr \left[ \frac{\rho_\tau}{n_\tau} - F(y_\tau) > \Delta_\tau - \sqrt{\frac{2\ln(8T/n_\tau)}{n_\tau}} \right] \\
\leq \exp \left\{ -2 \left( \sqrt{n_\tau \Delta_\tau} - \sqrt{2 \ln(8T/n_\tau)} \right)^2 \right\} \\
\leq \exp \left\{ -n_\tau \Delta_\tau^2 \right\}. \quad (4.71)
\]

Here Eq. (4.71) holds because \( \sqrt{n_\tau \Delta_\tau} \geq \sqrt{n_\tau \varepsilon_\tau} \geq \sqrt{8 \ln(8T\varepsilon_\tau^2)} \geq 2 \sqrt{2 \ln(8T/n_\tau)} \) by the choice of \( n_\tau \).

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 \( \ell \in \{1, \ldots, n_\tau\} \), meaning that \( \frac{\rho_\tau}{\ell} + \sqrt{\frac{2\ln(8T/\ell)}{\ell}} < y_\tau \). Note that because \( b_{\tau+1} = y_\tau < \theta^* \), by Lemma 54, we know \( F(y_\tau) \geq y_\tau \). By Lemma 55,

\[
\Pr[\mathcal{E}_3(\tau + 1)] = \Pr \left[ \exists \ell, \frac{\rho_\tau}{\ell} < y_\tau - \sqrt{\frac{2\ln(8T/\ell)}{\ell}} \right] \\
\leq \Pr \left[ \exists \ell, \left| \frac{\rho_\tau}{\ell} - F(y_\tau) \right| > \sqrt{\frac{2\ln(8T/\ell)}{\ell}} \right]
\]

143
2. For each $\tau$, each $t$ priori. We then consider two constructions of the unknown utility parameters: the revenue parameters $\psi_\tau$, conditioned on $\theta$. Fix $\alpha, \beta$ and explore $L_{y, \tau}$ for $\tau \geq \tau$, and the regret from exploring $L_{a, \tau}$.

For any $\tau' \in \{\tau, \tau + 1, \cdots, \tau_0\}$, the expected regret from exploring $L_{y, \tau}$ 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 $\alpha$ and $\beta$. This corresponds to the $O(\varepsilon_{\tau'}^{-1} \log(T \varepsilon_{\tau'}^2))$ term in Lemma 63.

We next upper bound the expected regret incurred by exploring $L_{a, \tau}$ for all $\tau' = \tau, \tau + 1, \cdots, \tau_0$. Because $a_\tau - F(a_\tau) = \beta$ by the definition of $\psi_\tau(\alpha, \beta)$, the expected regret incurred by exploring $L_{a, \tau}$ for $\tau' \in \{\tau, \tau + 1, \cdots, \tau_0\}$ is at most $\beta T$ 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 $\mathcal{W} = \{\tau'_1, \tau'_2, \cdots, \tau'_t\}$ 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 $\Delta_{\tau', \tau} T$, where $\Delta_{\tau', \tau} = y_{\tau'} - F(y_{\tau'}) \geq \varepsilon_{\tau'}$;
2. For each $\tau' \in \{\tau, \tau + 1, \cdots, \tau_0\}$, the probability update $a_{\tau + 1} \leftarrow x_{\tau'}$ is applied 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 $L_{a, \tau}$ at subsequent iterations $\tau' \geq \tau$ can be upper bounded by $\beta T + \sum_{\tau' = \tau}^{\tau_0} \sup_{\Delta \geq \varepsilon_{\tau'}} \Delta T \exp\{-n_{\tau'} \Delta^2\}$.

Proof of Lemma 65 Because $a_\tau = \theta^* - \alpha < \theta^*$, by Lemma 54 we have $F(a_\tau) \geq a_\tau = \theta^* - \alpha = F(\theta^*) - \alpha$. Subsequently, $F(S^*) - F(a_\tau) \leq \alpha$ thanks to Lemma 52. Also note that conditioned on $E_\tau(\tau)$, the revenue levels explored or exploited at each time epoch $t \in T(\tau')$, $\tau \leq \tau' \leq \tau_0$ are sandwiched between $a_\tau$ and $\theta^*$, and therefore $R(S^*) - R(S_t) \leq \alpha$. Hence, $\psi_\tau^3(\alpha, \beta) \leq \alpha \cdot \mathbb{E} \sum_{\tau' = \tau}^{\tau_0} |T(\tau')| \leq \alpha T$.

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 $\{v_i\}_{i=1}^N$:

$$P_0 : \quad v_1 = 1 - 1/4 \sqrt{T}, \quad v_2 = 1, \quad v_3 = \cdots = v_N = 0;$$
$$P_1 : \quad v_1 = 1 + 1/4 \sqrt{T}, \quad v_2 = 1, \quad v_3 = \cdots = v_N = 0.$$

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 $j \in \{0, 1\}$. 

$$\frac{n_{\tau}}{T} \leq \frac{\ln(T \varepsilon_{\tau}^2)}{T \varepsilon_{\tau}^2}.$$  (4.72)
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 $\hat{\psi} \in \{0, 1\}$ whose inputs are $T$ random purchasing actions $i_1, \ldots, i_T$, it holds that $\max_{j \in \{0, 1\}} P_j[\hat{\psi} \neq j] \geq 1/3$.

On the other hand, the following lemma shows that, if the policy $\pi$ can achieve a small regret under both $P_0$ and $P_1$, then one can construct an estimator based on $\pi$ 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 $\pi$ satisfies $\text{Regret}(\{S_t\}_{t=1}^T) < \sqrt{T}/384$ for both $P_0$ and $P_1$. Then there exists an estimator $\hat{\psi} \in \{0, 1\}$ such that $P_j[\hat{\psi} \neq j] \leq 1/4$ for both $j = 0$ and $j = 1$.

Lemma 67 is proved by explicitly constructing a classifier (tester) $\hat{\psi}$ from any sequence of low regret. In particular, for any assortment sequence $\{S_t\}_{t=1}^T$, we construct $\hat{\psi}$ as $\hat{\psi} = 0$ if $rac{1}{T} \sum_{t=1}^T \mathbb{I}[1 \in S_t, 2 \notin S_t] \geq 1/2$ and $\hat{\psi} = 1$ otherwise. Using Markov’s inequality and the construction of $\{r_i, v_i\}$, it can be shown that if $\text{Regret}(\{S_t\}_{t=1}^T) > \sqrt{T}/384$ then $\hat{\psi}$ 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 $\text{KL}(P_0(S) \parallel P_1(S)) \leq 1/18T$.

**Proof of Lemma 68.** If $1 \notin S$ then $P_0(S) = P_1(S)$ and therefore $\text{KL}(P_0(S) \parallel P_1(S)) = 0$. In addition, because $v_i = r_i = 0$ for all $i \geq 3$, 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 $\text{KL}(P_0(\{1\}) \parallel P_1(\{1\}))$ and $\text{KL}(P_0(\{1, 2\}) \parallel P_1(\{1, 2\}))$.

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 $p_1, \ldots, p_J$ and $q_1, \ldots, q_J$, respectively. Denote also $\epsilon_j := p_j - q_j$. Then $\text{KL}(P \parallel Q) \leq \sum_{j=1}^J \epsilon_j^2 / q_j$.

We first consider $\text{KL}(P_0(\{1\}) \parallel P_1(\{1\}))$. By definition, $P_0(i = 1|\{1\}) \leq 1/2 - 1/24\sqrt{T}$ and $P_1[i = 2|\{2\}] \leq 1/2 + 1/24\sqrt{T}$. Also, $\min_{i=0,1} P_i(i|\{1\}) \geq 1/3$. Subsequently,

$$\text{KL}(P_0(\{1\}) \parallel P_1(\{1\})) \leq 2 \times \frac{1/144T}{1/3} \leq \frac{1}{24T} \leq \frac{1}{18T}. \quad (4.73)$$

We next consider $\text{KL}(P_0(\{1, 2\}) \parallel P_1(\{1, 2\}))$. 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]) \leq 1/3 - 1/48\sqrt{T}, P_1(i = 1[1, 2]) \geq 1/3 + 1/48\sqrt{T} and \min_{0 \leq i \leq 2}\{P_1(i[1, 2])\} \geq 1/4. Subsequently,

$$KL(P_0([1, 2])\|P_1([1, 2])) \leq 3 \times \frac{1/576T}{1/4} \leq \frac{1}{48T} \leq \frac{1}{18T}. \quad (4.74)$$

The lemma is thus proved. \qed

We are now ready to prove Lemma 66.

**Proof of Lemma 66.** Denote \(\|P - Q\|_{TV} := 2\sup_A |P(A) - Q(A)|\) 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

$$\|P_0^{\otimes T} - P_1^{\otimes T}\|_{TV} \leq \sqrt{2KL(P_0^{\otimes T}\|P_1^{\otimes T})} \leq \sup_{S^{(1)}, \ldots, S_T} \sqrt{2\prod_{t=1}^T KL(P_0(S_t)\|P_1(S_t))} \leq \sqrt{2T} \cdot \sup_S \sqrt{KL(P_0(S)\|P_1(S))} \leq \sqrt{2T} \cdot \sqrt{1/18T} \leq 1/3. \quad (4.76)$$

Using Le Cam’s inequality we have

$$\inf_{\tilde{\psi}_{j=0,1}} P_j[\tilde{\psi} \neq j] \geq \frac{1}{2} (1 - \|P_0^{\otimes T} - P_1^{\otimes T}\|_{TV}) \geq \frac{1}{2} > \frac{1}{3}; \quad (4.77)$$

\qed

**Proof of Lemma 67** Denote \(\varphi_0 := 1/T \cdot \sum_{t=1}^T \mathbb{I}[1 \in S_t, 2 \notin S_t], \varphi_1 := 1/T \cdot \sum_{t=1}^T \mathbb{I}[1, 2 \in S_t], \varphi_2 := 1/T \cdot \sum_{t=1}^T \mathbb{I}[2 \in S_t, 1 \notin S_t]\) and \(\overline{\varphi} := 1/T \cdot \sum_{t=1}^T \mathbb{I}[1, 2 \notin S_t]\). Because the four events partition the entire probability space, we have \(\varphi_0 + \varphi_1 + \varphi_2 + \overline{\varphi} = 1\). In addition, it is easy to verify that \(S^* = \{1\}\) under \(P_0\) and under \(P_1\). Subsequently,

$$\frac{\text{Regret}_n(T)}{T} \leq \frac{\varphi_0}{12\sqrt{T}} + \frac{\varphi_2 + \overline{\varphi}}{24} \quad \text{under } P_0;$$

$$\frac{\text{Regret}_n(T)}{T} \leq \frac{\varphi_1}{48\sqrt{T}} + \frac{\varphi_2 + \overline{\varphi}}{6} \quad \text{under } P_1.$$

Using Markov’s inequality and the fact that \(\text{Regret}_n(T) \leq \sqrt{T}/384\) under both \(P_0\) and \(P_1\), we have

$$P_0\left[\frac{\varphi_0}{12\sqrt{T}} + \frac{\varphi_2 + \overline{\varphi}}{24} > \frac{1}{96\sqrt{T}}\right] \leq \frac{1}{4} \quad \text{and} \quad P_1\left[\frac{\varphi_1}{48\sqrt{T}} + \frac{\varphi_2 + \overline{\varphi}}{6} > \frac{1}{96\sqrt{T}}\right] \leq \frac{1}{4}. \quad (4.78)$$

Subsequently, because \(\varphi_0 + \varphi_1 + \varphi_2 + \overline{\varphi} = 1\), we know that \(\varphi_0 > 1/2\) with probability \(\geq 2/3\) under \(P_0\) and \(\varphi_0 < 1/2\) with probability \(\geq 2/3\) under \(P_1\). Define \(\hat{\psi}\) as

$$\hat{\psi} := \begin{cases} 0 & \text{if } \varphi_0 \geq 1/2; \\ 1 & \text{if } \varphi_0 < 1/2. \end{cases} \quad (4.79)$$

The estimator \(\hat{\psi}\) 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 \( v_1, \ldots, v_N \in \{1/K, (1+\epsilon)/K\} \) for some parameter \( \epsilon \in (0, 1/2) \) to be specified later. For any subset \( S \subseteq [N] \), we use \( \theta_S \) 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 \( S_K \) to denote all subsets of \([N]\) of size \( K \); that is, \( S \in S_K \) implies \( |S| = K \). Clearly, \( |S_K| = \binom{N}{K} \). We use \( P_S \) and \( \mathbb{E}_S \) to denote the law and expectation under the parameterization \( \theta_S \).

The first step in our proof is to show that under problem parameter \( \theta_{S_0} \) for some fixed \( S_0 \in S_K \), any assortment selection \( \tilde{S}_t \in S_K \) that differs significantly from \( S_0 \) would incur a large one-stage regret. This is formalized in Lemma 69, which shows that, if a \( \delta \) proportion of items differ between \( S_0 \) and \( \tilde{S}_t \), then the assortment \( \tilde{S}_t \) 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 \( \tilde{N}_i \) in our proof.

At the second step we show, through a “neighboring argument” detailed in Eq. (4.83), the question of bounding \( \mathbb{E}[\tilde{N}_i] \) can be reduced to upper bounding the discrepancy between \( \mathbb{E}_S[\tilde{N}_i] \) and \( \mathbb{E}_{S'}[\tilde{N}_i] \) under two “neighboring” parameterizations \( \theta_S \) 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 \( \epsilon \) which scales with \( N, T \) and \( K \) (more specifically, \( \epsilon \) 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 \( \tilde{S}_t \in S_K \):

**Lemma 69.** Fix arbitrary \( S_0 \in S_K \) and let \( v \) be the parameter associated with \( \theta_{S_0} \); that is, \( v_i = (1+\epsilon)/K \) for \( i \in S_0 \) and \( v_i = 1/K \) for \( i \notin S_0 \), where \( \epsilon \in (0, 1/2) \). For any \( \tilde{S}_t \in S_K \) it holds that

\[
\max_{S \subseteq S_K} \{ R_v(S) \} - R_v(\tilde{S}_t) \geq \frac{\delta \epsilon}{9},
\]

where \( \delta = 1 - (|\tilde{S}_t \cap S_0|/K) \).

**Proof.** By construction of \( v \), it is clear that \( \max_{S \subseteq S_K} \{ R_v(S) \} = R_v(S_0) = (1+\epsilon)/(2+\epsilon) \). On the other hand, \( R_v(\tilde{S}_t) = (1 + (1-\delta)\epsilon)/(2 + (1-\delta)\epsilon) \). Subsequently,

\[
\max_{S \subseteq S_k} \{ R_v(S) \} - R_v(\tilde{S}_t) = \frac{1+\epsilon}{2+\epsilon} - \frac{1 + (1-\delta)\epsilon}{2 + (1-\delta)\epsilon} = \frac{\delta \epsilon}{(2+\epsilon)(2 + (1-\delta)\epsilon)} \geq \frac{\delta \epsilon}{9},
\]

where the last inequality holds because \( 0 < \epsilon < 1/2 \).

For each assortment selection \( S_t \subseteq [N], |S_t| \leq K \), let \( \tilde{S}_t \supseteq S_t \) be an arbitrary subset of size \( K \) that contains \( S_t \); that is, \( \tilde{S}_t \supseteq S_t, \tilde{S}_t \subseteq [N] \) and \( |\tilde{S}_t| = K \). For example, when \( |S_t| = K \)

147
one may directly set $\bar{S}_t = S_t$. Define $\bar{N}_i := \sum_{t=1}^{T} \mathbb{I}[i \in \bar{S}_t]$. Using Lemma 69 and the fact that $\{ \bar{S}_t \}_{t=1}^{T}$ suffers less regret than $\{ S_t \}_{t=1}^{T}$, we have

$$
\max_{S \in S_K} \mathbb{E}_S \left[ \sum_{t=1}^{T} R_v(S) - R_v(S_t) \right] \geq \max_{S \in S_K} \mathbb{E}_S \left[ \sum_{t=1}^{T} R_v(S) - R_v(\bar{S}_t) \right]
$$

$$
\geq \frac{1}{|S_K|} \sum_{S \in S_K} \mathbb{E}_S \left[ \sum_{t=1}^{T} R_v(S) - R_v(\bar{S}_t) \right]
$$

$$
\geq \frac{1}{|S_K|} \sum_{S \in S_K} \sum_{\tilde{S} \in \mathcal{S}_K, \tilde{S} \neq S} \mathbb{E}_S[\bar{N}_i] \cdot \frac{\epsilon}{9K}.
$$

(4.80) Here Eq. (4.80) holds because the maximum regret is always lower bounded by the average regret (averaging over all parameterization $\theta_S$ for $S \in S_K$), Eq. (4.81) follows from Lemma 69, and Eq. (4.82) holds because $\sum_{i=1}^{N} \mathbb{E}_S[\bar{N}_i] = \mathbb{E}_S \left[ \sum_{i=1}^{N} \bar{N}_i \right] = TK$ for any $S \subseteq [N]$. The lower bound proof is then reduced to finding the largest $\epsilon$ 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 $\sum_{i \in S} \mathbb{E}_S[\bar{N}_i]$ term. Ideally, we expect this term to be small (e.g., around $K/N$ fraction of $\sum_{i=1}^{N} \mathbb{E}_S[\bar{N}_i] = TK$) because $S \in S_K$ is of size $K$. However, a bandit assortment selection algorithm, with knowledge of $S$, could potentially allocate its assortment selections so that $\bar{N}_i$ 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 $\sum_{i=1}^{N} \mathbb{E}_S[\bar{N}_i] = NK$ property and Pinsker’s inequality (Tsybakov, 2009) to bound the discrepancy in expectations under different parameterization.

Let $S_{K-1}^{(i)} = S_{K-1} \cap \{ S \subseteq [N] : i \notin S \}$ be all subsets of size $K - 1$ that do not include $i$. Re-arranging summation order we have

$$
\frac{1}{|S_K|} \sum_{S \in S_K} \frac{1}{K} \sum_{\tilde{S} \in S} \mathbb{E}_S[\bar{N}_i] = \frac{1}{K} \sum_{i=1}^{N} \frac{1}{|S_K|} \sum_{S \in S_K, \tilde{S} \neq S} \mathbb{E}_S[\bar{N}_i]
$$

$$
= \frac{1}{K} \sum_{i=1}^{N} \frac{1}{|S_K|} \sum_{S \in S_{K-1}^{(i)}} \mathbb{E}_S[\bar{N}_i].
$$

(4.83) Denote $P = P_S$ and $Q = P_{S_{K-1}^{(i)}}$. Also note that $0 \leq \bar{N}_i \leq T$ almost surely under both $P$ and $Q$. Using Pinsker’s inequality we have that

$$
\left| \mathbb{E}_P[\bar{N}_i] - \mathbb{E}_Q[\bar{N}_i] \right| \leq \sum_{j=0}^{T} j \cdot \left| P[\bar{N}_i = j] - Q[\bar{N}_i = j] \right|
$$

148
\[ T \cdot \sum_{j=0}^{T} \left| P[\tilde{N}_i = j] - Q[\tilde{N}_i = j] \right| \leq T \cdot \|P - Q\|_{TV} \leq T \cdot \sqrt{\frac{1}{2} \text{KL}(P\|Q)}. \]

Here \( \|P - Q\|_{TV} = \sup_A |P(A) - Q(A)| \) and \( \text{KL}(P\|Q) = \int (\log \frac{dP}{dQ})dP \) are the total variation and the Kullback-Leibler (KL) divergence between \( P \) and \( Q \), respectively. Subsequently,

\[ \frac{1}{|S_K|} \sum_{S \in S_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S[\tilde{N}_i] \leq \frac{1}{K} \sum_{i=1}^{N} \frac{1}{|S_K|} \sum_{S' \in S_{K-1}} \left( \mathbb{E}_{S'}[\tilde{N}_i] + T \sqrt{\frac{1}{2} \text{KL}(P_{S'}\|P_{S' \cup \{i\}})} \right). \quad (4.84) \]

The first term on the right-hand side of Eq. (4.120) is easily bounded:

\[ \frac{1}{|S_K|} \sum_{S \in S_K} \frac{1}{K} \sum_{i \in S} \mathbb{E}_S[\tilde{N}_i] = \frac{1}{|S_K|} \sum_{S' \in S_{K-1}} \frac{1}{K} \sum_{i \notin S'} \mathbb{E}_{S'}[\tilde{N}_i] \leq \frac{|S_{K-1}|}{K|S_K|} \cdot TK = \frac{N}{K} \cdot TK = \frac{TK}{N-K+1} \leq \frac{T}{3}. \quad (4.85) \]

Here the last inequality holds because \( K \leq N/4 \) and hence \( \frac{TK}{N-K+1} \leq \frac{TK}{3K+1} \leq \frac{T}{3} \). Combining all inequalities we have that

\[ \max_{S \in S_K} \mathbb{E}_S \left[ \sum_{t=1}^{T} R_{u}(S^*_u) - R_{u}(S_t) \right] \geq \epsilon \left( \frac{2T}{3} - \frac{T}{|S_K|} \sum_{S \in S_{K-1}} \frac{1}{K} \sum_{i \notin S'} \sqrt{\frac{1}{2} \text{KL}(P_{S'}\|P_{S' \cup \{i\}})} \right). \quad (4.86) \]

It remains to bound the KL divergence between two “neighboring” parameterization \( \theta_{S'} \) and \( \theta_{S' \cup \{i\}} \) for all \( S' \in S_{K-1} \) and \( i \notin S' \), which we elaborate in the next section.

**KL-divergence between assortment selections**

Define \( N_i := \sum_{t=1}^{T} \mathbb{1}[i \in S_t] \). Note that because \( S_t \subseteq \tilde{S}_t \), we have \( N_i \leq \tilde{N}_i \) almost surely and hence \( \sum_{i=1}^{N} \mathbb{E}_S[N_i] \leq \sum_{i=1}^{N} \mathbb{E}_S[\tilde{N}_i] = TK \) for all \( S \subseteq [N] \).

**Lemma 70.** Suppose \( \epsilon \in (0, 1/2] \). For any \( S' \in S_{K-1} \) and \( i \notin S' \), it holds that \( \text{KL}(P_{S'}\|P_{S' \cup \{i\}}) \leq \mathbb{E}_{S'}[N_i] \cdot 63\epsilon^2 / K \).
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

$$\text{KL}(P\|Q) \leq \sum_{j=0}^{J} \frac{\varepsilon_j^2}{q_j}.$$  

**Proof.** We have that

$$\text{KL}(P\|Q) = \sum_{j=0}^{J} (p_j + \varepsilon_j) \log \frac{p_j + \varepsilon_j}{q_j}$$

$$\leq \sum_{j=0}^{J} (p_j + \varepsilon_j) \frac{\varepsilon_j}{q_j} \leq \sum_{j=0}^{J} \frac{\varepsilon_j^2}{q_j}.$$  

Here (a) holds because $\log(1 + x) \leq x$ for all $x \geq -1$ and (b) holds because $\sum_{j=0}^{J} \varepsilon_j = 0$.  

We are now ready to prove Lemma 70.

**Proof.** It is clear that for any $S_t \subseteq [N], |S_t| \leq K$ such that $i \notin S_t$, we have $\text{KL}(P_{S'}(\cdot|S_t) \| P_{S \cup \{i\}}(\cdot|S_t)) = 0$. Therefore, we shall focus only on those $S_t \subseteq [N]$ with $i \in S_t$, which happens for $E_{S'}[N_t]$ epochs in expectation. Define $K' := |S_t| \leq K$ and $J := |S_t \cap S'| \leq K - 1$. Re-write the probability of $i_t = j$ as $p_j = v_j/(a + J\varepsilon/K)$ and $q_j = v_j/(a + (J + 1)\varepsilon/K)$ under $P_{S'}$ and $P_{S \cup \{i\}}$, respectively, where $a = 1 + K'/K \in (1, 2]$. We then have that

$$|p_0 - q_0| = \left| \frac{1}{a + J\varepsilon/K} - \frac{1}{a + (J + 1)\varepsilon/K} \right| \leq \frac{\varepsilon}{K};$$

$$|p_j - q_j| \leq \frac{1 + \varepsilon}{K} \left| \frac{1}{a + J\varepsilon/K} - \frac{1}{a + (J + 1)\varepsilon/K} \right| \leq \frac{2\varepsilon}{K^2},$$

if $1 \leq j \leq N, j \neq i;$

$$|p_j - q_j| \leq \left| \frac{1}{K} \frac{1}{a + J\varepsilon/K} - \frac{1 + \varepsilon}{K} \frac{1}{a + (J + 1)\varepsilon/K} \right|$$

$$\leq \frac{\varepsilon}{K} \frac{1}{a + (J + 1)\varepsilon/K} + \frac{1}{K} \left| \frac{1}{a + J\varepsilon/K} - \frac{1}{a + (J + 1)\varepsilon/K} \right|$$

$$\leq \frac{\varepsilon}{K} + \frac{1}{K} \frac{\varepsilon}{K^2} \leq \frac{\varepsilon}{K^2} + \frac{\varepsilon}{K} \leq \frac{4\varepsilon}{K},$$

if $j = i.$
Note that $q_0 \geq 1/3$ and $q_j \geq 1/(3K)$ for $j \geq 1$, because $\epsilon \in (0, 1/2]$, $a \in (1, 2]$ and $J \leq K - 1$. Invoking Lemma 71 we have that

$$\text{KL}(P_{S_i} \cdot |S_i) P_{S_{\cup (i)}} \cdot |S_i) \leq \frac{3\epsilon^2}{K^2} + 3K \cdot \frac{4J\epsilon^2}{K^2} + 3K \cdot \frac{16\epsilon^2}{K^2} \leq \frac{3\epsilon^2}{K^2} + \frac{12\epsilon^2}{K^2} + \frac{48\epsilon^2}{K} \leq \frac{63\epsilon^2}{K}.$$

\[\square\]

**Putting everything together**

Using Hölder’s inequality, we have that

$$\frac{T}{|S_K|} \sum_{S \in S_{K-1}} \frac{1}{K} \sum_{i \notin S} \sqrt{\frac{1}{2} \text{KL}(P_{S_i} \cdot |S_i) P_{S_{\cup (i)}} \cdot |S_i)} \leq \frac{T|S_{K-1}|}{K|S_K|} \cdot \max_{S \in S_{K-1}} \sum_{i \notin S} \sqrt{\frac{1}{2} \text{KL}(P_{S_i} \cdot |S_i) P_{S_{\cup (i)}} \cdot |S_i)}.$$

By Jensen’s inequality and the concavity of the square root, we have

$$\frac{1}{N - K + 1} \sum_{i \notin S} \sqrt{\frac{1}{2} \text{KL}(P_{S_i} \cdot |S_i) P_{S_{\cup (i)}} \cdot |S_i)} \leq \sqrt{\frac{1}{2(N - K + 1)} \sum_{i \notin S} \text{KL}(P_{S_i} \cdot |S_i) P_{S_{\cup (i)}} \cdot |S_i)}.$$

Invoking Lemma 70, we obtain

$$\frac{1}{N - K + 1} \sum_{i \notin S} \text{KL}(P_{S_i} \cdot |S_i) P_{S_{\cup (i)}} \cdot |S_i) \leq \frac{1}{N - K + 1} \sum_{i \notin S} \mathbb{E}_{S_i}[N_i] \cdot \frac{63\epsilon^2}{K},$$

$$\leq \frac{63\epsilon^2}{K(N - K + 1)} \sum_{i = 1}^{N} \mathbb{E}_{S_i}[N_i] \leq \frac{126\epsilon^2}{NK} \cdot TK = \frac{126T\epsilon^2}{N}.$$

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 $\epsilon T/27 \geq \min \{0.001\sqrt{NT}, T/54\}$. 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 \( \hat{\phi}_{i,\theta}, \tilde{u}_{i,\theta} \) concentrate around the true values \( \phi_{i,\theta}, u_{i,\theta} \).

**Lemma 72.** Suppose \( T(i, \theta) \geq 96 \ln(2MTK) \). With probability \( 1 - T^{-1} \) uniformly over all \( i \in [M], \theta \in \mathcal{K}_i \) and \( t \in [T] \)

\[
|\tilde{u}_{i,\theta} - u_{i,\theta}| \leq \min \left\{ U, \sqrt{\frac{96 \max(\tilde{u}_{i,\theta}, \overline{u}_{i,\theta}) \ln(2MTK)}{T(i, \theta)} + \frac{144 \ln(2MTK)}{T(i, \theta)}} \right\}; \tag{4.87}
\]

\[
|\tilde{\phi}_{i,\theta} - \phi_{i,\theta}| \leq \min \left\{ 1, \sqrt{\frac{\ln(2MTK)}{T(i, \theta)\tilde{u}_{i,\theta}}} \right\}. \tag{4.88}
\]

The following corollary is an immediate consequence of Lemma 72:

**Corollary 7.** With probability \( 1 - T^{-1}, \overline{u}_{i,\theta} \geq u_{i,\theta} \) and \( \overline{\phi}_{i,\theta} \geq \phi_{i,\theta} \) 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 \( R \). Recall that \( \overline{R}'(\theta) = \left[ \sum_{i=1}^{M} \overline{\phi}_{i,\theta} \overline{u}_{i,\theta} \right] / \left[ 1 + \sum_{i=1}^{M} \overline{u}_{i,\theta} \right] \) and \( \overline{R}'(\theta) = \left[ \sum_{i=1}^{M} \phi_{i,\theta} u_{i,\theta} \right] / \left[ 1 + \sum_{i=1}^{M} u_{i,\theta} \right] \).

**Corollary 8.** With probability \( 1 - T^{-1}, \overline{R}'(\hat{\theta}) \geq \overline{R}(\hat{\theta}) \) and \( \overline{R}'(\theta^*) \geq \overline{R}'(\theta^*) \), where \( \hat{\theta}, \theta^* \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M \) 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).

\[
\text{Regret}(\{\hat{\theta}^{(t)}\}_{t=1}^{T}) = \mathbb{E} \sum_{t=1}^{T} \overline{R}'(\theta^*) - \overline{R}'(\hat{\theta}^{(t)})
\]

\[
\leq \mathbb{E} \sum_{t=1}^{T} \overline{R}'(\theta^*) - \overline{R}'(\hat{\theta}^{(t)}) \bigg| \mathcal{A} \right] \Pr[\mathcal{A}] + O(T) \cdot \Pr[\mathcal{A}^c]
\]

\[
\leq O(1) + \mathbb{E} \sum_{t=1}^{T} \overline{R}'(\theta^*) - \overline{R}'(\hat{\theta}^{(t)}) \bigg| \mathcal{A} \right] \tag{4.89}
\]

\[
\leq O(1) + \mathbb{E} \sum_{t=1}^{T} \overline{R}'(\theta^{(t)}) - \overline{R}'(\hat{\theta}^{(t)}) \bigg| \mathcal{A} \right]. \tag{4.90}
\]

\[
= O(1) + \mathbb{E} \left[ \sum_{\tau} \mathbb{E}_\tau \cdot \left( \overline{R}'(\hat{\theta}^{(\tau)}) - \overline{R}'(\hat{\theta}^{(\tau)}) \right) \bigg| \mathcal{A} \right]. \tag{4.91}
\]

152
Here, $\tilde{\theta}^{(r)}$ denotes any $\theta^{(t)}$ in the $\tau$-th epoch $\mathcal{E}_\tau$. We also note that Eq. (4.89) holds because $\Pr[A^c] \leq T^{-1}$ and $\overline{R}'(\theta^*) \geq R'(\theta^*)$, and Eq. (4.90) holds because $\overline{R}'(\theta^{(t)}) \geq R'(\theta^*)$, since $\theta^{(t)}$ is the maximizer of $\overline{R}'$ at time $t$.

It remains to upper bound the discrepancy between $\overline{R}'(\tilde{\theta}^{(r)})$ and $R'(\tilde{\theta}^{(r)})$ at every epoch $\tau$. 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 $\theta = (\theta_1, \ldots, \theta_M) \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$,\[
|\overline{R}'(\theta) - R'(\theta)| \leq \frac{1}{1 + \sum_{i=1}^{M} u_i, \theta_i} \left[ \sum_{i=1}^{M} \frac{\overline{\theta}_i, \theta_i - u_i, \theta_i}{1 + u_i, \theta_i} + \sum_{i=1}^{M} u_i, \theta_i (\overline{\phi}_i, \theta_i - \phi_i, \theta_i) \right]. \tag{4.92}
\]

Note that $\mathbb{E}[\mathcal{E}_\tau] = 1 + \sum_{i=1}^{M} \mathbb{E}[\mathcal{E}_t, \tau] = 1 + \sum_{i=1}^{M} u_i, \theta_i$. Combining Lemma 73 with Eq. (4.91) we obtain

$$\text{Regret}(\{\tilde{\theta}^{(r)}\}_{r=1}^{T}) \leq O(1) + \sum_{\tau} \mathbb{E} \left[ \sum_{i=1}^{M} \frac{\overline{\theta}_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{4.93}\]

The following lemmas upper bound (asymptotically) the two terms in Eq. (4.93) separately.

**Lemma 74.** Conditioned on event $A$, it holds that

$$\sum_{\tau} \sum_{i=1}^{M} \frac{\overline{\theta}_i, \theta_i - u_i, \theta_i}{1 + u_i, \theta_i} \leq \sqrt{MKT \log(MTK)} + M\log^2(MTK). \tag{4.94}\]

**Lemma 75.** Conditioned on event $A$, it holds that

$$\sum_{\tau} \sum_{i=1}^{M} u_i, \theta_i (\overline{\phi}_i, \theta_i - \phi_i, \theta_i) \leq \sqrt{MKT \log(MTK)} + M\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 $|\tilde{u}_{i, \theta} - u_{i, \theta}|$ for fixed $i \in [M]$ and $\theta \in \mathcal{K}_i$.

**Case 1: $u_{i, \theta} \leq 1$.** Let $\delta > 0$ be a parameter to be specified later. Applying Lemma 94, we have

$$\Pr[|\tilde{u}_{i, \theta} - u_{i, \theta}| > \delta u_{i, \theta}] \leq \exp \left\{ -\frac{nu_{i, \theta}\delta^2}{2(1 + \delta) \times 4} \right\} + \exp \left\{ -\frac{nu_{i, \theta}\delta^2}{6 \times 4} \left( 3 - \frac{2\delta u_{i, \theta}}{2} \right) \right\} \leq \exp \left\{ -\frac{nu_{i, \theta}\delta^2}{16 \max(1, \delta)} \right\} + \exp \left\{ -\frac{nu_{i, \theta}\delta^2}{24} (3 - \delta u_{i, \theta}) \right\} \tag{4.96}\]

4Recall that in Algorithm 11, $\tilde{\theta}^{(t)}$ does not change within the same epoch $\mathcal{E}_\tau$. We write $\tilde{\theta}^{(\tau)}$ to highlight that $\tilde{\theta}^{(\tau)}$ is the maximizer of $\overline{R}'$ in the $\tau$-th epoch (see Step 5 of Algorithm 11).
Suppose in addition that \( \delta u_{i,\theta} \leq 2 \). Then

\[
\Pr[|\hat{u}_{i,\theta} - u_{i,\theta}| > \delta u_{i,\theta}] \leq 2 \exp \left\{ -\frac{nu_{i,\theta} \min\{\delta, \delta^2\}}{24} \right\}
\]

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 \( 1 - T^{-1} \)

\[
|\hat{u}_{i,\theta} - u_{i,\theta}| \leq \delta u_{i,\theta} \leq \sqrt{\frac{48u_{i,\theta} \ln(2MTK)}{T(i, \theta)}} + \frac{48 \ln(2MTK)}{T(i, \theta)}.
\]  

(4.96)

Note that if \( T(i, \theta) \geq 48 \ln(2MTK) \) the condition \( \delta u_{i,\theta} \leq 2 \) is met. Replacing all occurrences of \( u_{i,\theta} \) in Eq. (4.96) by \( \hat{u}_{i,\theta} \) and using the fact that \( \sqrt{a+b} \leq \sqrt{a} + \sqrt{b} \), we have

\[
|\hat{u}_{i,\theta} - u_{i,\theta}| \leq \sqrt{\frac{48\hat{u}_{i,\theta} \ln(2MTK)}{T(i, \theta)}} + \frac{48 \ln(2MTK)}{T(i, \theta)} + \sqrt{\frac{48|\hat{u}_{i,\theta} - u_{i,\theta}| \ln(2MTK)}{T(i, \theta)}}
\]

\[
\leq \sqrt{\frac{48\hat{u}_{i,\theta} \ln(2MTK)}{T(i, \theta)}} + \frac{48 \ln(2MTK)}{T(i, \theta)} + \sqrt{\frac{48 \ln(2MTK)}{T(i, \theta)}}
\]

\[
\leq \sqrt{\frac{48\hat{u}_{i,\theta} \ln(2MTK)}{T(i, \theta)}} + \frac{144 \ln(2MTK)}{T(i, \theta)}.
\]  

(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

\[
\Pr[|\hat{u}_{i,\theta} - u_{i,\theta}| > \delta u_{i,\theta}] \leq \exp \left\{ -\frac{T(i, \theta)u_{i,\theta}^2 \delta^2}{6 \times 4u_{i,\theta}^2} \left( 3 - \frac{2u_{i,\theta} \delta}{2 \times u_{i,\theta}} \right) \right\} + \exp \left\{ -\frac{T(i, \theta)u_{i,\theta}^2 \delta^2}{2 \times 4u_{i,\theta}^2} \right\}
\]

\[
\leq 2 \exp\{-T(i, \theta)\delta^2/12\}.
\]

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 $1 - T^{-1}$,

$$|\hat{u}_{i,\theta} - u_{i,\theta}| \leq \delta u_{i,\theta} \leq \sqrt{\frac{24u_{i,\theta}^2 \ln(2MTK)}{T(i,\theta)}}.  \tag{4.98}$$

Note that $\delta \leq 1$ holds if $T(i,\theta) \geq 24 \ln(2MTK)$. In addition, if $T(i,\theta) \geq 96 \ln(2MTK)$ we have $|\hat{u}_{i,\theta} - u_{i,\theta}| \leq 0.5u_{i,\theta}$ and hence $\hat{u}_{i,\theta} \geq 0.5u_{i,\theta}$. Subsequently, Eq. (4.98) implies

$$|\hat{u}_{i,\theta} - u_{i,\theta}| \leq \sqrt{\frac{96u_{i,\theta}^2 \ln(2MTK)}{T(i,\theta)}}.  \tag{4.99}$$

Finally, combining Eqs. (4.97,4.99) we proved the upper bound on $|\hat{u}_{i,\theta} - u_{i,\theta}|$.

We next prove the upper bound on $|\hat{\phi}_{i,\theta} - \phi_{i,\theta}|$. Recall that for each $\tau \in T(i,\theta)$, $\hat{\tau}_{ik}$ is the sum of $\hat{n}_{ik}$ i.i.d. random variables with mean $\phi_{i,\theta}$ and within range $[0,1]$ almost surely. Also note that $\sum_{\tau \in T(i,\theta)} \hat{n}_{ik} = T(i,\theta)\hat{u}_{i,\theta}$. Applying Hoeffding’s inequality (Lemma 89) we have for any $\delta > 0$ that

$$\Pr \left[|\hat{\phi}_{i,\theta} - \phi_{i,\theta}| > \delta \right] \leq 2 \exp \left\{-2\delta^2 \cdot T(i,\theta)\hat{u}_{i,\theta} \right\}. $$

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

$$|\hat{\phi}_{i,\theta} - \phi_{i,\theta}| \leq \sqrt{\frac{\ln(2MTK)}{T(i,\theta)\hat{u}_{i,\theta}}}.  \tag{4.100}$$

**Proof of Corollary 8** We first prove $\overline{R}(\theta) \geq R'(\hat{\theta})$. By definition, $\sum_{i=1}^M (\overline{\phi}_{i,\hat{\theta}} - \overline{R}(\hat{\theta}))\pi_{i,\hat{\theta}_i} = \overline{R}(\hat{\theta})$. In addition, because $\overline{R}(\hat{\theta})$ is the maximizer of $\overline{R}$, setting $\lambda = \overline{R}(\hat{\theta})$ and by the second property of Lemma 58 we know that $\psi_\lambda(\theta) = \lambda$ and $\psi_\lambda(\theta) \leq \lambda$ for all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$, where $\psi_\lambda(\theta) = \sum_{i=1}^M (\overline{\phi}_{i,\hat{\theta}_i} - \lambda)\pi_{i,\hat{\theta}_i}$.

We claim that $\overline{\phi}_{i,\hat{\theta}_i} \geq \overline{R}(\theta)$ whenever $\pi_{i,\hat{\theta}_i} > 0$. Assume the contrary, that $\overline{\phi}_{i,\hat{\theta}_i} < \overline{R}(\theta) = \lambda$ and $\pi_{i,\hat{\theta}_i} > 0$ for some $i \in [M]$. Consider $\tilde{\theta}' = (\tilde{\theta}_1', \ldots, \tilde{\theta}_M')$ defined as $\tilde{\theta}_i' = \infty$ and $\tilde{\theta}_i' = \hat{\theta}_i$ for all $i' \neq i$. Because $\tilde{\theta}_i' = \infty$ we know that $\pi_{i,\tilde{\theta}_i'} = u_{i,\hat{\theta}_i} = 0$. Subsequently, $\psi_\lambda(\tilde{\theta}') = \psi_\lambda(\tilde{\theta}') - (\overline{\phi}_{i,\hat{\theta}_i} - \lambda)\pi_{i,\hat{\theta}_i} > \psi_\lambda(\tilde{\theta}) = \lambda$. This contradicts the condition that $\psi_\lambda(\theta) \leq \lambda$ for all $\theta \in \mathcal{K}_1 \times \cdots \times \mathcal{K}_M$.

Define $\psi^0_\lambda(\theta) := \sum_{i=1}^M (\phi_{i,\theta_i} - \lambda)u_{i,\theta_i}$, which is similar to the definition of $\psi_\lambda$ except all occurrences of $\overline{\phi}_{i,\hat{\theta}_i}$ and $\pi_{i,\hat{\theta}_i}$ are replaced by their true values $\phi_{i,\theta_i}, u_{i,\theta_i}$. Because $\overline{\phi}_{i,\hat{\theta}_i} \geq \overline{R}(\theta)$ for all $\pi_{i,\hat{\theta}_i} > 0$, and $\overline{\phi}_{i,\hat{\theta}_i}, u_{i,\theta_i}$ are upper bounds of $\phi_{i,\theta_i}, u_{i,\theta_i}$, we conclude that $\psi^0_\lambda(\hat{\theta}) \leq \psi^0_\lambda(\theta) = \lambda$, implying that $\sum_{i=1}^M (\phi_{i,\hat{\theta}_i} - \lambda)u_{i,\hat{\theta}_i} \leq \lambda$. Re-arranging terms we have $R'(\theta) = \frac{\sum_{i=1}^M \phi_{i,\theta_i}u_{i,\theta_i}}{1 + \sum_{i=1}^M u_{i,\theta_i}} \leq \lambda = \overline{R}(\hat{\theta})$.

We next prove $R'(\theta^*) \geq R'(\theta^*)$. Recall that $R'(\theta^*) = \frac{\sum_{i=1}^M \phi_{i,\theta^*}u_{i,\theta^*}}{1 + \sum_{i=1}^M u_{i,\theta^*}}$. Hence, $\psi^0_\lambda(\theta^*) = \lambda$ for $\lambda = R'(\theta^*)$, meaning that $\sum_{i=1}^M (\phi_{i,\theta^*} - \lambda)u_{i,\theta^*} = \lambda$. By similar analysis we know that $\phi_{i,\theta^*} \geq \lambda$ for all $u_{i,\theta^*} > 0$ too. Because $\overline{\phi}_{i,\hat{\theta}_i}, u_{i,\theta_i}$ are upper bounds of $\phi_{i,\theta_i}, u_{i,\theta_i}$ and...
Lemma 72, the expectation of the first term in Eq. (4.93) can be upper bounded by,

\[ R_1(\theta^*) = \sum_{i=1}^{M} (\overline{\phi}_i, \theta^*) - \lambda \overline{u}_i, \theta^* \geq \sum_{i=1}^{M} (\phi_i, \theta^*) - \lambda \overline{u}_i, \theta^* = R_1(\theta^*) = \frac{\left[ \sum_{i=1}^{M} (\overline{\phi}_i, \theta^*) \overline{u}_i, \theta^* \right]}{[1 + \sum_{i=1}^{M} \overline{u}_i, \theta^*]} \geq R'(\theta^*). \]

**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

\[
\left( 1 + \sum_{i=1}^{M} u_i \right) [\overline{R}' - R'] = \left( 1 + \sum_{i=1}^{M} u_i \right) \left[ \frac{\sum_{i=1}^{M} \overline{u}_i, \theta_i - \sum_{i=1}^{M} u_i, \phi_i}{1 + \sum_{i=1}^{M} \overline{u}_i} \right] = \sum_{i=1}^{M} \overline{u}_i (1 + \sum_{i=1}^{M} \overline{u}_i - u_i) + \sum_{i=1}^{M} u_i (\overline{\phi}_i - \phi_i)
\]

The first term on the right-hand side of Eq. (4.101) can be further upper bounded by

\[
\sum_{i=1}^{M} \frac{1 + \sum_{i=1}^{M} u_i \overline{u}_i}{1 + \sum_{i=1}^{M} u_i \overline{u}_i} - u_i = \sum_{i=1}^{M} \sum_{i=1}^{M} u_i, \theta_i (1 + \sum_{i=1}^{M} u_i) - \sum_{i=1}^{M} u_i (1 + \sum_{i=1}^{M} u_i) \leq \sum_{i=1}^{M} \overline{u}_i - u_i.
\]

Here the last inequalities holds because the \( \sum_{i=1}^{M} u_i, \theta_i \) term cancels out, and \( 1 + \sum_{i=1}^{M} u_i \overline{u}_i \geq 1 + \overline{u}_i \geq 1 + u_i. \)

**Proof of Lemma 74** Define \( T_r(i, \theta) \) as the size of \( T(i, \theta) \) after epoch \( \tau \), and \( T_0(i, \theta) \) as the final size of \( T(i, \theta) \) when Algorithm 11 terminates (i.e., the total number of epochs in which singleton \( \theta \in K_i \) was offered in nest \( i \)). Define also \( \overline{u}_i^{(\tau)} \) to be the estimate of \( u_i, \theta_i \) at epoch \( \tau \). Using Lemma 72, the expectation of the first term in Eq. (4.93) can be upper bounded by,

\[
\sum_{r=1}^{M} \sum_{i=1}^{M} U \cdot 1\{T_r(i, \overline{\theta}_i^{(\tau)}) < 96 \ln(2MTK)\} + \sqrt{\frac{96 \max(\overline{u}_i^{(\tau)}, [\overline{u}_i^{(\tau)}]^2) \ln(2MTK)}{T_r(i, \overline{\theta}_i^{(\tau)}) \cdot (1 + u_i^{(\tau)})^2}}
\]

\[
+ \frac{144 \ln(2MTK)}{T_r(i, \overline{\theta}_i^{(\tau)})} \cdot 1\{T_r(i, \overline{\theta}_i^{(\tau)}) \geq 96 \ln(2MTK)\}
\]

\[
= \sum_{r=1}^{M} \sum_{i=1}^{M} \sum_{\ell=0}^{T_0(i, \theta)} U \cdot 1\{r < 96 \ln(2MTK)\} + \sqrt{\frac{96 \max(\overline{u}_i^{(\tau)}, [\overline{u}_i^{(\tau)}]^2) \ln(2MTK)}{\ell(1 + u_i, \theta)^2}}
\]

\[
+ \frac{144 \ln(2MTK)}{\ell} \cdot 1\{r \geq 96 \ln(2MTK)\}
\]

156
\[
\sum_{i=1}^{M} \sum_{\theta \in K_i} U \log(MTK) \leq \sqrt{u_i,\theta T_0(i, \theta) \log(MTK)} + \log T_0(i, \theta) \log(MTK) \quad (4.103)
\]

\[
\leq MKU \log^2(MTK) + \sum_{i=1}^{M} \sum_{\theta \in K_i} \sqrt{u_i,\theta T_0(i, \theta) \log(MTK)}. \quad (4.104)
\]

Here Eq. (4.103) holds by plugging in upper bounds on $|\hat{u}_{i,\theta} - u_{i,\theta}|$ (Lemma 72) and noting that, $\max\{a, a^2\}/(1 + a)^2 \leq a_i, \sum_{\ell \in T_0(i, \theta)} \ell^{-1/2} \leq \sqrt{T_0(i, \theta)}$ and $\sum_{\ell \in T_0(i, \theta)} \ell^{-1} \leq \log T_0(i, \theta)$, Eq. (4.104) holds by replacing $\log T_0$ with $\log(MTK)$.

Applying Cauchy-Schwartz inequality and the fact that $E[\hat{u}_{i,\tau}] = u_{i,\theta_1}, E[\hat{u}_{i,\tau}] = 1 + \sum_{i=1}^{M} E[\hat{u}_{i,\tau}]$, the summation term in (4.104) can be further bounded by

\[
\sum_{i=1}^{M} \sum_{\theta \in K_i} \sqrt{u_i,\theta T_0(i, \theta) \log(MTK)} \leq \sqrt{M|K|} \cdot \sum_{i=1}^{N} \sum_{\theta \in K_i} u_i,\theta T_0(i, \theta) \log(MTK)
\]

\[
= \sqrt{M|K|} \cdot \sum_{i=1}^{N} \sum_{\theta \in K_i} \sum_{\tau \in T(i, \theta)} E[\hat{u}_{i,\tau}] \log(MTK)
\]

\[
\leq \sqrt{M|K|} \cdot \sum_{\tau} E[\hat{u}_{i,\tau}] \log(MTK)
\]

Subsequently,

\[
\sum_{\theta \in K_i} \sum_{\tau = 1}^{M} \frac{\hat{u}_{i,\theta}^{(r)}(\tau) - u_{i,\theta}^{(r)}(\tau)}{1 + u_{i,\theta}^{(r)}(\tau)} \leq \sqrt{MKT \log(MTK)} + MKU \log^2(MTK).
\]

**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 $T(i, \theta) \hat{u}_{i,\theta}$ 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 $i \in [M], \theta \in K_i$ such that $u_{i,\theta} \geq 768 \ln(2MTK)/T(i, \theta)$ and $T(i, \theta) \geq 96 \ln(2MTK)$, we have $T(i, \theta) \hat{u}_{i,\theta} \geq 0.5T(i, \theta) u_{i,\theta}$.

**Proof.** First consider the case of $u_{i,\theta} \geq 1$. By Eq. (4.98) in the proof of Lemma 72, if $T(i, \theta) \geq 96 \ln(2MTK)$ we have $|\hat{u}_{i,\theta} - u_{i,\theta}| \leq 0.5u_{i,\theta}$ and therefore $T(i, \theta) \hat{u}_{i,\theta} \geq 0.5T(i, \theta) u_{i,\theta}$.

In the rest of the proof we consider the case of $768 \ln(2MTK)/T(i, \theta) \leq u_{i,\theta} \leq 1$. By Eq. (4.96) in the proof of Lemma 72, we have

\[
T(i, \theta) \hat{u}_{i,\theta} \geq T(i, \theta) u_{i,\theta} - \sqrt{48T(i, \theta) u_{i,\theta} \ln(2MTK)} - 48 \ln(2MTK).
\]

Under the condition that $u_{i,\theta} \geq 768 \ln(2MTK)/T(i, \theta)$, the above inequality yields $T(i, \theta) \hat{u}_{i,\theta} \geq 0.5T(i, \theta) u_{i,\theta}$. \qed
Combining Corollary 9 with Lemma 72 and noting that \(|\bar{\phi}_{i,\theta} - \phi_{i,\theta}| \leq 1\) always holds, the second term on the right-hand side of Eq. (4.93) can be upper bounded by

\[
\sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} T_{0}(i, \theta) \sum_{\ell=0}^{T_{\theta}(i, \theta)} U\{\ell < 96 \ln(2MTK)\} + \left[ \frac{768 \ln(2MTK)}{\ell} \right] + u_{i, \theta} \sqrt{\frac{2 \ln(2MTK)}{T(i, \theta) u_{i, \theta}}} \cdot 1\{\ell \geq 96 \ln(2MTK)\}
\]

\[
\leq \sum_{i=1}^{M} \sum_{\theta \in \mathcal{K}_i} U \log(MKT) + \sqrt{u_{i, \theta} T_{0}(i, \theta) \log(MKT)} + \log T_{0}(i, \theta) \log(MKT).
\]

(4.105)

Using similar derivation as in Eq. (4.94), we have

\[
\sum_{\tau=1}^{M} \sum_{\theta \in \mathcal{K}_i} u_{i, \theta} [\bar{\phi}_{i, \theta} - \phi_{i, \theta}] \leq \sqrt{MKT \log(MTK)} + MKU \log^2(MTK).
\]

### 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 \(\tilde{\theta}_i^* := \lfloor \theta_i^*/\delta \rfloor \cdot \delta\) for all \(i \in [M]\) and \(\tilde{\theta}^* := (\tilde{\theta}_1^*, \cdots, \tilde{\theta}_M^*)\). It is easy to verify that \(\tilde{\theta}^* \in \tilde{\mathcal{K}}_1 \times \cdots \times \tilde{\mathcal{K}}_M\).

Therefore, it suffices to prove that \(R'(\tilde{\theta}^*) \geq R^* - \delta\) where \(R^* = R'(\theta^*)\).

To simplify notations, abbreviate \(R_i = R_i(\mathcal{L}_i(\theta^*_i))\), \(V_i = V_i(\mathcal{L}_i(\theta^*_i))\), \(\tilde{R}_i = R_i(\mathcal{L}_i(\tilde{\theta}^*_i))\) and \(\tilde{V}_i = V_i(\mathcal{L}_i(\tilde{\theta}^*_i))\), where \(R_i(\cdot)\) and \(V_i(\cdot)\) are defined in Eqs. (4.11,4.13). Denote also that \(x_i := \tilde{V}_i - V_i\). By definition of \(R_i\) and \(\tilde{R}_i\), we have \(R_i V_i = \sum_{r_{ij} \geq \theta^*_i} r_{ij} v_{ij}\) and \(\tilde{R}_i \tilde{V}_i = \sum_{r_{ij} \geq \theta^*_i} r_{ij} v_{ij}\). Subsequently,

\[
\tilde{R}_i \tilde{V}_i = R_i V_i + \sum_{\theta^*_i \geq r_{ij} \geq \tilde{\theta}^*_i} r_{ij} v_{ij} \geq R_i V_i + x_i (\theta^*_i - \delta).
\]

(4.106)

Here the last inequality holds because \(|\theta^*_i - \tilde{\theta}^*_i| \leq \delta\) and \(\sum_{\theta^*_i \geq r_{ij} \geq \tilde{\theta}^*_i} v_{ij} = \tilde{V}_i - V_i = x_i\). Subsequently,

\[
\tilde{V}_i^{-\gamma_i} \left[ \tilde{R}_i - (R^* - \delta) \right] = (V_i + x_i)^{\gamma_i} \left[ \tilde{R}_i - (R^* - \delta) \right]
\]

\[
\geq (V_i + x_i)^{\gamma_i} \left[ \frac{R_i V_i + x_i (\theta^*_i - \delta)}{V_i + x_i} - R^* + \delta \right]
\]

(4.108)

\[
\geq (V_i + x_i)^{\gamma_i} \left[ \frac{R_i V_i + x_i \theta^*_i}{V_i + x_i} - R^* \right].
\]

(4.109)

Here in Eq. (4.108) we apply Eq. (4.106), and Eq. (4.109) holds because \(x_i/(V_i + x_i) \leq 1\).

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 \(\Delta\) for \(\Delta \geq 0\).
Proof. Note that \( h_i(\Delta) = (V_i + \Delta)^{\gamma_i-1}(R_iV_i + \Delta \theta_i^*) - (V_i + \Delta)^{\gamma_i} R^* \). Differentiating \( h_i \) with respect to \( \Delta \) we have

\[
h'_i(\Delta) = (\gamma_i - 1)(V_i + \Delta)^{\gamma_i-2}(R_iV_i + \Delta \theta_i^*) + \theta_i^*(V_i + \Delta)^{\gamma_i-1} - \gamma_i(V_i + \Delta)^{\gamma_i-1} R^*. \tag{4.110}
\]

Using the second property of Lemma 56 that \( \theta_i^* \geq \gamma_i R^* + (1 - \gamma_i) R_i(S_i^*) \), we have for all \( \Delta \geq 0 \) that

\[
h'_i(\Delta) \geq (\gamma_i - 1)(V_i + \Delta)^{\gamma_i-2}(R_iV_i + \Delta \theta_i^*) + \gamma_i R^* + (1 - \gamma_i) R_i \Delta^{\gamma_i-1}
\]

\[
= (1 - \gamma_i)(V_i + \Delta)^{\gamma_i-2}[R_i(V_i + \Delta) - R_iV_i - \Delta \theta_i^*]
\]

\[
= (1 - \gamma_i)(V_i + \Delta)^{\gamma_i-2} \cdot (R_i - \theta_i^*) \Delta \geq 0. \tag{4.112}
\]

The proposition is then proved, because \( h'_i(\Delta) \geq 0 \) for all \( \Delta \geq 0 \). \( \square \)

Invoking Proposition 19, we have that for all \( i \in [M] \),

\[
\tilde{V}_i^{\gamma_i} \left[ \tilde{R}_i - (R^* - \delta) \right] \geq (V_i + x_i)^{\gamma_i} \frac{R_iV_i + x_i \theta_i^*}{V_i + x_i} - R^* \geq V_i^{\gamma_i} [R_i - R^*]. \tag{4.114}
\]

Summing over \( i \in [M] \) on both sides of the above inequality and using the definition that \( R^* = (\sum_{i \in [M]} R_iV_i^{\gamma_i})/(1 + \sum_{i \in [M]} V_i^{\gamma_i}) \),

\[
\sum_{i \in [M]} \tilde{V}_i^{\gamma_i} \left[ \tilde{R}_i - (R^* - \delta) \right] \geq \sum_{i \in [M]} R_iV_i^{\gamma_i} - \left( \sum_{i \in [M]} V_i^{\gamma_i} \right) R^* = R^* \geq R^* - \delta. \tag{4.115}
\]

Re-organizing terms we have

\[
R'(\tilde{\theta}^*) = \frac{\sum_{i = 1}^{M} \phi_i \tilde{\theta}_i^* u_i \tilde{\theta}_i^*}{1 + \sum_{i = 1}^{M} u_i \tilde{\theta}_i^*} = \frac{\sum_{i \in [M]} R_i(L_i(\tilde{\theta}_i^*))V_i(L_i(\tilde{\theta}_i^*))^{\gamma_i}}{1 + \sum_{i \in [M]} V_i(L_i(\tilde{\theta}_i^*))^{\gamma_i}} \geq \frac{\sum_{i \in [M]} \tilde{R}_i \tilde{V}_i^{\gamma_i}}{1 + \sum_{i \in [M]} \tilde{V}_i^{\gamma_i}} \geq R^* - \delta,
\]

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 $\rho$ is set to $(1 + \sqrt{2}) = 0.694774$.

<table>
<thead>
<tr>
<th></th>
<th>Type A Nest</th>
<th>Type B Nest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Revenues $r_{ij}$</td>
<td>$1$</td>
<td>$\rho$</td>
</tr>
<tr>
<td>Preferences $v_{ij}$</td>
<td>$(1 + \epsilon)/M^2$</td>
<td>$(1 - \epsilon)/M^2$</td>
</tr>
</tbody>
</table>

**Lemma 76.** Let $U \subseteq [M]$ be the set of Type A nests, and by construction $[M] \setminus U$ are all Type B nests. For any $S = (S_1, \cdots, S_M) \subseteq [N]^M$, define $m^U_T(S) := \sum_{i \in U} 1\{S_i \neq \{1, 2\} \} + \sum_{i \notin U} 1\{S_i \neq \{1, 2, 3\} \}$. Then there exists a numerical constant $C > 0$ such that for all $S$, $R(S^*) - R(S) \geq m^U_T(S) \cdot C/\rho$, 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 $\pi$, we want to show a lower bound on the worst-case regret

$$\sup_{\{r_{ij}, v_{ij}\}} \sum_{t=1}^T R^* - \mathbb{E}^\pi \left[ R(S^{(t)}) \right].$$

Let $M_0 = M/4$ be an integer (because $M$ is divisible by 4) and $S_{M_0}$ 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 $[M] \setminus U$ are Type B. The following inequalities show a reduction to average-case regret:

$$\sup_{\{r_{ij}, v_{ij}\}} \sum_{t=1}^T R^* - \mathbb{E}^\pi \left[ R(S^{(t)}) \right] \geq \sup_{U \subseteq S_{M_0}} \sum_{t=1}^T R^* - \mathbb{E}^\pi_U \left[ R(S^{(t)}) \right] \geq \frac{1}{|S_{M_0}|} \sum_{U \subseteq S_{M_0}} \sum_{t=1}^T R^* - \mathbb{E}^\pi_U \left[ R(S^{(t)}) \right].$$

Here we use the $\mathbb{E}^\pi_U$ 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 $\pi$ itself.

For any $i \in [M]$ and $S \subseteq [N]$, denote $n_S(i) := \sum_{t=1}^T 1\{S^{(t)}_i = S\}$ as the random variable of the number of times assortment $S$ is offered in nest $i$. Let $\mathbb{E}^\pi_U[n_S(i)]$ be the expectation of $n_S(i)$, with expectation taken under model parameters setting $U$ (recall that $U$ is the set of all Type A nests) and policy $\pi$. Invoking Lemma 76 and noting that $\sum_{i=1}^M \sum_{S \subseteq [N]} \mathbb{E}^\pi_U[n_S(i)] = MT$ for any $U \subseteq [M]$ and policy $\pi$, the right-hand side of Eq. (4.117) can be lower bounded by

$$\frac{1}{|S_{M_0}|} \sum_{U \subseteq S_{M_0}} \sum_{t=1}^T \mathbb{E}^\pi_U \left[ m^U_T(S^{(t)}) \cdot C/\rho \right] = \frac{C}{\rho} \sum_{U \subseteq S_{M_0}} \left[ \sum_{i \in U} \sum_{S \neq \{1, 2\}} \mathbb{E}^\pi_U[n_S(i)] + \sum_{S \neq \{1, 2, 3\}} \mathbb{E}^\pi_U[n_S(i)] \right].$$
Then for any (see e.g., Csiszar & Körner (2011); Tsybakov (2009)). Note that in the last term

We upper bound $\mathbb{E}_{\mathcal{U}' \cup \{i\}}[n_{1,2}(i)]$ and $\mathbb{E}_{(\mathcal{U}' \cup \{i\})^c}[n_{1,2,3}(i)]$ by comparing them with $\mathbb{E}_{\mathcal{U}'}[n_{1,2}(i)]$ and $\mathbb{E}_{(\mathcal{U}')^c}[n_{1,2,3}(i)]$. In particular, let $P_{\mathcal{U}}^\pi$ denote the probabilistic law under $\mathcal{U}$ and policy $\pi$. Then for any $S \subseteq [N]$,

$$\mathbb{E}_{\mathcal{U}}[n_{S}(i)] - \mathbb{E}_{\mathcal{W}}[n_{S}(i)] \leq \sum_{j=0}^{T} j \cdot |P_{\mathcal{U}}^\pi[n_{S}(i) = j] - P_{\mathcal{W}}^\pi[n_{S}(i) = j]|$$

$$\leq T \cdot \sum_{j=0}^{T} |P_{\mathcal{U}}^\pi[n_{S}(i) = j] - P_{\mathcal{W}}^\pi[n_{S}(i) = j]|$$

$$= T \|P_{\mathcal{U}}^\pi - P_{\mathcal{W}}^\pi\|_{TV} \leq T \sqrt{\frac{1}{2} \min \{KL(P_{\mathcal{U}}^\pi \| P_{\mathcal{W}}^\pi), KL(P_{\mathcal{W}}^\pi \| P_{\mathcal{U}}^\pi)\}}$$

$$\leq T \sqrt{\frac{T}{2} \min \{\max_{S} KL(P_{\mathcal{U}}(\cdot|S) \| P_{\mathcal{W}}(\cdot|S)), \max_{S} KL(P_{\mathcal{W}}(\cdot|S) \| P_{\mathcal{U}}(\cdot|S))\}}.$$
not have superscript $\pi$, because conditioned on a particular assortment combination $S$ the KL divergence no longer depends on $\pi$.

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 $S = (S_1, \ldots, S_M)$.

**Lemma 77.** Suppose $|U \Delta W| = 1$, where $U \Delta W = (U \setminus W) \cup (W \setminus U)$ denotes the symmetric difference between subsets $U, W \subseteq [M]$. Then there exists a constant $C' > 0$ such that for any $S = (S_1, \ldots, S_M)$, $\min\{KL(P_U(\cdot|S)||P_W(\cdot|S)), KL(P_W(\cdot|S)||P_U(\cdot|S))\} \leq C' \epsilon^2 / M$.

Invoking Lemma 77, the right-hand side of Eq. (4.121) can be further upper bounded by

$$T \sqrt{\frac{T}{2} \cdot \frac{C' \epsilon^2}{M}} \leq T \sqrt{T \epsilon^2 / M}. \quad (4.122)$$

Invoking Eq. (4.119) and noting that $|S_{M_0-1}| = (\frac{M}{M_0-1}) = \frac{M_0}{M_0+1}(\frac{M}{M_0}) \leq |S_{M_0}|/3$ and $\sum_{i=1}^{M} \mathbb{E}_{U}[n_S(i)] \leq MT$ for all $U' \subseteq [M]$ and $S \subseteq [N]$, we have

$$\frac{1}{|S_{M_0}|} \sum_{U \in S_{M_0}} \sum_{i=1}^{T} R^* - \mathbb{E}_{U}[R(S^{(i)})] \geq C \epsilon \cdot T - \frac{C \epsilon}{|S_{M_0}|} \sum_{U \in S_{M_0}} \frac{1}{M} \sum_{U \neq U'} \left( \mathbb{E}_{U''}[n_{S_{1,2}}(i)] + \mathbb{E}_{(U' \cup \{i\})}[n_{S_{1,2,3}}(i)] \right)$$

$$\geq C \epsilon \cdot T - \frac{C \epsilon}{3} \sup_{U \in S_{M_0}} \frac{1}{M} \sum_{i=1}^{M} \left( \mathbb{E}_{U'}[n_{S_{1,2}}(i)] + \mathbb{E}_{(U' \cup \{i\})}[n_{S_{1,2,3}}(i)] \right)$$

$$\geq C \epsilon \cdot T - \frac{C \epsilon}{3} \left( \frac{1}{M} \cdot 2MT + \frac{1}{M} \cdot M \cdot O \left( T \sqrt{T \epsilon^2 / M} \right) \right). \quad (4.123)$$

Setting $\epsilon = \epsilon_0 \sqrt{M/T}$ for some sufficiently small positive constant $\epsilon_0 > 0$, we complete the proof of Theorem 18.

**Proof of Lemma 76** For any $U \subseteq [M]$, $S \subseteq \{1, 2, 3\}$ and $S = (S_1, \ldots, S_M) \in [3]^M$, define $m_{U,S}(S) := \sum_{i \in U} 1\{S_i = S\}$ and similarly $m_{U',S}(S) := \sum_{i \notin U} 1\{S_i = S\}$. Denote also $S^* = (S_1^*, \ldots, S_M^*)$ 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 $R_U(\cdot), V_U(\cdot), R_{U^*}(\cdot), V_{U^*}(\cdot)$ be revenue and preference of assortment selections in nests of Type A ($R_U(\cdot)$ and $V_U(\cdot)$) or Type B ($R_{U^*}(\cdot)$ and $V_{U^*}(\cdot)$), respectively. Recall that $|U| = M/4$ and $|U^*| = 3M/4$. We then have

$$R(S^*) - R(S) = \frac{R_U(\{1, 2\})V_U(\{1, 2\})^{1/2} \cdot M/4 + R_{U^*}(\{1, 2, 3\})V_{U^*}(\{1, 2, 3\})^{1/2} \cdot 3M/4}{1 + V_U(\{1, 2\})^{1/2} \cdot M/4 + V_{U^*}(\{1, 2, 3\})^{1/2} \cdot 3M/4}$$

$$- \frac{\sum_{S \subseteq \{1, 2, 3\}} m_{U,S}(S) \cdot R_U(S)V_U(S)^{1/2} + m_{U^*,S}(S) \cdot R_{U^*}(S)V_{U^*}(S)^{1/2}}{1 + \sum_{S \subseteq \{1, 2, 3\}} m_{U,S}(S) \cdot V_U(S)^{1/2} + m_{U^*,S}(S) \cdot V_{U^*}(S)^{1/2}}. \quad (4.124)$$
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.

$S = \emptyset$ : $V_U(S)^{1/2} = 0$, $R_U(S) = 0$, $V_{U^c}(S)^{1/2} = 0$, $R_{U^c}(S) = 0$;

$S = \{1\}$ : $V_U(S)^{1/2} = \frac{\sqrt{1 + \epsilon}}{M}$, $R_U(S) = 1$, $V_{U^c}(S)^{1/2} = \frac{\sqrt{1 - \epsilon}}{M}$, $R_{U^c}(S) = 1$;

$S = \{2\}$ : $V_U(S)^{1/2} = \frac{\sqrt{1 - \epsilon}}{M}$, $R_U(S) = 0.8$, $V_{U^c}(S)^{1/2} = \frac{\sqrt{1 + \epsilon}}{M}$, $R_{U^c}(S) = 0.8$;

$S = \{3\}$ : $V_U(S)^{1/2} = \frac{1}{M}$, $R_U(S) = \rho$, $V_{U^c}(S)^{1/2} = \frac{1}{M}$, $R_{U^c}(S) = \rho$;

$S = \{1, 2\}$ : $V_U(S)^{1/2} = \frac{\sqrt{2}}{M}$, $R_U(S) = .9 + .1\epsilon$, $V_{U^c}(S)^{1/2} = \frac{\sqrt{2}}{M}$, $R_{U^c}(S) = .9 - .1\epsilon$;

$S = \{1, 3\}$ : $V_U(S)^{1/2} = \frac{\sqrt{1 + \epsilon}}{M}$, $R_U(S) = \frac{1 + \rho + \epsilon}{2 + \epsilon}$,

$V_{U^c}(S)^{1/2} = \frac{\sqrt{1 - \epsilon}}{M}$, $R_{U^c}(S) = \frac{1 + \rho - \epsilon}{2 - \epsilon}$;

$S = \{2, 3\}$ : $V_U(S)^{1/2} = \frac{\sqrt{1 - \epsilon}}{M}$, $R_U(S) = \frac{.8 + \rho - .8\epsilon}{2 - \epsilon}$,

$V_{U^c}(S)^{1/2} = \frac{\sqrt{1 + \epsilon}}{M}$, $R_{U^c}(S) = \frac{.8 + \rho + .8\epsilon}{2 + \epsilon}$;

$S = \{1, 2, 3\}$ : $V_U(S)^{1/2} = \frac{\sqrt{3}}{M}$, $R_U(S) = \frac{1.8 + \rho + .2\epsilon}{3}$,

$V_{U^c}(S)^{1/2} = \frac{\sqrt{3}}{M}$, $R_{U^c}(S) = \frac{1.8 + \rho - .2\epsilon}{3}$.

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 $\text{KL}(P_U(\cdot|S)||P_W(\cdot|S))$ can then be written as

$$\text{KL}(P_U(\cdot|S)||P_W(\cdot|S)) = -\mathbb{E}_U \left[ \log \frac{P_W(i,j|S)}{P_U(i,j|S)} \right] = -\mathbb{E}_U \left[ \log \frac{P_W(i|S)}{P_U(i|S)} \right] - \mathbb{E}_U \left[ \log \frac{P_W(j|i,S)}{P_U(j|i,S)} \right].$$

(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)), $\text{KL}(p||q)$.
can be upper bounded as
\[
\text{KL}(p\|q) = -\sum_{i=0}^{M} p_i \log \frac{q_i}{p_i} \leq \sum_{i=0}^{M} \frac{|p_i - q_i|^2}{q_i}.
\]

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 \(|p_i - q_i| \leq \epsilon/M\) for \(i \in \{0, i_0\}\), \(|p_i - q_i| \leq \epsilon/M^2\) if \(i \notin \{0, i_0\}\), \(q_0 \geq \Omega(1)\) and \(q_i \geq 1/M\) for all \(i \geq 1\). Subsequently,
\[
\text{KL}(p\|q) \leq \epsilon^2/M. \tag{4.126}
\]

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 \(\text{KL}(P_U(\cdot|i_0, S_{i_0})\|P_W(\cdot|i_0, S_{i_0})) \leq \epsilon^2\) for all \(S_{i_0} \subseteq [N]\). In addition, \(\max\{P_U(i_0|S), P_W(i_0|S)\} \leq 1/M\). Subsequently,
\[
-\mathbb{E}_U \left[ \log \frac{P_W(j|i, S)}{P_U(j|i, S)} \right] = P_U(i_0|S) \cdot \text{KL}(P_U(\cdot|i_0, S_{i_0})\|P_W(\cdot|i_0, S_{i_0})) \leq \epsilon^2/M. \tag{4.127}
\]

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 \(\theta^*\) obtained from the pure exploration phase of Algorithm 13, and show as a corollary that the true model \(\theta_0\) is feasible to all subsequent local MLE formulations with high probability (see Corollary 10). In the second step, we use an \(\epsilon\)-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 \(\hat{\theta}_t - \theta_0\) implies an upper bound on the estimation error of the expected revenue \(R_t(S)\), hence showing that \(R_t(S)\) 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 \(\theta^*\), 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
\[
\|\theta^* - \theta_0\|_2 \leq \frac{2}{\kappa \sqrt{d + \log(1/\delta)/\lambda_{\min}(V)}} \quad \text{where} \quad \kappa = \frac{1}{2e(1 + \rho)} \quad \text{and} \quad V = \sum_{t=1}^{T_0} v_{t,i}v_{t,i}^T.
\] (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 $\inf_{|x| \leq 1} |\theta - \theta_0| \leq 1 \geq 2e(1 + \rho)$ where $\sigma(x) = 1/(1 + e^{-x})$ is the sigmoid link function. Because $\sigma'(x) = \sigma(x)(1 - \sigma(x))$, we have $\sigma'(x^\top) = \varphi(1 - \varphi) \geq 0.5\varphi$ where $\varphi = \min\{p_\theta(1), 1 - p_\theta(1)\}$ and $p_\theta(1) = \sigma(x^\top) = 1/(1 + \exp\{-x^\top\})$. By (A2), we know that $\varphi_\theta \geq 1/(1 + \rho)$. Subsequently, for any $\|x\|_2 \leq 1$ and $\|\theta - \theta_0\|_2 \leq 1$, we have
\[
\varphi \geq \frac{1}{1 + \exp\{-x^\top\}} = \frac{1}{1 + \exp\{-x^\top(\theta - \theta_0)\}} \exp\{-x^\top\theta_0\} \geq \frac{1}{e} \frac{1}{1 + \exp\{x^\top\theta_0\}} \geq \frac{1}{e(1 + \rho)}.
\]

Lemma 78 is then an immediate consequence of (Li et al., 2017b, Eq. (18)). \hfill \Box

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 $T_0 \geq C_0 \max\{\nu^2d \log T/\lambda_0^2, \rho^2(d + \log T)/(\tau^2\lambda_0)\}$ then with probability $1 - O(T^{-1})$, $\|\theta^* - \theta_0\|_2 \leq \tau$.

Proof. Denote $\Lambda := \mathbb{E}_\mu x^\top$ and $\Lambda := V/T_0 = \frac{1}{T_0} \sum_{t=1}^{T_0} x_{t,i}x_{t,i}^\top$. Clearly $\mathbb{E}\Lambda = \Lambda$. In addition, because $\|v_{t,j}\|_2 \leq \nu$ almost surely, $v_{t,j}$ are sub-Gaussian random variables with parameter $\nu^2$. By standard concentration inequalities (see, e.g., (Vershynin, 2012, Proposition 2.1)), we have with probability $1 - O(T^{-2})$ that $\|\Lambda - \Lambda\|_{op} \leq \nu \sqrt{d \log T/T_0}$. Hence, if $T_0 \geq C_0 \nu^2 d \log T/\lambda_0^2$ for some sufficiently large universal constant $C_0$, we have $\|\Lambda - \Lambda\|_{op} \leq 0.5\lambda_0 = \lambda_{\min}(\Lambda)$ and therefore $\lambda_{\min}(V) = T_0 \lambda_{\min}(\Lambda) \geq 0.5T_0\lambda_0$. The corollary then immediately follows Lemma 78. \hfill \Box

The purpose of Corollary 10 is to establish a connection between the number of pure exploration iterations $T_0$ and the critical radius $\tau$ 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 $\tau$ with high probability, which certifies that the true model $\theta_0$ is also a feasible local estimator in our MLE-UCB policy. This is an important property for later analysis of local MLE solutions $\hat{\theta}_{t-1}$.

Analysis of the local MLE

The following lemma upper bounds a Mahalanobis distance between $\hat{\theta}_t$ and $\theta_0$. For convenience, we adopt the notation that $r_{t0} = 0$ and $v_{t0} = 0$ for all $t$ throughout this section. We also define
\[
I_t(\theta) := \sum_{t' = 1}^{t} M_{t'}(\theta), \quad (4.129)
\]
\[
M_{t'}(\theta) := -\mathbb{E}_{\theta_0, t} [\nabla_{\theta}^2 \log p_{\theta, t'}(j|S_{t'})]
\]
Lemma 79. Suppose $\tau \leq 1/\sqrt{8\nu v^2 K^2}$. Then there exists a universal constant $C > 0$ such that with probability $1 - O(T^{-1})$ the following holds uniformly over all $t = T_0, \ldots, T - 1$:

$$
(\hat{\theta}_t - \theta_0)^\top I_t(\theta_0)(\hat{\theta}_t - \theta_0) \leq C \cdot d \log(\nu/TK).
$$

(4.130)

Remark 29. For $\theta = \theta_0$, the expression of $M_v(\theta)$ can be simplified as $M_v(\theta_0) = \mathbb{E}_{\theta_0,v}[v_{\ell}^\top v_{\ell}^\top] - \{\mathbb{E}_{\theta_0,v}v_{\ell}v_{\ell}^\top\}^\top$.

We next state the proof of Lemma 79. For any $\theta \in \mathbb{R}^d$ define

$$
f_v(\theta) := \mathbb{E}_{\theta_0,v} \left[ \log \frac{p_{\theta,v}(j|S_v)}{p_{\theta_0,v}(j|S_v)} \right] = \sum_{j \in S_v \setminus \{0\}} p_{\theta_0,v}(j|S_v) \log \frac{p_{\theta,v}(j|S_v)}{p_{\theta_0,v}(j|S_v)}.
$$

By simple algebra calculations, the first and second order derivatives of $f_v$ with respect to $\theta$ can be computed as

$$
\nabla_\theta f_v(\theta) = \mathbb{E}_{\theta_0,v}[v_{\ell}] - \mathbb{E}_{\theta,v}[v_{\ell}];
$$

$$
\nabla_\theta^2 f_v(\theta) = -\mathbb{E}_{\theta_0,v}[v_{\ell}v_{\ell}^\top] + \{\mathbb{E}_{\theta_0,v}v_{\ell}v_{\ell}^\top\}^\top
+ \{\mathbb{E}_{\theta,v}v_{\ell}v_{\ell}^\top\}^\top - \{\mathbb{E}_{\theta,v}v_{\ell}v_{\ell}^\top\}^\top.
$$

(4.131)

(4.132)

In the rest of the section we drop the subscript in $\nabla_\theta$, $\nabla_\theta^2$, and the $\nabla$, $\nabla^2$ notations should always be understood as with respect to $\theta$.

Define $F_t(\theta) := \sum_{\ell=1}^t f_v(\theta)$. It is easy to verify that $-F_t(\theta)$ is the Kullback-Leibler divergence between the conditional distribution of $(i_1, \ldots, i_t)$ parameterized by $\theta$ and $\theta_0$, respectively. Therefore, $F_t(\theta)$ is always non-positive. Note also that $F_t(\theta_0) = 0$, $\nabla F_t(\theta_0) = 0$, $\nabla^2 f_v(\theta) = -M_v(\theta)$ and $\nabla^2 F_t(\theta) = -I_t(\theta)$. By Taylor expansion with Lagrangian remainder, there exists $\tilde{\theta}_t = \alpha \theta_0 + (1 - \alpha) \hat{\theta}_t$ for some $\alpha \in (0, 1)$ such that

$$
F_t(\hat{\theta}_t) = \frac{1}{2}(\hat{\theta}_t - \theta_0)^\top I_t(\theta_0)(\hat{\theta}_t - \theta_0).
$$

(4.133)

Our next lemma shows that, if $\tilde{\theta}_t$ is close to $\theta_0$ (guaranteed by the constraint that $\|\hat{\theta}_t - \theta^*\|_2 \leq \tau$), then $I_t(\tilde{\theta}_t)$ can be spectrally lower bounded by $I_t(\theta_0)$. It is proved in the supplementary material.

Lemma 80. Suppose $\tau \leq 1/\sqrt{8\nu v^2 K^2}$. Then $I_t(\tilde{\theta}_t) \geq \frac{1}{2} I_t(\theta_0)$ for all $t$.

Proof. Because $\hat{\theta}_t$ is a feasible solution of the local MLE, we know $\|\hat{\theta}_t - \theta^*\|_2 \leq \tau$. Also by Corollary 10 we know that $\|\theta^* - \theta_0\|_2 \leq \tau$ with high probability. By triangle inequality and the definition of $\tilde{\theta}_t$ we have that $\|\tilde{\theta}_t - \theta_0\|_2 \leq 2\tau$.

166
To prove $I_t(\tilde{\theta}_t) \geq \frac{1}{2} L(\theta_0)$ we only need to show that $M_{t'}(\tilde{\theta}_t) - M_t(\theta_0) \leq \frac{1}{2} M_{t'}(\theta_0)$ for all $1 \leq t' \leq t$. This reduces to proving

$$\left\{ \mathbb{E}_{t,v} v_{t,j} - \mathbb{E}_{t,v} v_{t,j} \right\} \left\{ \mathbb{E}_{t,v} v_{t,j} - \mathbb{E}_{t,v} v_{t,j} \right\}^\top \leq \frac{1}{2} \mathbb{E}_{t,v} \left[ (v_{t,j} - \mathbb{E}_{t,v} v_{t,j})(v_{t,j} - \mathbb{E}_{t,v} v_{t,j})^\top \right].$$

(4.134)

Fix arbitrary $S_{t'} \subseteq [N]$, $|S_{t'}| = J \leq K$ and for convenience denote $x_1, \ldots, x_J \in \mathbb{R}^d$ 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 $p_{\tilde{\theta}_t}(j)$ be the probability of choosing action $j \in [J]$ corresponding to $x_j$ parameterized by $\theta_0$ or $\tilde{\theta}_t$. Define $\bar{x} := \sum_{j=1}^J p_{\theta_0}(j) x_j$, $w_j := x_j - \bar{x}$ and $\delta_j := p_{\tilde{\theta}_t}(j) - p_{\theta_0}(j)$. Recall also that $x_0 = 0$ and $w_0 = -\bar{x}$. Eq. (4.134) is then equivalent to

$$\left\{ \sum_{j=0}^J \delta_j w_j \right\} \left\{ \sum_{j=0}^J \delta_j w_j \right\}^\top \leq \frac{1}{2} \sum_{j=0}^J p_{\theta_0}(j) w_j w_j^\top. \tag{4.135}$$

Let $L = \text{span}\{w_j\}_{j=0}^J$ and $H \in \mathbb{R}^{L \times d}$ be a whitening matrix such that $H (\sum_{j=0}^J p_{\theta_0}(j) w_j w_j^\top) H^\top = I_{L \times L}$, where $I_{L \times L}$ is the identity matrix of size $L$. Denote $\tilde{w}_j := H w_j$. We then have $\sum_{j=0}^J p_{\theta_0}(j) \tilde{w}_j \tilde{w}_j^\top = I_{L \times L}$. Eq. (4.135) is then equivalent to

$$\left\| \sum_{j=0}^J \delta_j \tilde{w}_j \right\|_2^2 \leq \frac{1}{2}. \tag{4.136}$$

On the other hand, by (A2) we know that $p_{\theta_0}(j) \geq 1/\rho K$ for all $j$ and therefore $\|\tilde{w}_j\|_2 \leq \sqrt{\rho K}$ for all $j$. Subsequently, we have

$$\left\| \sum_{j=0}^J \delta_j \tilde{w}_j \right\|_2^2 \leq \left( \max_j |\delta_j| \cdot \sum_{j=0}^J \|\tilde{w}_j\|_2 \right)^2 \leq \max_j |\delta_j|^2 \cdot \rho K^2. \tag{4.137}$$

Recall that $\delta_i = p_{\tilde{\theta}_t}(i) - p_{\theta_0}(i)$ where $p_{\theta_0}(i) = \exp\{x_i^\top \theta\} / (1 + \sum_{j \in S_{t'}} \exp\{x_j^\top \theta\})$. Simple algebra yields that $\nabla_{\theta} p_{\theta_0}(i) = p_{\theta_0}(i) [x_i - \mathbb{E}_{\theta} x_j]$, where $\mathbb{E}_{\theta} x_j = \sum_{j \in S_{t'}} p_{\theta_0}(j) x_j$. Using the mean-value theorem, there exists $\tilde{\theta}_t = \tilde{\alpha} \tilde{\theta}_t + (1 - \tilde{\alpha}) \theta_0$ for some $\tilde{\alpha} \in (0, 1)$ such that

$$\delta_i = \langle \nabla_{\theta} p_{\theta_0}(i), \tilde{\theta}_t - \theta_0 \rangle = p_{\theta_0}(i) \langle x_i - \mathbb{E}_{\tilde{\theta}_t} x_j, \tilde{\theta}_t - \theta_0 \rangle. \tag{4.138}$$

Because $\|x_i\|_2 \leq \nu$ almost surely for all $t \in [T]$ and $i \in [N]$, we have

$$\max_j |\delta_j|^2 \cdot \rho K^2 \leq 4 \cdot \max_i \|x_i\|_2^2 \cdot \|\tilde{\theta}_t - \theta_0\|_2^2 \cdot \rho K^2 \leq 4 \nu^2 K^2 \cdot \tau^2. \tag{4.139}$$

The lemma is then proved by plugging in the condition on $\tau$. \hfill \square

As a corollary of Lemma 80, we have

$$F_t(\tilde{\theta}_t) \leq -\frac{1}{4} (\tilde{\theta}_t - \theta_0)^\top I_t(\theta_0) (\tilde{\theta}_t - \theta_0). \tag{4.140}$$
On the other hand, consider the “empirical” version \( \hat{F}_t(\theta) := \sum_{i=1}^{t} \hat{f}_i(\theta) \), where

\[
\hat{f}_i(\theta) := \log \frac{p_{\theta, i}(\theta \mid S_i)}{p_{\theta_i, i}(\theta \mid S_i)}.
\] (4.141)

It is easy to verify that \( \hat{F}_t(\theta_0) = 0 \) remains true; in addition, for any fixed \( \theta \in \mathbb{R}^d \), \( \{\hat{F}_t(\theta)\}_t \) forms a martingale \(^5\) and satisfies \( \mathbb{E} \hat{F}_t(\theta) = F_t(\theta) \) for all \( t \). This leads to our following lemma, which upper bounds the uniform convergence of \( \hat{F}_t(\theta) \) towards \( F_t(\theta) \) for all \( \|\theta - \theta_0\| \leq 2\tau \).

**Lemma 81.** Suppose \( \tau \leq 1/\sqrt{8\rho^2 \nu^2 K^2} \). 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 \( \|\theta - \theta_0\| \leq 2\tau \):

\[
|F_t(\theta)| \leq C \left[ \log(TK) + \sqrt{F_t(\theta) \log(TK)} \right]. \tag{4.142}
\]

**Proof.** We first consider a fixed \( \theta \in \mathbb{R}^d \), \( \|\theta - \theta_0\| \leq 2\tau \). Define

\[
\mathcal{M} := \max_{t' < t} \left| \hat{f}_{t'}(\theta) \right| \quad \text{and} \quad \mathcal{V}^2 := \sum_{t' = 1}^{t} \mathbb{E}_{j \sim \theta_{t'}, \theta} \left| \log \frac{p_{\theta, i}(j \mid S_{t'})}{p_{\theta_{t'}, i}(j \mid S_{t'})} \right|^2. \tag{4.143}
\]

Using an Azuma-Bernstein type inequality (see, for example, (Fan et al., 2015, Theorem A), (Freedman, 1975, Theorem (1.6))), we have

\[
|F_t(\theta) - F_t(\theta)| \leq \mathcal{M} \log(1/\delta) + \sqrt{\mathcal{V}^2 \log(1/\delta)} \quad \text{with probability } 1 - \delta. \tag{4.144}
\]

The following lemma upper bounds \( \mathcal{M} \) and \( \mathcal{V}^2 \) using \( F_t(\theta) \) and the fact that \( \theta \) is close to \( \theta_0 \). It will be proved right after this proof.

**Lemma 82.** If \( \tau \leq 1/\sqrt{8\rho^2 \nu^2 K^2} \) then \( \mathcal{M} \leq 1 \) and \( \mathcal{V}^2 \leq 8|F_t(\theta)| \).

**Corollary 11.** Suppose \( \tau \) satisfies the condition in Lemma 82. Then for any \( \|\theta - \theta_0\| \leq 2\tau \)

\[
|\hat{F}_t(\theta) - F_t(\theta)| \leq \log(1/\delta) + \sqrt{|F_t(\theta)| \log(1/\delta)} \quad \text{with probability } 1 - \delta. \tag{4.145}
\]

Our next step is to construct an \( \epsilon \)-net over \( \{\theta \in \mathbb{R}^d : \|\theta - \theta_0\| \leq 2\tau \} \) and apply union bound on the constructed \( \epsilon \)-net. This together with a deterministic perturbation argument delivers uniform concentration of \( \hat{F}_t(\theta) \) towards \( F_t(\theta) \).

For any \( \epsilon > 0 \), let \( \mathcal{H}(\epsilon) \) be a finite covering of \( \{\theta \in \mathbb{R}^d : \|\theta - \theta_0\| \leq 2\tau \} \) in \( \|\cdot\|_2 \) up to precision \( \epsilon \). That is, \( \sup_{\|\theta - \theta_0\| \leq 2\tau} \min_{\theta \in \mathcal{H}(\epsilon)} \|\theta - \theta'\| \leq \epsilon \). 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 \( \log |\mathcal{H}(\epsilon)| \leq d \log(\tau/\epsilon) \). Subsequently, by Corollary 11 and the union bound, we have with probability \( 1 - O(T^{-1}) \) that

\[
|\hat{F}_t(\theta) - F_t(\theta)| \leq d \log(T/\epsilon) + \sqrt{|F_t(\theta)| d \log(T/\epsilon)} \quad \forall T_0 < t \leq T, \theta \in \mathcal{H}(\epsilon). \tag{4.146}
\]

\(^5\{X_k\}_k\) 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 \(\|\theta - \theta'\|_2 \leq \epsilon\) that

\[
|\hat{F}_t(\theta) - \hat{F}_t(\theta')| \leq t \cdot \sup_{\nu \in S_t' \cup \{0\}} \left| \log \frac{p_{\theta,\nu}(j|S')}{p_{\theta',\nu}(j|S')} \right| \leq \frac{d \log(\rhoTK)}{\mathbb{E}F_t(\theta)|d \log(\rhoTK)}.
\]  

(4.147)

Here Eq. (4.147) holds because \(\log(1 + x) \leq x\); Eq. (4.148) holds because \(p_{\theta,\nu}(j|S') \geq p_{\theta_0,\nu}(j|S') - |p_{\theta,\nu}(j|S') - p_{\theta_0,\nu}(j|S')| \geq 1/2\rho TK\) thanks to (A2) and Eq. (4.139).

Combining Eqs. (4.146,4.149) and setting \(\epsilon = 1/(\rho TK)\) we have with probability \(1 - O(T^{-1})\) that

\[
|\hat{F}_t(\theta) - F_t(\theta)| \leq d \log(\rho TK) + \sqrt{F_t(\theta)|d \log(\rho TK)} \quad \forall T_0 < t \leq T, \|\theta - \theta_0\|_2 \leq 2\tau,
\]  

(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 \(p_{\theta_0,\nu}(j|S') \geq 1/\rho TK\) for all \(j\). Also, Eqs. (4.138,4.139) shows that \(|p_{\theta,\nu}(j|S') - p_{\theta_0,\nu}(j|S')| \leq 4\nu^2 \cdot \tau^2\). If \(\tau^2 \leq 1/\sqrt{8\rho TK}\) we have \(|p_{\theta,\nu}(j|S') - p_{\theta_0,\nu}(j|S')| \leq 0.5p_{\theta_0,\nu}(j|S')\) and therefore \(\hat{F}_t(\theta) \leq \log 2^2 \leq 1\).

We next give upper bounds on \(V^2\). Fix arbitrary \(t'\), and for notational simplicity let \(p_j = p_{\theta_0,\nu}(j|S')\) and \(q_j = p_{\theta,\nu}(j|S')\). Because \(\log(1 + x) \leq x\) for all \(x \in (-1, \infty)\), we have

\[
\mathbb{E}_{j \sim \theta_0,\nu} \left| \log \frac{p_{\theta,\nu}(j|S')}{p_{\theta_0,\nu}(j|S')} \right|^2 \leq \sum_{j \in S_t' \cup \{0\}} p_j \log^2 \left( 1 + \frac{q_j - p_j}{p_j} \right) \leq \sum_{j \in S_t' \cup \{0\}} \frac{(q_j - p_j)^2}{p_j}. \tag{4.151}
\]

On the other hand, by Taylor expansion we know that for any \(x \in (-1, \infty)\), there exists \(\bar{x} \in (0, x)\) such that \(\log(1 + x) = x - x^2/2(1 + \bar{x})^2\). Subsequently,

\[
-f'_{\nu}(\theta) = -\mathbb{E}_{j \sim \theta_0,\nu} \left[ \log \frac{p_{\theta,\nu}(j|S')}{p_{\theta_0,\nu}(j|S')} \right] = -\sum_{j \in S_t' \cup \{0\}} p_j \log \left( 1 + \frac{q_j - p_j}{p_j} \right)
\]  

(4.152)

\[
= -\sum_{j \in S_t' \cup \{0\}} p_j \left( \frac{q_j - p_j}{p_j} - \frac{1}{2(1 + \bar{\theta}_j)^2} \frac{(q_j - p_j)^2}{p_j^2} \right).
\]  

(4.153)

\[
\geq \frac{1}{2(1 + \max_j |p_j - q_j/p_j|^2)} \sum_{j \in S_t' \cup \{0\}} \frac{(q_j - p_j)^2}{p_j}.
\]  

(4.154)

Here \(\bar{\theta}_j \in (0, (q_j - p_j)/p_j)\) 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 $|q_j - p_j|^2 \leq 4\nu^2 \cdot \tau^2$. In addition, (A2) implies that $p_j \geq 1/\rho K$ for all $j$. Therefore, if $\tau \leq 1/\sqrt{4\rho^2\nu^2K^2}$ we have $|q_j - p_j|/p_j \leq 1$ for all $j$ and hence

$$
\mathbb{E}_{j \sim \theta_0,v} \left| \log \frac{p_{\theta,v}(j|S_v)}{p_{\theta_0,v}(j|S_v)} \right|^2 \leq \sum_{j \in S_v \cup [0]} \frac{(q_j - p_j)^2}{p_j} \leq 8f_v(\theta). \quad (4.155)
$$

Summing over all $t' = 1, \ldots, t$ and noting that $f_v(\theta)$ 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 $\hat{F}_t(\hat{\theta}_t) \leq 0 \leq F_t(\hat{\theta}_t)$, we have

$$
|F_t(\hat{\theta}_t)| \leq |\hat{F}_t(\hat{\theta}_t) - F_t(\hat{\theta}_t)| \leq d \log(\rho \nu TK) + \sqrt{|F_t(\hat{\theta}_t)|d \log(\rho \nu TK)}.
$$

Subsequently,

$$
|F_t(\hat{\theta}_t)| \leq d \log(\rho \nu NT). \quad (4.157)
$$

In addition, because $F_t(\hat{\theta}_t) \leq 0$, by Eq. (4.140) we have

$$
-\frac{1}{2}(\hat{\theta}_t - \theta_0)^T I_t(\theta_0)(\hat{\theta}_t - \theta_0) \geq F_t(\hat{\theta}_t) \geq d \log(\rho \nu TK).
$$

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 $\tau$ satisfies the condition in Lemma 79. With probability $1 - O(T^{-1})$ the following holds uniformly for all $t > T_0$ and $S \subseteq [N], |S| \leq K$ such that

1. $\overline{R}_t(S) \geq R_t(S)$;

2. $|\overline{R}_t(S) - R_t(S)| \leq \min\{1, \omega \sqrt{\|I_{t-1}^{-1/2}(\theta_0)M_t(\theta_0|S)I_{t-1}^{-1/2}(\theta_0\|_op}\}$.

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 $\varphi$ of parameters, i.e., $|\varphi(\hat{\theta}_{t-1}) - \varphi(\theta_0)|$ using the estimation error of the parameters themselves $\theta_{t-1} - \theta_0$.

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 $\tau$ 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 $R_t(S) := \mathbb{E}_{\theta_0,t}[\tau_{t,j}|S]$ and $\hat{R}_t(S) := \mathbb{E}_{\hat{\theta}_{t-1},t}[\hat{r}_{t,j}|S]$ using $I_{t-1}(\theta_0)$ and $M_t(\theta_0|S)$, where $I_{t-1}(\theta) = \sum_{j=1}^{t-1} M_{\nu}(\theta)$ and $M_{\nu}(\theta) = \mathbb{E}_{\theta,v}[v_{\nu,j}v_{\nu,j}^\top] - \mathbb{E}_{\theta,v}[v_{\nu,j}]\mathbb{E}_{\theta,v}[v_{\nu,j}]^\top - \mathbb{E}_{\theta,v}[v_{\nu,j}]\mathbb{E}_{\theta,v}[v_{\nu,j}]^\top + \mathbb{E}_{\theta,v}[v_{\nu,j}]\mathbb{E}_{\theta,v}[v_{\nu,j}]^\top$. 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 $|\tilde{R}_t(S) - R_t(S)| \leq \sqrt{d \log(\nu TK)} \cdot \sqrt{\|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0)|S| I_{t-1}^{-1/2}(\theta_0)\|_{op}}$, where in $\leq$ notation we only hide numerical constants.

Below we state our proof of Lemma 84. Fix $S \subseteq [N]$. We use $\mathfrak{R}_t(\theta) = \mathbb{E}_{\theta,t}[r_{tj}] = \frac{\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta\}}{1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta\}}$ to denote the expected revenue of assortment $S$ at time $t$, evaluated using a specific model $\theta \in \mathbb{R}$. Then

$$
\nabla_\theta \mathfrak{R}_t(\theta) = \frac{\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta\} (1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta\})^2 - \left(\sum_{j \in S} r_{tj} \exp\{v_{tj}^\top \theta\}\right) \left(\sum_{j \in S} \exp\{v_{tj}^\top \theta\}\right)}{(1 + \sum_{j \in S} \exp\{v_{tj}^\top \theta\})^2}
= \mathbb{E}_{\theta,t}[r_{tj} v_{tj}] - \{\mathbb{E}_{\theta,t} r_{tj}\} \{\mathbb{E}_{\theta,t} v_{tj}\}. \tag{4.159}
$$

By the mean value theorem, there exists $\theta_{-1} = \theta_0 + \xi(\theta_{-1} - \theta_0)$ for some $\xi \in (0,1)$ such that

$$
|\tilde{R}_t(S) - R_t(S)| = \left|\mathfrak{R}_t(\theta_{-1}) - \mathfrak{R}_t(\theta_0)\right| = \left|\langle \nabla \mathfrak{R}_t(\theta_{-1}), \hat{\theta}_{-1} - \theta_0 \rangle \right|
= \sqrt{(\hat{\theta}_{-1} - \theta_0)^\top \left[\nabla \mathfrak{R}_t(\theta_{-1}) \nabla \mathfrak{R}_t(\hat{\theta}_{-1})\right] \hat{\theta}_{-1} - \theta_0}. \tag{4.160}
$$

Recall that $\nabla \mathfrak{R}_t(\theta_{-1}) = \mathbb{E}_{\theta_{-1,t}}[r_{tj} v_{tj}] - \{\mathbb{E}_{\theta_{-1,t}} r_{tj}\} \{\mathbb{E}_{\theta_{-1,t}} v_{tj}\} = \mathbb{E}_{\theta_{-1,t}}[(r_{tj} - \mathbb{E}_{\theta_{-1,t}} r_{tj})(v_{tj} - \mathbb{E}_{\theta_{-1,t}} v_{tj})]$. Subsequently, by Jensen’s inequality and the fact that $r_{tj} \in [0,1]$ almost surely,

$$
\nabla \mathfrak{R}_t(\theta_{-1}) \nabla \mathfrak{R}_t(\theta_{-1})^\top \leq \mathbb{E}_{\theta_{-1,t}} \left[ (r_{tj} - \mathbb{E}_{\theta_{-1,t}} r_{tj}) (v_{tj} - \mathbb{E}_{\theta_{-1,t}} v_{tj})^\top \right] \leq \mathbb{E}_{\theta_{-1,t}} \left[ (v_{tj} - \mathbb{E}_{\theta_{-1,t}} v_{tj}) (v_{tj} - \mathbb{E}_{\theta_{-1,t}} v_{tj})^\top \right] = \hat{M}_t(\theta_{-1} | S). \tag{4.161}
$$

Define $\hat{M}_t(\theta | S) := \mathbb{E}_{\theta,t}[\{v_{tj} - \mathbb{E}_{\theta,t} v_{tj}\} \{v_{tj} - \mathbb{E}_{\theta,t} v_{tj}\}^\top]$, where $S \subseteq [N]$ is the assortment supplied at iteration $t$. Combining Eqs. (4.160,4.161) with Lemma 79, we have

$$
|\tilde{R}_t(S) - R_t(S)| \leq \sqrt{d \log(\nu TK)} \cdot \sqrt{\|I_{t-1}(\theta_0)^{-1/2} \hat{M}_t(\theta_{-1} | S) I_{t-1}(\theta_0)^{-1/2}\|_{op}}. \tag{4.162}
$$

It remains to show that $\hat{M}_t(\theta_{-1} | S)$ and $M_t(\theta_0 | S)$ are close, for which we first recall the definitions of both quantities:

$$
\hat{M}_t(\theta_{-1} | S) = \mathbb{E}_{\theta_{-1,t}} \left[ (v_{tj} - \mathbb{E}_{\theta_{-1,t}} v_{tj}) (v_{tj} - \mathbb{E}_{\theta_{-1,t}} v_{tj})^\top \right];
M_t(\theta_0 | S) = \mathbb{E}_{\theta_0,t}[v_{tj} v_{tj}] - \{\mathbb{E}_{\theta_0,t} v_{tj}\} \{\mathbb{E}_{\theta_0,t} v_{tj}\}^\top = \hat{M}_t(\theta_0 | S).
$$

The next lemma shows that under suitable conditions $\hat{M}_t(\theta_{-1} | S)$ is close to $\hat{M}_t(\theta_0 | S) = M_t(\theta_0 | S)$, implying that $\frac{1}{d} M_t(\theta_0 | S) \leq \hat{M}_t(\theta_{-1} | S) \leq 4 M_t(\theta_0 | S)$. It is proved in the supplementary material.

Lemma 85. Suppose $\tau \leq 1/\sqrt{8 \rho^2 v_2 K^2}$. Then $\frac{1}{d} M_t(\theta_0 | S) \leq \hat{M}_t(\theta_{-1} | S) \leq 4 M_t(\theta_0 | S)$ for all $t$, $S$ and $\theta$. 171
Proof. Define \( \overline{M}_t(\theta|S) := \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T] \), where only the outermost expectation is replaced by taking with respect to the probability law under \( \theta_0 \). Denote also \( \tilde{w}_j := v_{tj} - \mathbb{E}_{\theta,t}v_{tj} \). Then \( \overline{M}_t(\theta|S) = \sum_p p_{\theta_0,t}(j) \tilde{w}_j \tilde{w}_j^T \) and \( \overline{M}_t(\theta|S) - \overline{M}_t(\theta|S) = \sum_j \delta_j \tilde{w}_j \tilde{w}_j^T \), where \( \delta_j = p_{\theta_0,t}(j) - p_{\theta,t}(j) \). By Eq. (4.138) and the fact that \( \|v_{tj}\|_2 \leq \nu \), \( \|\theta - \theta_0\|_2 \leq \tau \), we have

\[
\max_j |\delta_j| \leq \sqrt{4\nu^2 \cdot \tau}. \tag{4.163}
\]

On the other hand, by (A2) we know that \( \min_j p_{\theta_0,t}(j) \geq 1/\rho K \) and therefore

\[
\overline{M}_t(\theta|S) = \sum_j p_{\theta_0,t} \tilde{w}_j \tilde{w}_j^T \geq \frac{1}{\rho K} \sum_j \tilde{w}_j \tilde{w}_j^T. \tag{4.164}
\]

Combining Eqs. (4.163, 4.164) and the fact that \( \overline{M}_t(\theta|S) - \overline{M}_t(\theta|S) = \sum_j \delta_j \tilde{w}_j \tilde{w}_j^T \), we have \( \overline{M}_t(\theta|S) - \overline{M}_t(\theta|S) \leq \overline{M}_t(\theta|S)/2 \) and \( \overline{M}_t(\theta|S) - \overline{M}_t(\theta|S) \leq \overline{M}_t(\theta|S)/2 \), provided that \( \tau \leq 1/\sqrt{8\rho^2 \nu^2 K^2} \). This also implies \( \frac{1}{2} \overline{M}_t(\theta|S) \leq \overline{M}_t(\theta|S) \leq 2M_t(\theta|S) \).

We next prove that \( \frac{1}{2} M_t(\theta_0|S) \leq \overline{M}_t(\theta|S) \leq 2M_t(\theta_0|S) \) which, together with \( \frac{1}{2} \overline{M}_t(\theta|S) \leq 2\overline{M}_t(\theta|S) \) established in the previous section, implies Lemma 85. Recall the definitions that

\[
M_t(\theta_0|S) = \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T];
\]

\[
\overline{M}_t(\theta|S) = \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T].
\]

Adding and subtracting \( \mathbb{E}_{\theta,t}v_{tj}, \mathbb{E}_{\theta,t}v_{tj} \) terms, we have

\[
\overline{M}_t(\theta|S) - M_t(\theta_0|S)
\]

\[
= \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj} + \mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj} + \mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T]
\]

\[
- \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T]
\]

\[
= \mathbb{E}_{\theta,t}[(\mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T] + \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(\mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T]
\]

\[
+ (\mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(\mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T.
\]

By Eq. (4.134) in the proof of Lemma 80, we have that

\[
(\mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(\mathbb{E}_{\theta,t}v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T \leq \frac{1}{2} \mathbb{E}_{\theta,t}[(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})(v_{tj} - \mathbb{E}_{\theta,t}v_{tj})^T] = \frac{1}{2} M_t(\theta_0|S)
\]

provided that \( \tau \leq 1/\sqrt{8\rho^2 \nu^2 K^2} \), thus implying \( \frac{1}{2} \overline{M}_t(\theta|S) \leq M_t(\theta_0|S) \). \( \Box \)

As a consequence of Lemma 85, the right-hand side of Eq. (4.162) can be upper bounded by

\[
\sqrt{d\log(\rho \nu KT)} \cdot \sqrt{4\|I_{t-1}(\theta_0)^{-1/2}M_t(\theta_0|S)I_{t-1}(\theta_0)^{-1/2}\|_{\text{op}}}. \]

Lemma 84 is thus proved. We are now ready to prove Lemma 83. By Lemma 84, we know that with high probability

\[
|\tilde{R}_t(S) - R_t(S)| \leq \sqrt{d\log(\rho \nu KT)} \cdot \sqrt{\|I_{t-1}(\theta_0)^{-1/2}M_t(\theta_0|S)I_{t-1}(\theta_0)^{-1/2}\|_{\text{op}}} \tag{4.165}
\]

172
In addition, by Lemma 85 and the fact that \( \| \hat{\theta}_{t-1} - \theta_0 \|_2 \leq \tau \) thanks to the local MLE formulation, we have \( \frac{1}{2} M_t(\theta_0|S) \leq \hat{M}_t(\hat{\theta}_{t-1}|S) \leq 4 M_t(\theta_0|S) \) and subsequently \( \frac{1}{2} I_{t-1}(\theta_0) \leq I_{t-1}(\hat{\theta}_{t-1}) \leq 4 I_{t-1}(\theta_0) \) because \( I_{t-1}(\cdot) \) and \( I_{t-1}(\cdot) \) are summations of \( M_\ell(\cdot) \) and \( \hat{M}_\ell(\cdot) \) terms. Setting \( \omega \geq \sqrt{d \log(\rho \nu T K)} \) we proved that \( R_t(S) \geq R_t(S) \). The second property of Lemma 83 can be proved similarly, by invoking the spectral similarities between \( I_{t-1}(\cdot), M_\ell(\cdot) \) and \( I_{t-1}(\cdot), \hat{M}_\ell(\cdot) \).

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 \( R_t(S) \leq R_t(S) \) for all \( S \) (see Lemma 83), we have the following upper bound for each term in the regret (see Eq. (4.28)):

\[
R_t(S_t^*) - R_t(S_t) \leq (R_t(S_t^*) - R_t(S_t)) + (R_t(S_t) - R_t(S_t)),
\]

where the last inequality holds because \( R_t(S_t^*) - R_t(S_t) \leq 0 \) (note that \( S_t \) maximizes \( R_t(\cdot) \)).

Subsequently, invoking Lemma 83 and the Cauchy-Schwarz inequality, we have

\[
\sum_{t=T_0+1}^T R_t(S_t^*) - R_t(S_t) \leq \sqrt{d \log(\rho \nu T K) \cdot \sum_{t=T_0+1}^T \min\{1, \| I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0) \|_\text{op} \}}
\]
\[
\leq \sqrt{d T \log(\rho \nu T K) \cdot \sum_{t=T_0+1}^T \min\{1, \| I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0) \|_\text{op}^2 \}}.
\]

The following lemma is a key result that upper bounds \( \sum_{t=T_0+1}^T \min\{1, \| I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0) \|_\text{op}^2 \} \).

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

\[
\sum_{t=T_0+1}^T \min\{1, \| I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0) \|_\text{op}^2 \} \leq 4 \log \frac{\det I_T(\theta_0)}{\det I_{T_0}(\theta_0)} \leq d \log(\lambda_0^{-1} \rho \nu).
\]

**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 \( \sigma_1(A_t) \geq \cdots \geq \sigma_d(A_t) \geq 0 \). By simple algebra,

\[
\sum_{t=T_0+1}^T \min\{1, \| I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0) \|_\text{op}^2 \} = \sum_{t=T_0+1}^T \min\{1, \sigma_1(A_t)^2 \}
\]
\[
\leq \sum_{t=T_0+1}^T 2 \log(1 + \sigma_1(A_t)^2) \leq \sum_{t=T_0+1}^T 4 \log(1 + \sigma_1(A_t)).
\]

(4.168)
On the other hand, note that $I_t(\theta_0) = I_{t-1}(\theta_0) + M_t(\theta_0|S_t) = I_{t-1}(\theta_0) + \frac{1}{2} [I_{d \times d} + A_t] I_{t-1}(\theta_0)^{-1}$. Hence,

$$\log \det I_t(\theta_0) = \log \det I_{t-1}(\theta_0) + \sum_{j=1}^d \log(1 + \sigma_j(A_t)).$$  \hfill (4.169)

Comparing Eqs. (4.168) and (4.169), we have

$$\sum_{t=T_0+1}^T \min\{1, \|I_{t-1}^{-1/2}(\theta_0) M_t(\theta_0|S_t) I_{t-1}^{-1/2}(\theta_0)\|_2^2\} \leq 4 \log \frac{\det I_T(\theta_0)}{\det I_{T_0}(\theta_0)},$$  \hfill (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 ($t \leq T_0$), we have

$$I_{T_0}(\theta_0) = \sum_{t=1}^{T_0} p_{\theta_0,t}(j_t)(1 - p_{\theta_0,t}(j_t))^2 v_{t,j_t} v_{t,j_t}^T \geq \frac{1}{(1 + \rho)^3} \sum_{t=1}^{T_0} v_{t,j_t} v_{t,j_t}^T,$$  \hfill (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 $\lambda_{\min}(\sum_{t=1}^{T_0} v_{t,j_t} v_{t,j_t}^T) \geq 0.5T_0\lambda_0$, where $\lambda_0 > 0$ is a parameter specified in assumption (A1). Therefore,

$$\det I_{T_0}(\theta_0) \geq \left[ T_0\lambda_0/\rho^3 \right]^d.$$  \hfill (4.172)

On the other hand, because $\max_{t,j} \|v_{t,j}\|_2 \leq \nu$ we have $I_T(\theta_0) \leq T \cdot \nu^2$ and subsequently

$$\det I_T(\theta_0) \leq [\nu^2 T]^d.$$  \hfill (4.173)

Combining Eqs. (4.172) and (4.173) we proved the second inequality in Lemma 86. \hfill \Box

We are now ready to give the final upper bound on $\text{Regret}(\{S_t\}_{t=1}^T)$ 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 $\overline{R}_t(S) \leq R_t(S)$ 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

$$\text{Regret}(\{S_t\}_{t=1}^T) \leq T_0 + O(1) + \mathbb{E} \sum_{t=T_0+1}^T R_t(S^*_t) - R_t(S_t) \leq O(1) + \frac{\nu^2 d \log T}{\lambda_0^2} + \frac{\rho^2(d + \log T)}{\tau^2 \lambda_0} + d\sqrt{T} \cdot \log(\lambda_0^{-1} \nu T K) \leq O(1) + d^2 \lambda_0^{-2} \rho^4 \nu^2 K^2 \log T + d\sqrt{T} \cdot \log(\lambda_0^{-1} \nu T K).$$  \hfill (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

$$\theta \in \Theta := \{\theta_W : W \in \mathcal{W}_{d/4} := \{\theta_W : W \subseteq [d], |W| = d/4\}. \quad (4.175)$$

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 $\{v_i\}$ 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)

$$[v_U]_i = 1/\sqrt{d} \quad \text{for } i \in U; \quad [v_U]_i = 0 \quad \text{for } i \notin U. \quad (4.176)$$

It is easy to check that with the condition $\epsilon \in (0, 1/\sqrt{d}]$, $\|\theta_W\|_2 \leq 1$ and $\|v_U\|_2 \leq 1$ for all $W, U \in \mathcal{W}_{d/4}$. Hence the worst-case regret of any policy $\pi$ can be lower bounded by the worst-case regret of parameters belonging to $\Theta$, which can be further lower bounded by the “average” regret over a uniform prior over $\Theta$:

$$\sup_{v,\theta} \mathbb{E}_{v,\theta} \sum_{t=1}^T R(S^*_\theta) - R(S_t) \geq \max_{\theta_W \in \Theta} \mathbb{E}_{v,\theta_W} \sum_{t=1}^T R(S^*_\theta) - R(S_t)$$

$$\geq \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_{v,\theta_W} \sum_{t=1}^T R(S^*_\theta) - R(S_t). \quad (4.177)$$

Here $S^*_\theta$ is the optimal assortment of size at most $K$ that maximizes (expected) revenue under parameterization $\theta$. By construction, it is easy to verify that $S^*_\theta$ 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 derive an explicit lower bound on the Bayes risk in Eq. (4.177). For any sequences $\{S_t\}_{t=1}^T$ produced by the policy $\pi$, we first describe an alternative sequence $\{\tilde{S}_t\}_{t=1}^T$ that provably enjoys less regret under parameterization $\theta_W$, while simplifying our analysis.
Lemma 87. Suppose one may choose an arbitrary feature \( v_U \) with \( U, \cdots, U_L \in \mathcal{W}_{d/4} \). Let \( U^* \) be the subset among \( U_1, \cdots, U_L \) that maximizes \( \langle v_U^*, \theta_W \rangle \), where \( \theta_W \) is the underlying parameter. Let \( \tilde{S}_t \) be the assortment consisting of all \( K \) items corresponding to feature \( v_U^* \). We then have the following observation:

**Proposition 20.** \( R(S_t) \leq R(\tilde{S}_t) \) under \( \theta_W \).

**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 \( u_j = \exp\{v_j^T \theta_W\} \) under \( \theta_W \). Clearly \( R(S) \) is a monotonically non-decreasing function in \( u_j \). By replacing all \( v_j \in S_t \) with \( v_U^* \in \tilde{S}_t \), the \( u_j \) values do not decrease and therefore the Proposition holds true.

To simplify notation we also use \( \tilde{U}_t \) to denote the unique \( U^* \in \mathcal{W}_{d/4} \) in \( \tilde{S}_t \). We also use \( \mathbb{E}_W \) and \( \mathbb{P}_W \) to denote the law parameterized by \( \theta_W \) and policy \( \pi \). The following lemma gives a lower bound on \( R(\tilde{S}_t) - R(S^*_{\theta_W}) \) by comparing it with \( W \).

**Lemma 87.** Suppose \( \epsilon \in (0, 1/d\sqrt{d}) \) and define \( \delta := d/4 - |\tilde{U}_t \cap W| \). Then

\[
R(S^*_{\theta_W}) - R(\tilde{S}_t) \geq \frac{-\delta \epsilon}{4K\sqrt{d}}.
\]

**Proof.** Let \( v = v_W \) and \( \tilde{v} = v_{\tilde{U}_t} \) be the corresponding feature vectors. Then

\[
R(S^*_{\theta_W}) - R(\tilde{S}_t) = \frac{K \exp\{v^T \theta_W\}}{1 + K \exp\{v^T \theta_W\}} - \frac{K \exp\{\tilde{v}^T \theta_W\}}{1 + K \exp\{\tilde{v}^T \theta_W\}}
\]

\[
= \frac{K[\exp\{v^T \theta_W\} - \exp\{\tilde{v}^T \theta_W\}]}{(1 + K \exp\{v^T \theta_W\})(1 + K \exp\{\tilde{v}^T \theta_W\})}
\]

\[
\geq \frac{\exp\{v^T \theta_W\} - \exp\{\tilde{v}^T \theta_W\}}{2Ke}.
\]

Here the last inequality holds because \( \max(\exp\{v^T \theta_W\}, \exp\{\tilde{v}^T \theta_W\}) \leq \epsilon \). In addition, by Taylor expansion we know that \( 1 + x \leq e^x \leq 1 + x + x^2/2 \) for all \( x \in [0, 1] \). Subsequently,

\[
R(S^*_{\theta_W}) - R(\tilde{S}_t) \geq \frac{(v - \tilde{v})^T \theta_W - (\tilde{v}^T \theta_W)^2/2}{2Ke} \geq \frac{\delta \epsilon / \sqrt{d} - (\sqrt{d} \epsilon)^2/2}{2Ke}.
\]

Finally, noting that \( d(\epsilon / \sqrt{d})^2/2 \leq \delta \epsilon / 2\sqrt{d} \) provided that \( \epsilon \in (0, 1/d\sqrt{d}) \), we finish the proof of Lemma 87.

Define random variables \( \bar{N}_i := \sum_{t=1}^{T} 1\{i \in \tilde{U}_t\} \). Lemma 87 immediately implies

\[
\mathbb{E}_W \sum_{t=1}^{T} R(S^*_{\theta_W}) - R(\tilde{S}_t) \geq \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\bar{N}_i] \right), \quad \forall W \in \mathcal{W}_{d/4}.
\]

(4.178)
Averaging both sides of Eq. (4.178) with respect to all $P$ in $\mathcal{W}_{d/4}$ and swapping the summation order, we have

$$
\frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \frac{1}{t-1} \sum_{i=1}^{T} R(S_{\theta_W}^i) - R(S_t) \geq \frac{\epsilon}{4K\sqrt{d}} \sum_{W \in \mathcal{W}_{d/4}} \left( \frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\tilde{N}_t] \right)
$$

$$
= \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{4} - \frac{1}{|\mathcal{W}_{d/4}|} \sum_{i=1}^{d} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W[\tilde{N}_t] \right)
$$

$$
\geq \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{4} - \frac{|\mathcal{W}_{d/4-1}|}{|\mathcal{W}_{d/4}|} \max_{W \neq \mathcal{W}_{d/4-1}} \sum_{i \in W} \mathbb{E}_W[\tilde{N}_t] \right)
$$

$$
= \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{4} - \frac{|\mathcal{W}_{d/4-1}|}{|\mathcal{W}_{d/4}|} \max_{W \neq \mathcal{W}_{d/4-1}} \sum_{i \in W} \mathbb{E}_W[\tilde{N}_t] + \mathbb{E}_W[\tilde{N}_t] - \mathbb{E}_W[\tilde{N}_t] \right).
$$

Note that for any fixed $W$, $\sum_{i \notin W} \mathbb{E}_W[\tilde{N}_t] \leq \sum_{i=1}^{d} \mathbb{E}_W[\tilde{N}_t] \leq dT/4$. Also, $|\mathcal{W}_{d/4-1}|/|\mathcal{W}_{d/4}| = (d/d_{4-1})/(d/4) = d/4d/4 \leq 1/3$. Subsequently,

$$
\frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S_{\theta_W}^i) - R(S_t) \geq \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{6} - \max_{W \neq \mathcal{W}_{d/4-1}} \sum_{i \in W} \mathbb{E}_W[\tilde{N}_t] - \mathbb{E}_W[\tilde{N}_t] \right).
$$

Pinckers’s inequality

In this section we concentrate on upper bounding $\|\mathbb{E}_{W \cup \{i\}}[\tilde{N}_t] - \mathbb{E}_W[\tilde{N}_t]\|$ 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 $\theta_W$ and $\theta_{W \cup \{i\}}$, respectively. Then

$$
\left| \mathbb{E}_P[\tilde{N}_t] - \mathbb{E}_Q[\tilde{N}_t] \right| \leq \sum_{j=0}^{T} \left| P[\tilde{N}_t = j] - Q[\tilde{N}_t = j] \right|
$$

$$
\leq T \cdot \sum_{j=0}^{T} \left| P[\tilde{N}_t = j] - Q[\tilde{N}_t = j] \right|
$$

$$
\leq T \cdot \left\| P - Q \right\|_{TV} \leq T \cdot \sqrt{\frac{1}{2} \text{KL}(P||Q)},
$$

where $\left\| P - Q \right\|_{TV} = \sup_A |P(A) - Q(A)|$ is the total variation distance between $P$, $Q$, $\text{KL}(P||Q) = \int (\log dP/dQ)dP$ is the Kullback-Leibler (KL) divergence between $P$, $Q$, and the inequality $\left\| P - Q \right\|_{TV} \leq \sqrt{\frac{1}{2} \text{KL}(P||Q)}$ is the celebrated Pinckers’s inequality.
For every \( i \in [d] \) define random variables \( N_i := \sum_{t=1}^{T} \frac{1}{K} \sum_{u: i \in S_t} 1 \{ i \in U \} \). The next lemma upper bound the KL divergence:

**Lemma 8.8.** For any \( W \in \mathcal{W} \), and \( i \in [d] \), \( \text{KL}(P_W \| P_{W \cup \{i\}}) \leq C_{KL} \cdot |E_W| \cdot e^2 / d \) for some universal constant \( C_{KL} > 0 \).

**Proof.** Fix a time \( t \) with policy’s assortment choice \( S_t \), and define \( n_i(S_t) := \sum_{u: i \in S_t} 1 \{ i \in U \} / K \). Let \( \{ p_j \} \) be the probabilities of purchasing item \( j \) under parameterization \( \theta_W \) and \( \theta_{W \cup \{i\}} \), respectively. Then

\[
\text{KL}(P_W (\cdot | S_t) \| P_{W \cup \{i\}} (\cdot | S_t)) = \sum_{j \in S_t \cup \{0\}} p_j \log \frac{q_j}{p_j} \leq \sum_j p_j \frac{p_j - q_j}{q_j} \leq \sum_j |p_j - q_j|^2,
\]

(4.180)

where the only inequality holds because \( \log(1 + x) \leq x \) for all \( x > -1 \). Because \( q_j \geq e^{-1} / (1 + Ke) \geq 1 / (2Ke^2) \) for all \( j \in S_t \cup \{0\} \), Eq. (4.180) is reduced to

\[
\text{KL}(P_W (\cdot | S_t) \| P_{W \cup \{i\}} (\cdot | S_t)) \leq 2e^2 K \cdot \sum_{j \in S_t \cup \{0\}} |p_j - q_j|^2.
\]

(4.181)

We next upper bound \( |p_j - q_j| \) separately. First consider \( j = 0 \). We have

\[
|p_j - q_j| = 1 + \frac{1}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_W \}} - 1 + \frac{1}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_{W \cup \{i\}} \}} \leq \frac{1}{(1 + K/e)^2} \cdot 2 \sum_{j \in S_t} |v_j^\top (\theta_W - \theta_{W \cup \{i\}})| \leq \frac{2Kn_i(S_t)e/\sqrt{d}}{(1 + K/e)^2} \leq \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}}.
\]

Here the first inequality holds because \( e^x \leq 1 + 2x \) for all \( x \in [0, 1] \).

For \( j > 0 \) corresponding to \( v_j = v_U \) where \( i \notin U \), we have

\[
|p_j - q_j| = \frac{\exp \{ v_U^\top \theta_W \}}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_W \}} - \frac{\exp \{ v_U^\top \theta_{W \cup \{i\}} \}}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_{W \cup \{i\}} \}} \leq \frac{1}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_W \}} - \frac{1}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_{W \cup \{i\}} \}} \leq \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}}.
\]

Here the first inequality holds because \( \exp \{ v_U^\top \theta_W \} = \exp \{ v_U^\top \theta_{W \cup \{i\}} \} \leq 1 \), since \( i \notin U \).

For \( j > 0 \) corresponding to \( v_j = v_U \) and \( i \in U \), we have

\[
|p_j - q_j| = \frac{\exp \{ v_U^\top \theta_W \}}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_W \}} - \frac{\exp \{ v_U^\top \theta_{W \cup \{i\}} \}}{1 + \sum_{j \in S_t} \exp \{ v_j^\top \theta_{W \cup \{i\}} \}}
\]

178
\[
\exp\{v_u^\top \theta_{W \cup \{i\}}\} \cdot \frac{1}{1 + \sum_{j \in S_i} \exp\{v_j^\top \theta_W\}} - \frac{1}{1 + \sum_{j \in S_i} \exp\{v_j^\top \theta_{W \cup \{i\}}\}}
\]

\[
+ |\exp\{v_u^\top \theta_W\} - \exp\{v_u^\top \theta_{W \cup \{i\}}\}| \cdot \frac{1}{1 + \sum_{j \in S_i} \exp\{v_j^\top \theta_W\}} \leq \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}} + \frac{\epsilon}{\sqrt{d}} \cdot \frac{1}{1 + K/\epsilon} \leq \frac{8e^2 n_i(S_t)\epsilon}{K\sqrt{d}} + \frac{2e\epsilon}{K\sqrt{d}}.
\]

Combining all upper bounds on $|p_j - q_j|$ and Eq. (4.181), we have

\[
\KL(P_W(\cdot | S_i) \| P_{W \cup \{i\}}(\cdot | S_i)) \leq 2e^2 K \cdot \left[ \frac{128e^4 n_i(S_t)^2\epsilon^2}{K^2 d} (1 + K) + K n_i(S_t) \cdot \frac{8e^4\epsilon^2}{K^2 d} \right]
\]

\[
\leq n_i(S_t)\epsilon^2/d.
\]

Here the last inequality holds because $n_i(S_t) \leq 1$. 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

\[
\KL(P_W \| P_{W \cup \{i\}}) \leq \mathbb{E}_W[N_i] \cdot \epsilon^2/d,
\]

which is to be demonstrated. \[\square\]

Combining Lemma 88 and Eq. (4.179), we have

\[
\frac{1}{|W_{d/4}|} \sum_{W \in W_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S^n_{\theta_W}) - R(S_t) \geq \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{6} - T \sum_{i=1}^{d} \sqrt{\KL \mathbb{E}_W[N_i]\epsilon^2/d} \right).
\]

Further using Cauchy-Schwartz inequality, we have

\[
\sum_{i=1}^{d} \sqrt{\KL \mathbb{E}_W[N_i]\epsilon^2/d} \leq \sqrt{d} \cdot \sqrt{\sum_{i=1}^{d} \KL \mathbb{E}_W[N_i]\epsilon^2/d},
\]

which is further upper bounded by $\sqrt{d} \cdot \sqrt{\KL T\epsilon^2/4}$ because $\sum_{i=1}^{d} \mathbb{E}_W[N_i] \leq dT/4$. Subsequently,

\[
\frac{1}{|W_{d/4}|} \sum_{W \in W_{d/4}} \mathbb{E}_W \sum_{t=1}^{T} R(S^n_{\theta_W}) - R(S_t) \geq \frac{\epsilon}{4K\sqrt{d}} \left( \frac{dT}{6} - T \sqrt{\KL \frac{dT}{4}\epsilon^2} \right),
\]

(4.182)

where $\KL = \KL / 4$. Setting $\epsilon = \sqrt{d/144\KL T}$ we complete the proof of Theorem 20.
Chapter 5

Conclusion and discussion

In this thesis, we study various problems under the general theme of selective data acquisition 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 rise 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 ($K \leq N/4$) 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; Elmahraby & 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; Elmaghraby & 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 $a \leq X_i \leq b$ almost surely. Then for any $t > 0$,

$$\Pr \left[ \frac{1}{n} \sum_{i=1}^{n} X_i - \mathbb{E}X > t \right] \leq 2 \exp \left\{ -\frac{nt^2}{2(b-a)^2} \right\}.$$

**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 $t > 0$,

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

**Lemma 91** (Bernstein's inequality). Suppose $X$ is a sub-exponential random variable with parameters $\nu$ and $\alpha$.

$$\Pr \left[ |X - \mathbb{E}X| > t \right] \leq \begin{cases} 2 \exp \left\{ -t^2/2\nu^2 \right\}, & 0 < t \leq \nu^2/\alpha; \\ 2 \exp \left\{ -t/2\alpha \right\}, & t > \nu^2/\alpha. \end{cases}$$

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 $\mathbb{E}[X_j^2|X_1, \ldots, X_{j-1}] \leq \sigma^2$ for all $t = 1, \ldots, n$. Further assume that $\mathbb{E}[|X_j^k|X_1, \ldots, X_{j-1}] \leq \frac{1}{2} k! \sigma^2 b^{k-2}$ for all integers $k \geq 3$. Then for all $t > 0$,

$$\Pr \left[ \left| \sum_{j=1}^{n} X_j \right| \geq t \right] \leq 2 \exp \left\{ -\frac{t^2}{2(n\sigma^2 + bt)} \right\}.$$
The following lemma is the Hoeffding’s maximal inequality, by Hoeffding (1963).

**Lemma 93** (Hoeffding’s maximal inequality). Let \( X_1, \ldots, X_n \) be i.i.d. random variables with mean \( \mu \) and satisfy \( a \leq X_i \leq b \) almost surely for all \( i \in [n] \). Then for any \( t > 0 \),

\[
\Pr \left[ \forall i \in [n], X_1 + \cdots + X_i \geq i \cdot \mu + t \right] \leq \exp \left\{ - \frac{2t^2}{n(b-a)^2} \right\}. \tag{A.1}
\]

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 \( X_1, \ldots, X_n \) are i.i.d. geometric random variables with parameters \( p > 0 \), meaning that \( \Pr[X_i = k] = (1 - p)^k p \) for \( k = 0, 1, 2, \ldots \). Define \( \mu := \mathbb{E} X_i = (1 - p)/p \). Then

\[
\Pr \left[ \frac{1}{n} \sum_{i=1}^n X_i > (1 + \delta)\mu \right] \leq \left\{ \begin{array}{ll}
\exp \left\{ - \frac{n \mu^2}{2(1 + \delta)(1 + \mu)^2} \right\}, & \text{if } \mu \leq 1, \\
\exp \left\{ - \frac{n \delta^2 \mu^2}{6(1 + \mu)^2} \left( 3 - \frac{2 \delta \mu}{1 + \mu} \right) \right\}, & \text{if } \mu > 1, \delta \in (0, 1);
\end{array} \right.
\]

\[
\Pr \left[ \frac{1}{n} \sum_{i=1}^n X_i < (1 - \delta)\mu \right] \leq \left\{ \begin{array}{ll}
\exp \left\{ - \frac{n \delta^2 \mu^2}{6(1 + \mu)^2} \left( 3 - \frac{2 \delta \mu}{1 + \mu} \right) \right\}, & \text{if } \mu \leq 1, \\
\exp \left\{ - \frac{n \delta^2 \mu^2}{2(1 + \mu)^2} \right\}, & \text{if } \mu > 1.
\end{array} \right.
\]

### A.2 Matrix/vector concentration inequalities

**Lemma 95** (Rudelson & Vershynin (2007)). Let \( x \) be a \( p \)-dimensional random vector such that \( \|x\|_2 \leq M \) almost surely and \( \|\mathbb{E} xx^\top\|_2 \leq 1 \). Let \( x_1, \ldots, x_n \) be i.i.d. copies of \( x \). Then for every \( t \in (0, 1) \)

\[
\Pr \left[ \frac{1}{n} \sum_{i=1}^n x_i x_i^\top - \mathbb{E} xx^\top \right]_2 > t \right] \leq 2 \exp \left\{ - C \cdot \frac{nt^2}{M^2 \log n} \right\},
\]

where \( C > 0 \) is some universal constant.

**Lemma 96** (Corollary 5.2 of Mackey et al. (2014)). Let \( (Y_k)_{k \geq 1} \) be a sequence of random \( d \)-dimensional Hermitian matrices that satisfy

\[
\mathbb{E} Y_k = 0 \quad \text{and} \quad \|Y_k\|_2 \leq R \quad \text{a.s.}
\]

Define \( X = \sum_{k \geq 1} Y_k \). The for any \( t > 0 \),

\[
\Pr \left[ \|X\|_2 \geq t \right] \leq d \cdot \exp \left\{ - \frac{t^2}{3\sigma^2 + 2Rt} \right\} \quad \text{for} \quad \sigma^2 = \left\| \sum_{k \geq 1} \mathbb{E} Y_k^2 \right\|_2.
\]

**Lemma 97** (Corollary 10.3 of Mackey et al. (2014)). Let \( A_1, \ldots, A_n \) be a sequence of deterministic \( d \)-dimensional Hermitian matrices that satisfy

\[
\sum_{k=1}^n A_k = 0 \quad \text{and} \quad \sup_{1 \leq k \leq n} \|A_k\|_2 \leq R.
\]
Define random matrix \( X = \sum_{j=1}^{m} A_{\sigma(j)} \) for \( m \leq n \), where \( \sigma \) is a random permutation from \([n]\) to \([n]\). Then for all \( t > 0 \),

\[
\Pr[\|X\|_2 \geq t] \leq d \exp\left\{-\frac{t^2}{12\sigma^2 + 4\sqrt{2}Rt}\right\} \quad \text{for} \quad \sigma^2 = \frac{m}{n} \left( \sum_{k=1}^{n} A_k^2 \right)_2.
\]

**Lemma 98** (Tropp (2015), simplified). Suppose \( A_1, \ldots, A_n \) are i.i.d. positive semidefinite random matrices of dimension \( d \) and \( \|A_i\|_{op} \leq R \) almost surely. Then for any \( t > 0 \),

\[
\Pr \left[ \left\| \frac{1}{n} \sum_{i=1}^{n} A_i - \mathbb{E}A \right\|_{op} > t \right] \leq 2 \exp\left\{-\frac{nt^2}{8R^2}\right\}.
\]

**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 \( \max_{1 \leq i \leq d} |\sigma_i - \sigma'_i| \leq \|E\|_{op} \).
Bibliography


190


191


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


Fedorov, V. V. (1972). *Theory of optimal experiments*. Elsevier. 1, 19, 181


Pukelsheim, F. (2006). *Optimal design of experiments*. SIAM. 1, 5, 6, 19, 181


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


200


201


