
Online Assortment and Price Optimization Under Contextual Choice Models

Yigit Efe Erginbas
UC Berkeley

Thomas A. Courtade
UC Berkeley

Kannan Ramchandran
UC Berkeley

Abstract

We consider an assortment selection and pricing problem in which a seller has N different items available for sale. In each round, the seller observes a d -dimensional contextual preference information vector for the user, and offers to the user an assortment of K items at prices chosen by the seller. The user selects at most one of the products from the offered assortment according to a multinomial logit choice model whose parameters are unknown. The seller observes which, if any, item is chosen at the end of each round, with the goal of maximizing cumulative revenue over a selling horizon of length T . For this problem, we propose an algorithm that learns from user feedback and achieves a revenue regret of order $\tilde{\mathcal{O}}(d\sqrt{KT}/L_0)$ where L_0 is the minimum price sensitivity parameter. We also obtain a lower bound of order $\Omega(d\sqrt{T}/L_0)$ for the regret achievable by any algorithm.

1 INTRODUCTION

In online marketplaces, dynamic assortment selection and pricing for sequentially arriving buyers presents a challenge for online learning. Since the preferences of buyers are varying and uncertain, adaptive strategies are essential to meet their needs and maximize the effectiveness of offers. To address this problem, we investigate the application of online learning techniques for contextual assortment selection and pricing. Assortment selection involves the seller choosing a subset of items from a vast catalog to present to buyers, and dynamically assigning prices to the offered items. The overall goal is to maximize revenue over the course of repeated interactions.

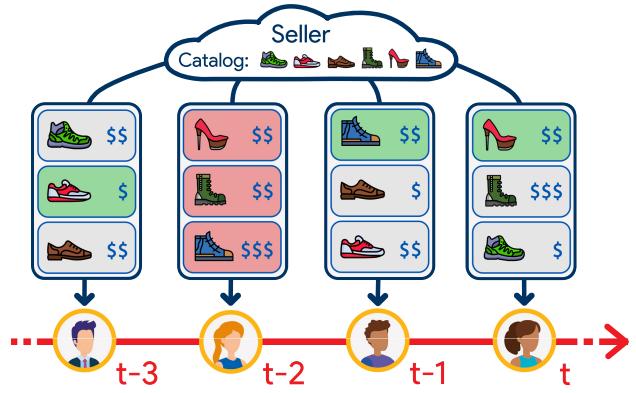


Figure 1: A seller has access to a catalog (set) of $N = 6$ distinct items, from which it can advertise to sequentially arriving users. In each round, the seller offers an assortment of $K = 3$ items at well-chosen prices. The user selects one of the products from the offered assortment (represented with a green background), or rejects all offered items (represented with a red background).

Dynamic assortment selection and pricing strategies are deployed in a variety of online sectors including e-commerce (e.g., Amazon), food delivery (e.g., Uber Eats), and hospitality (e.g., Airbnb). With similar systems becoming ubiquitous in our daily lives, there is a growing opportunity to deliver tailored product recommendations and pricing adjustments. Therefore, it is crucial to consider data-driven approaches that can enhance user experiences and boost profitability in today's highly competitive digital industry.

We design sequential *assortment selection and pricing* algorithms that offer a sequence of assortments (menus) of up to K items from a catalog of N possible items. The learning agent (seller) sequentially selects assortments to offer and sets prices for the included items. After making assortment and pricing decisions in each round, the learning agent receives user feedback, which consists of the specific item chosen from the offered assortment. We assume that the item choice follows a multinomial logistic (MNL) model (McFadden, 1978), which is one of the most widely used models in dynamic assortment optimization literature (Caro and Gallien,

2007; Agrawal et al., 2017; Aouad et al., 2018). Because assortment-based offers are relevant to many industries that involve access to additional information about users, contextual choice models have gained significant traction in recent years (Chen et al., 2020; Javanmard et al., 2020). In alignment with this approach, we assume that the utility parameters in the MNL choice model are linear functions of d -dimensional context vectors that are revealed at each round.

To address a range of real-world scenarios where price optimization is essential for maximal revenue, we incorporate the *pricing* of items as a second component of the seller’s problem. This largely differs from previous literature on sequential assortment selection, wherein prices are assumed to be predetermined (Chen et al., 2020; Oh and Iyengar, 2021). The main challenge in our work is **the complex interdependence between assortment and pricing decisions**, an issue that existing methods are not designed to address.

In the process of offering a sequence of assortments with judiciously chosen prices, the seller’s goal is to maximize the expected revenue accumulated over a time horizon of T rounds. However, since the seller does not have knowledge of the parameters of the contextual choice model ahead of time, the decisions involve a trade-off between learning the choice model in order to increase long-term revenues and earning short-term revenues by leveraging the already-acquired information.

1.1 Overview of Our Algorithm

Our algorithm selects *optimistic* assortment and prices that balance the trade-off between exploration and exploitation. This is accomplished by deriving tight upper bounds for the utility functions in the MNL model. In contrast to the dynamic assortment selection literature, which only establishes a pointwise upper bound for the value of an assortment, we construct price-dependent functions that upper bound the values across all price points. This construction allows us to quantify the varying uncertainty for different prices and successfully assess the trade-off involved in joint optimization of assortment and prices.

To construct these utility upper bounds, we need to obtain estimates of the parameters in the MNL model. However, dynamic estimation of MNL parameters has been exclusively studied under the assumption of fixed prices and the state-of-the-art techniques result in a dependence on a problem-dependent parameter κ (Oh and Iyengar, 2021)¹. If we consider extending these analyses to include price selection, we observe that the κ parameter strongly depends on the assortment size K and the minimum price sensitivity L_0 ². In particular, the worst-case dependency is $\kappa = K^{2+1/L_0}$

which would translate into a $\mathcal{O}(K^{2+1/L_0} d \sqrt{T})$ regret bound. Hence, a direct extension of existing approaches is far from optimal, especially when the minimum price sensitivity parameter L_0 is small.

We tackle this issue by constructing better estimates of the Fisher Information Matrix for the parameters of the MNL model, which enables us to eliminate the κ dependence. The key to our analysis is a novel Bernstein-type inequality for self-normalized vector-valued martingales which we derive based on techniques introduced in Faury et al. (2020).

Consistent with the sequential decision-making literature, we measure the performance of algorithms using a relevant notion of regret, defined as the difference between the expected revenue generated by the algorithm and the offline optimal expected revenue when all parameters are known. We show that our algorithm achieves a revenue regret of order $\tilde{\mathcal{O}}(d\sqrt{KT}/L_0)$, which, as we show, is the best possible up to logarithmic factors in d , T , and minimum price sensitivity L_0 ².

1.2 Our Contributions

To the best of our knowledge, we are the first to address the problem of dynamic contextual assortment selection and pricing **simultaneously**. Our contributions are:

- *Formulation:* We introduce and formalize the problem of sequential assortment and price optimization under contextual multinomial logit choice model.
- *Regret upper bound:* We develop an algorithm for the contextual assortment selection and pricing problem (Algorithm 2). We show that it achieves $\tilde{\mathcal{O}}(d\sqrt{KT}/L_0)$ regret in T rounds where d is the dimension of the context vectors, K is the assortment size, and L_0 is the minimum price sensitivity. We further improve the time and space complexity of our algorithm by leveraging online Newton step (ONS) techniques for parameter estimation in Algorithm 3.
- *Regret lower bound:* We show that for any algorithm, there exists an adversarial problem instance such that it incurs $\Omega(d\sqrt{T}/L_0)$ regret. Therefore, Algorithm 2 enjoys optimal regret up to logarithmic terms in d , T , N , and L_0 .
- *Assortment and price optimization algorithm:* As a part of our solution, we develop an efficient algorithm (Algorithm 1) to find the optimal assortment and prices under the MNL model with any differentiable and strictly decreasing utility function.

¹The parameter κ inversely scales with the minimum probability of each item being chosen. For a precise definition of κ and additional details, please refer to Appendix E.

²The minimum price sensitivity L_0 is a lower bound for the rate of decay of the utility as a function of the prices.

Table 1: Comparison of related works, provided regret bounds, and computational complexity per time step of given algorithms. T is the number of rounds, K is the assortment size, N is the total number of items, d is the feature dimension. The big- \mathcal{O} and big- Ω notations denote the regret upper and lower bounds, respectively. To the best of our knowledge, we are the first to address the problem of simultaneous contextual assortment selection and pricing.

	Context	Assortment	Pricing	Regret	Computational Complexity ³
Agrawal et al. (2019)	No	Yes	No	$\tilde{\mathcal{O}}(\sqrt{NT}), \Omega(\sqrt{NT/K})$	$\Theta(N)$
Miao and Chao (2018)	No	Yes	Yes	$\tilde{\mathcal{O}}(\sqrt{NT})$ ⁴	unknown ⁵
Chen et al. (2020)	Yes	Yes	No	$\tilde{\mathcal{O}}(d\sqrt{T}), \Omega(d\sqrt{T}/K)$	$\Theta(KT + \binom{N}{K})$
Oh and Iyengar (2021)	Yes	Yes	No	$\tilde{\mathcal{O}}(\kappa d\sqrt{T})$	$\Theta(N)$
Javanmard et al. (2020)	Yes	No	Yes	$\mathcal{O}(\log(dT)\sqrt{T})$	$\Theta(N\sqrt{T})$
Perivier and Goyal (2022)	Yes	No	Yes	$\tilde{\mathcal{O}}(d\sqrt{T})$ ⁶	$\Theta(N)$
Perivier and Goyal (2022)	Yes	Yes	No	$\tilde{\mathcal{O}}(dK\sqrt{T})$	unknown ⁷
CAP (Algorithm 2)	Yes	Yes	Yes	$\tilde{\mathcal{O}}(d\sqrt{KT}/L_0), \Omega(d\sqrt{T}/L_0)$	$\Theta(KT + N)$
CAP-ONS (Algorithm 3)	Yes	Yes	Yes	$\tilde{\mathcal{O}}(dK\sqrt{T}/L_0)$	$\Theta(N)$

Remark 1.1. The gap between our upper and lower bounds for regret is on the order of $\mathcal{O}(\sqrt{K})$, but given that the maximum assortment size is typically small (e.g., 5 to 20) in most real-world scenarios, this difference might be considered non-critical.

1.3 Related Works

Generalized Linear Bandits Linear bandits, generalized linear bandits, and their variants have been extensively studied in the context of sequential decision-making with contextual information (Abbasi-Yadkori, 2011; Chu et al., 2011; Li et al., 2017). Building on this literature, recent works by Ban and Keskin (2021); Xu and Wang (2024); Wang et al. (2025) have explored parametric contextual pricing for a single item under generalized linear demand models, where demand depends solely on the item’s own price. In contrast, the MNL model we consider captures demand through a choice model, accounting for the influence of all item prices in the assortment. Another line of research examines combinatorial variants of the contextual bandit problem, often incorporating semi-bandit or cascading feedback (Chen et al., 2013; Qin et al., 2014; Kveton et al., 2015; Zong et al., 2016). However, these approaches cannot account for substitution effects, as their choice models fail to consider which other items are included in the assortment.

Dynamic Assortment Selection There has been an emerging body of literature on multinomial logit (MNL) bandits in both non-contextual (Cheung and Simchi-Levi, 2017; Agrawal et al., 2019) and contextual settings (Oh and Iyengar, 2019; Chen et al., 2020; Agrawal et al., 2020; Oh and Iyengar, 2021). While

these studies address the sequential assortment selection problem under the MNL choice model, they assume fixed prices for the items.

Incorporating variable prices directly into these algorithms, such as in Chen et al. (2020) or Agrawal et al. (2020), proves impractical as they compute separate upper confidence bounds for the value of each of the $\binom{N}{K}$ possible assortments and choose the one with maximum value. With the addition of pricing into the problem, these upper bounds become functions of the prices for all items and make the optimization even harder. Oh and Iyengar (2021) offers a polynomial-time contextual MNL-bandit algorithm that computes upper confidence bounds for the value of each item rather than each assortment. However, their algorithm and analysis translates into a $\mathcal{O}(K^{2+1/L_0}d\sqrt{T})$ regret bound when we introduce price optimization (see Appendix E for details). Hence, this approach is also far from optimal. Recently, Perivier and Goyal (2022) has also provided an assortment selection algorithm with improved regret bounds. However, their analysis only works under the assumption of uniform prices across items, which does not hold in our setting.

Dynamic Pricing The problem of dynamic pricing has been typically modeled as a variant of the multi-armed bandit problem that aims to maximize revenue from selling copies of a single good to sequentially ar-

³We provide the amortized computational complexity per time step with respect to parameters N , K , and T .

⁴The given regret bounds are for Bayesian regret.

⁵It depends on the prior of the parameter.

⁶This result considers an adversarial arrival model.

⁷Their confidence sets are expensive to compute.

riving users (Kleinberg and Leighton, 2003; Besbes and Zeevi, 2009; Bubeck et al., 2019; Paes Leme and Schneider, 2018; Xu and Wang, 2021). Our contribution stands out by considering the combinatorial aspect of the assortment selection problem faced in simultaneously offering multiple items. Recent studies by Javanmard et al. (2020) and Perivier and Goyal (2022) consider the problem of pricing multiple items that are offered under the MNL choice model. However, in contrast to our work, these frameworks assume that all available items are offered to the buyer. To the best of our knowledge, there is only one work (Miao and Chao, 2018) which considers the joint problem of assortment optimization and pricing under unknown demand information. However, this work does not utilize a contextual model and assumes that the unknown parameter is randomly drawn from a prior distribution known to the algorithm.

2 PROBLEM DEFINITION

Notation: We use bold lowercase font for vectors \mathbf{x} and bold uppercase font for matrices \mathbf{X} . For a vector \mathbf{x} , we denote its i -th entry by x_i and we use $\|\mathbf{x}\|$ to denote its ℓ^2 -norm. For two vectors \mathbf{x} and \mathbf{y} , we use $(\mathbf{x}; \mathbf{y})$ to denote their concatenation and use $\langle \mathbf{x}, \mathbf{y} \rangle$ to denote their inner product. For a vector \mathbf{x} and a positive-definite matrix \mathbf{W} , we use $\|\mathbf{x}\|_{\mathbf{W}}$ to denote the weighted ℓ^2 -norm. For any positive integer N , we use $[N]$ to denote the set $\{1, 2, \dots, N\}$.

We consider the problem of online assortment selection and pricing for selling items to sequentially arriving buyers. We denote the set of available items by $[N]$ and consider that the seller is constrained to offer at most K items to each arriving buyer. Accordingly, we let $\mathcal{S}_K := \{S \subseteq [N] : |S| \leq K\}$ denote the set of all possible assortments that the seller can choose to offer.

At each time $t \in [T]$, the seller observes random feature vectors $\mathbf{x}_{ti} \in \mathbb{R}^d$ for each item $i \in [N]$. Given this contextual information, the seller offers an assortment of items $S_t \in \mathcal{S}_K$ and chooses a price $p_{ti} \in \mathbb{R}$ for each offered item $i \in S_t$. At the end of each round t , the seller observes only the purchase decision $i_t \in S_t \cup \{0\}$ of the buyer and obtains revenue p_{ti_t} . Here, $\{0\}$ represents the no-purchase option (or outside option), which indicates that the user did not choose any item offered in S_t , resulting in revenue $p_{t0} = 0$. For convenience, we let $\mathbf{p}_t \in \mathbb{R}^N$ denote the collection of prices chosen for all items.

For a given assortment S_t and price vector \mathbf{p}_t , the buyer's decision i_t is a categorical random variable with support $S_t \cup \{0\}$. We model this decision via the widely used multinomial logit (MNL) choice model (McFadden, 1978) under a linear contextual utility

function. Formally, the choice probability for each item $i \in S_t$ (and the no-purchase option) is assumed to be given as in the following assumption.

Assumption 2.1 (Multinomial logit choice under linear contextual utility). The utility of the buyer at time t for item i is given by the linear model

$$u_{ti}(p) = \langle \boldsymbol{\psi}^*, \mathbf{x}_{ti} \rangle - \langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle \cdot p$$

where $\boldsymbol{\psi}^* \in \mathbb{R}^d$ and $\boldsymbol{\phi}^* \in \mathbb{R}^d$ are time-invariant parameter vectors unknown to the seller. In this model, the $\alpha_{ti} := \langle \boldsymbol{\psi}^*, \mathbf{x}_{ti} \rangle$ term represents the buyer's base valuation of the item while the $\beta_{ti} := \langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle$ term represents the buyer's price sensitivity.

Then, given an assortment S_t with prices \mathbf{p}_t , the probability that the buyer selects item $i \in S_t$ is

$$q_t(i|S_t, \mathbf{p}_t) := \frac{\exp\{u_{ti}(p_{ti})\}}{1 + \sum_{j \in S_t} \exp\{u_{tj}(p_{tj})\}}.$$

Consequently, the probability of no purchase is

$$q_t(0|S_t, \mathbf{p}_t) := \frac{1}{1 + \sum_{j \in S_t} \exp\{u_{tj}(p_{tj})\}}.$$

Under this model, the expected revenue at time t is

$$R_t(S_t, \mathbf{p}_t) := \sum_{i \in S_t} p_{ti} \cdot q_t(i|S_t, \mathbf{p}_t) \quad (1)$$

for any selection of assortment $S_t \in \mathcal{S}_K$ and price vector $\mathbf{p}_t \in \mathbb{R}^N$. Thus, for a sequence of assortments $S_t \in \mathcal{S}_K$ and price vectors $\mathbf{p}_t \in \mathbb{R}^N$ chosen over time, the cumulative expected revenue can be written as $\sum_{t=1}^T R_t(S_t, \mathbf{p}_t)$.

After the seller decides on the assortment $S_t \in \mathcal{S}_K$ and prices $\mathbf{p}_t \in \mathbb{R}^N$ to offer to the user at each time t , the user reports the item $i_t \in S_t \cup \{0\}$ that they have decided to purchase. We denote by H_t the history $\{(\mathbf{x}_{\tau i})_{i \in [N]}, S_\tau, \mathbf{p}_\tau, i_\tau\}_{\tau=1}^{t-1}$ of observations available to the seller when choosing the next set of assortment $S_t \in \mathcal{S}_K$ along with the next price vector \mathbf{p}_t . Then, the seller agent employs a policy $\pi = \{\pi^t | t \in [T]\}$, which is a sequence of functions, each mapping the history H_t and the context vectors $\{\mathbf{x}_{ti}\}_{i \in [N]}$ to an action $(S_t, \mathbf{p}_t) \in \mathcal{S}_K \times \mathbb{R}^N$.

Given the contextual information at every round t , the task of the seller is to sequentially offer the items to users at well-chosen prices such that it can achieve maximal revenue. To evaluate policies in achieving this objective, we define the *regret* metric that measures the gap between the expected revenue of policy π and that of the offline optimal sequence of assortments and prices. The regret \mathcal{R}_T for a time horizon of T periods is defined as

$$\mathcal{R}_T := \sum_{t=1}^T R_t(S_t^*, \mathbf{p}_t^*) - \sum_{t=1}^T R_t(S_t, \mathbf{p}_t),$$

where (S_t^*, \mathbf{p}_t^*) denotes an offline optimal assortment and price selection that satisfies

$$(S_t^*, \mathbf{p}_t^*) \in \operatorname{argmax}_{\substack{S \in \mathcal{S}_K \\ \mathbf{p} \in \mathbb{R}^N}} R_t(S, \mathbf{p}). \quad (2)$$

Based on the definition of the regret metric, we see that regret minimization is equivalent to maximizing the cumulative expected revenue.

3 ASSORTMENT AND PRICE OPTIMIZATION

As stated in Assumption 2.1, we assume that buyers' purchase decisions are given by a multinomial logit (MNL) model. Therefore, the assortment and price optimization at time t can be formulated as

$$\max_{\substack{S_t \in \mathcal{S}_K \\ \mathbf{p}_t \in \mathbb{R}^N}} R_t(S_t, \mathbf{p}_t) = \max_{\substack{S_t \in \mathcal{S}_K \\ \mathbf{p}_t \in \mathbb{R}^N}} \frac{\sum_{i \in S_t} p_{ti} \exp\{u_{ti}(p_{ti})\}}{1 + \sum_{j \in S_t} \exp\{u_{tj}(p_{tj})\}}.$$

We recall that the utility functions are given by linear form $u_{ti}(p) = \langle \psi^*, \mathbf{x}_{ti} \rangle - \langle \phi^*, \mathbf{x}_{ti} \rangle \cdot p$ and make the following regularity assumption.

Assumption 3.1 (Minimum Price Sensitivity). There exists a constant $L_0 > 0$ such that utility functions satisfy $u'_{ti}(p) = -\langle \phi^*, \mathbf{x}_{ti} \rangle \leq -L_0$ for all $t \in [T]$ and $i \in [N]$, almost surely.

This assumption ensures that the utility function $u_{ti}(p)$ is strictly decreasing in price and hence infinity is a so-called null price, i.e. $\lim_{p \rightarrow \infty} p e^{u_{ti}(p)} = 0$, so that the objective function $R_t(S_t, \mathbf{p}_t)$ has a finite maximum.

Characterization of Optimality Even though the true utility functions are assumed to be linear, our learning algorithm will require us to solve for the optimum assortment and prices under broader classes of utility functions. Hence, in the next proposition, we characterize optimality under any differentiable and strictly decreasing utility function $h_{ti}(p)$.

Proposition 3.2 (Optimum assortment and prices). Suppose utility functions $h_{ti}(p)$ are differentiable and strictly decreasing for all items $i \in [N]$. Let B_t be the unique solution of the fixed point equation

$$B = \max_{S \in \mathcal{S}_K} \sum_{i \in S} v_{ti}(B) \quad (3)$$

where $v_{ti}(B) := \max_{p \in \mathbb{R}} \{f_{ti}(p) : p + 1/h'_{ti}(p) = B\}$ and $f_{ti}(p) := -e^{h_{ti}(p)}/h'_{ti}(p)$. Then, the optimum assortment S_t^* is the assortment S that achieves the maximum in the optimization problem (3), the optimum prices are

$$p_{ti}^* = \operatorname{argmax}_{p \in \mathbb{R}} \{f_{ti}(p) : p + 1/h'_{ti}(p) = B_t\},$$

and the optimum revenue achieved by (S_t^*, \mathbf{p}_t^*) is B_t .

Proof. (Sketch) First, we write the first-order necessary conditions for the optimality of prices as $\nabla_{\mathbf{p}} \{\sum_{i \in S} p_i q_t(i|S, \mathbf{p})\} = \mathbf{0}$ under any fixed assortment S . Using the structure of the MNL model, this necessary condition reduces to $\sum_{i \in S} p_i q_t(i|S, \mathbf{p}) = p_j + \frac{1}{h'_{tj}(p_j)}$ for all $j \in S$. Note that the left-hand side of the equation is equal to the revenue obtained at prices \mathbf{p} . Therefore, the pricing problem can be written as maximizing B subject to $B = \sum_{i \in S} p_i q_{ti}(\mathbf{p})$ and $B = p_i + \frac{1}{h'_i(p_i)}$ for all $i \in S$.

Furthermore, using the form of MNL, we can show that the condition $B = \sum_{i \in S} p_i q_{ti}(\mathbf{p})$ is equivalent to $B = \sum_{i \in S} f_{ti}(p_i)$ where $f_{ti}(p) = -e^{h_{ti}(p)}/h'_{ti}(p)$. Therefore, the pricing problem can be written as maximizing B subject to conditions (a) $B = \sum_{i \in S} f_{ti}(p_i)$ and (b) $B = p_i + \frac{1}{h'_i(p_i)}$ for all i .

To convert this problem into a fixed point equation, we define $v_{ti}(B) = \max_{p \in \mathbb{R}} \{f_{ti}(p) : p + 1/h'_{ti}(p) = B\}$, which corresponds to the maximum value the right hand side of condition (a) can take when the condition (b) is satisfied. As we show in our proof, $v_{ti}(B)$ is a continuous and strictly decreasing function of B . This implies that the optimum B value uniquely satisfies the fixed point equation $B = \sum_{i \in S} v_{ti}(B)$. Lastly, we incorporate the assortment selection into this optimization problem and show that the assortment and price optimization can be achieved by solving the fixed point equation 3. See Appendix B for details. \square

Remark 3.3. Wang (2013) provides a weaker version of Proposition 3.2 that requires the additional assumption that the utility functions $h_{ti}(p)$ are twice-differentiable and concave in p . Even though this assumption holds for linear utility functions, the learning algorithm that we will introduce in the following sections requires us to solve the assortment and price optimization problem under non-concave utility functions.

Optimization Algorithm To solve the fixed point equation (3), we start by showing that its right-hand side is a positive and strictly decreasing function in B . We also show that if the utility functions satisfy $h_{ti}(0) \leq 1$ and $h'_{ti}(p) \leq -L_0$ for all $p \in \mathbb{R}$, then the solution to (3) lies in the interval $[0, P_0]$ for some $P_0 = \mathcal{O}(\log K/L_0)$. Note that this condition holds for the true utility function $u_{ti}(p)$. Under this condition, we can use a binary-search based algorithm to find the fixed point over the interval $[0, P_0]$. For future reference, we describe this procedure in Algorithm 1.

Computational Complexity The main difficulty in running Algorithm 1 is finding the set $\mathcal{P}_{ti}(B)$ that contains the solutions for the equation $p + 1/h'_{ti}(p) = B$ for any given B . Fortunately, for utility functions $h_{ti}(p)$

Algorithm 1 Assortment and price optimization

```

1: Input: utility functions  $h_{ti}(p)$  for  $i \in [n]$ 
2: Input: precision parameter  $\epsilon$ 
3: Input: search interval  $[0, P_0]$ 
4:  $B_\ell = 0, B_r = P_0$ 
5: while  $B_r - B_\ell > \epsilon$  do
6:    $B \leftarrow (B_r + B_\ell)/2$ 
7:   for  $i \in [N]$  do
8:     Find  $\mathcal{P}_{ti}(B) = \{p : p + 1/h'_{ti}(p) = B\}$ 
9:      $v_{ti} \leftarrow \max\{f_{ti}(p) : p \in \mathcal{P}_{ti}(B)\}$ 
10:     $B^* = \max_{S \in \mathcal{S}_K} \sum_{i \in S} v_{ti}$ 
11:    if  $B > B^*$  then  $B_r \leftarrow B$  else  $B_\ell \leftarrow B$ 
12: Output:  $B^*$ 

```

that we will use in Algorithms 2 and 3, we can show that there are only a small number of solutions (i.e., $\Theta(1)$), and these solutions can be efficiently computed. (See Appendix B.1 for details.) Since each iteration of this binary-search based algorithm requires us to compute the v_{ti} value for all $i \in [N]$, the algorithm has an overall computational complexity of $\Theta(N \log(P_0/\epsilon))$ for any arbitrary precision ϵ .

4 ONLINE LEARNING

In this section, we discuss how to estimate parameters based on user choices, introduce our online learning algorithms, and provide our regret bounds.

4.1 MLE for Multinomial Logistic Regression

Since the seller does not have access to problem parameters $\psi^* \in \mathbb{R}^d$ and $\phi^* \in \mathbb{R}^d$, it cannot directly compute the optimum assortments and prices. Therefore, it needs to construct an estimate of the parameters based on the history H_t of observations.

For convenience, we let $\boldsymbol{\theta} = (\psi, \phi)$ and $\tilde{\mathbf{x}}_{ti} = (\mathbf{x}_{ti}, -p_{ti}\mathbf{x}_{ti})$ denote the extended parameter and feature vectors such that $\langle \boldsymbol{\theta}, \tilde{\mathbf{x}}_{ti} \rangle = \langle \psi, \mathbf{x}_{ti} \rangle - \langle \phi, \mathbf{x}_{ti} \rangle \cdot p_{ti}$.

Then, we write the MNL choice probabilities under some parameter $\boldsymbol{\theta} = (\psi, \phi)$ using the notation

$$q_{ti}(\boldsymbol{\theta}) = \frac{e^{\langle \boldsymbol{\theta}, \tilde{\mathbf{x}}_{ti} \rangle}}{1 + \sum_{j \in S_t} e^{\langle \boldsymbol{\theta}, \tilde{\mathbf{x}}_{tj} \rangle}}.$$

With this notation, the negative log-likelihood function for the observations up to time t is given by

$$\ell_t(\boldsymbol{\theta}) := - \sum_{s=1}^{t-1} \log q_{si}(\boldsymbol{\theta}). \quad (4)$$

The maximum likelihood estimator is the parameter $\hat{\boldsymbol{\theta}}$ that minimizes $\ell_t(\boldsymbol{\theta})$ over the parameter space. Since

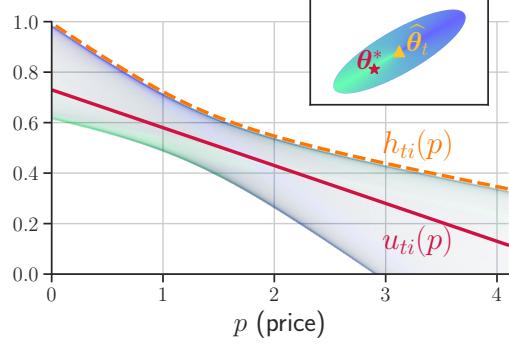


Figure 2: The confidence region depicted in the top right corner contains the true parameter $\boldsymbol{\theta}^*$ with high probability. Each parameter in the confidence set corresponds to a different linear function and we construct $h_{ti}(p)$ as a tight upper bound to $u_{ti}(p)$.

$\ell_t(\boldsymbol{\theta})$ is convex, we can use gradient-based convex optimization methods to find an MLE solution (Boyd and Vandenberghe, 2004). See Appendix A for details.

4.2 Algorithm

Our core idea is to construct a tight, high-confidence upper bound for the revenue $R_t(S, \mathbf{p})$ as a function of $S \in \mathcal{S}_K$ and $\mathbf{p} \in \mathbb{R}^N$, and to determine the assortments and prices according to the *optimism* principle in order to ensure low regret. The upper confidence bound (UCB) techniques and the *optimism in the face of uncertainty* principle have been widely known to be effective in balancing the exploration and exploitation in many bandit problems (Lattimore and Szepesvári, 2020; Abbasi-Yadkori, 2011; Li et al., 2017). The key distinction of our approach lies in the construction and analysis of functional upper bounds, which capture the continuous dependence of revenue on prices. In particular, we construct a pointwise confidence upper bound $h_{ti}(p)$ for each utility function $u_{ti}(p)$, i.e., $h_{ti}(p) \geq u_{ti}(p)$ for all $p \in \mathbb{R}$ with high probability. In order to achieve low regret rates, it is crucial to obtain tight upper bounds as depicted in Figure 2.

We offer randomly selected assortments and prices for the first T_0 rounds to ensure that our maximum likelihood estimates $\hat{\boldsymbol{\theta}}_t = (\hat{\psi}_t, \hat{\phi}_t)$ in subsequent rounds are sufficiently close to the true parameter $\boldsymbol{\theta}^*$. This allows us to construct a matrix \mathbf{V}_t as a tight estimate of the Fisher Information Matrix around $\boldsymbol{\theta}^*$ (please refer to Algorithm 2 for the definition of \mathbf{V}_t). Then, we obtain confidence regions of the form $\{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_t} \leq \alpha_t\}$ for some confidence radius α_t such that $\boldsymbol{\theta}^*$ is contained within the region with high probability. In contrast to prior works (Chen et al., 2020; Oh and Iyengar, 2021), we use estimated choice probabilities $q_{ti}(\hat{\boldsymbol{\theta}}_t)$ in our \mathbf{V}_t construction, which is the key in achieving a better scaling of α_t with respect to K and L_0 .

Algorithm 2 CAP: Contextual Assortment and Pricing under MNL Model

```

1: Input: initialization rounds  $T_0$ , confidence parameters  $\{\alpha_t\}_{t \in [T]}$ , minimum price sensitivity  $L_0$ 
2:  $\mathbf{V}_1 \leftarrow \mathbf{0} \in \mathbb{R}^{2d \times 2d}$ 
3: for  $t = 1, 2, \dots, T_0 - 1$  do ▷ Initialization rounds
4:   Choose  $S_t$  uniformly at random from  $\{S \subseteq [n] : |S| \leq K\}$ 
5:   Choose  $p_{ti}$  independently and uniformly at random from  $[1, 2]$  for all  $i \in S_t$ 
6:   Offer assortment  $S_t$  at price  $\mathbf{p}_t$  and observe  $i_t$ 
7:    $\mathbf{V}_{t+1} \leftarrow \mathbf{V}_t + \frac{1}{K^2} \sum_{i \in S_t} \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top$ 
8: for  $t = T_0, T_0 + 1, \dots, T$  do ▷ MLE computation
9:   Compute  $\hat{\boldsymbol{\theta}}_t = (\hat{\boldsymbol{\psi}}_t, \hat{\boldsymbol{\phi}}_t)$  by minimizing (4) ▷ Price-dependent confidence function
10:  Let  $g_{ti}(p) := \alpha_t \|(\mathbf{x}_{ti}, -p \mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}}$  for all  $i \in [n]$ 
11:  Let  $\tilde{h}_{ti}(p) := \langle \hat{\boldsymbol{\psi}}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\boldsymbol{\phi}}_t, \mathbf{x}_{ti} \rangle \cdot p + g_{ti}(p)$  for all  $i \in [n]$ 
12:  Let  $h_{ti}(p) := \min_{p' \leq p} \{\tilde{h}_{ti}(p') - L_0(p - p')\}$  for all  $i \in [n]$  ▷ Utility function estimate
13:  Choose  $(S_t, \mathbf{p}_t)$  using Algorithm 1 with estimated utility functions  $h_{ti}(p)$ 
14:  Offer assortment  $S_t$  at price  $\mathbf{p}_t$  and observe  $i_t$ 
15:   $\mathbf{V}_{t+1} \leftarrow \mathbf{V}_t + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{tj}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top$  ▷ Information estimate

```

Based on these confidence regions for the parameter, we obtain an intermediate utility upper bound

$$\tilde{h}_{ti}(p) := \langle \hat{\boldsymbol{\psi}}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\boldsymbol{\phi}}_t, \mathbf{x}_{ti} \rangle \cdot p + g_{ti}(p)$$

where $g_{ti}(p) := \alpha_t \|(\mathbf{x}_{ti}, -p \mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}}$ is a *price-dependent* confidence bonus. Note that $\tilde{h}_{ti}(p)$ is a convex and differentiable function. However, it is not necessarily a decreasing function and hence we cannot immediately use our Proposition 3.2 to find optimum assortments and prices under $\tilde{h}_{ti}(p)$. To resolve this problem, we use the fact that $u'_{ti}(p) \leq -L_0$ for all $p \in \mathbb{R}$, and construct a tighter upper bound

$$h_{ti}(p) := \min_{p' \leq p} \{\tilde{h}_{ti}(p') - L_0(p - p')\}.$$

As a result, we can replace each $u_{ti}(p)$ in (1) with $h_{ti}(p)$ to obtain an upper bound for the revenue function as

$$\tilde{R}_t(S, \mathbf{p}) := \frac{\sum_{i \in S_t} p_{ti} \exp\{h_{ti}(p_{ti})\}}{1 + \sum_{j \in S_t} \exp\{h_{tj}(p_{tj})\}}. \quad (5)$$

As we verify in our proofs, this estimate satisfies $\tilde{R}_t(S, \mathbf{p}) \geq R_t(S, \mathbf{p})$ for any $S \in \mathcal{S}_K$ and any $\mathbf{p} \in \mathbb{R}^N$. Using \tilde{R}_t as a proxy for R_t , we choose the assortments and prices according to

$$(S_t, \mathbf{p}_t) \in \operatorname{argmax}_{\substack{S \in \mathcal{S}_K \\ \mathbf{p} \in \mathbb{R}_+^n}} \tilde{R}_t(S, \mathbf{p}). \quad (6)$$

As discussed in Section 3, we can efficiently solve this optimization problem using Algorithm 1 since $h_{ti}(p)$ are differentiable and strictly decreasing.

4.3 Regret Analysis

Our main result presented in Theorem 4.2 concerns the regret upper bound for Algorithm 2. We show this result under the following regularity assumption on the context process which is a standard assumption made in the generalized linear bandit (Li et al., 2017) and MNL contextual bandit (Chen et al., 2020; Oh and Iyengar, 2021) literature.

Assumption 4.1 (Stochastic and bounded features). Each feature vector \mathbf{x}_{ti} is an independent random variable with unknown distribution; they satisfy $\|\mathbf{x}_{ti}\| \leq 1$, and there exists a constant $\sigma_0 > 0$ such that $\mathbb{E}[\mathbf{x}_{ti} \mathbf{x}_{ti}^\top] \succcurlyeq \sigma_0 \mathbf{I}$. Furthermore, parameter vectors satisfy $\|(\boldsymbol{\psi}^*, \boldsymbol{\phi}^*)\| \leq 1$.

Accordingly, we can demonstrate in Theorem 4.2 that Algorithm 2 enjoys $\tilde{\mathcal{O}}(d\sqrt{KT}/L_0)$ regret bound in terms of key problem primitives N , K , d , L_0 , and T . This regret rate is independent of the number of items N , and is thus applicable in settings with a large number of candidate items.

Theorem 4.2. *Suppose Assumptions 2.1, 3.1, and 4.1 hold and we run CAP (Algorithm 2) with initialization length T_0 given in (13) and confidence width α_t given in (15). Then, the expected regret for a sufficiently large time horizon T is upper-bounded as*

$$\mathcal{R}_T \leq C_1 \cdot \frac{\log K}{L_0} d \sqrt{KT \log T \log(T/d)}$$

for a constant C_1 independent of N , K , d , L_0 , and T .

Proof. (Sketch) In proving our regret bounds, we first show that the optimum prices p_{ti}^* are bounded within

$[0, P]$ for some $P = \mathcal{O}(\log K/L_0)$ under our utility estimations $h_{ti}(p)$. Then, we show that $T_0 = \Theta(\sigma_0^{-3}dP^2K\log^2 T)$ initialization steps are enough to ensure $\|\hat{\theta}_t - \theta^*\|_2 = \mathcal{O}(1/P)$ for all $t \geq T_0$. This result enables us to estimate the Fisher Information Matrix around θ^* within a constant factor using \mathbf{V}_t . Next, we establish a confidence region $\{\theta : \|\theta - \theta_t\|_{\mathbf{V}_t} \leq \alpha_t\}$ for θ^* with $\alpha_t = \mathcal{O}(\sigma_0^{-1}d\log T)$, importantly noting that α_t is independent of both K and L_0 . Here, we use a novel Bernstein-type inequality for self-normalized vector-valued martingales which allows us to fully capture the correlation structure between our observations with the help of $q_{ti}(\hat{\theta}_t)$. Based on these confidence regions, we construct optimistic utility estimate functions $h_{ti}(p)$ as described in Section 4.2. The selection of assortment and prices according to $h_{ti}(p)$ allows us to obtain an upper bound for the regret incurred at each time step, and hence an upper bound for \mathcal{R}_T . Please see Appendix C for details. \square

Remark 4.3. Our analysis in this work assumes that the L_0 parameter, or a lower bound of it, is known to the algorithm. However, as we describe in Appendix C.2, it is possible to relax this assumption and estimate L_0 to achieve the same regret rates.

4.4 Extension to Online Parameter Update

Algorithm 2 is simple to implement and enjoys provable regret bounds as shown in Theorem 4.2. However, the computation of the MLE at each round requires access to all feature vectors corresponding to previous assortments. To reduce the time and space complexities of our algorithm, we can instead use a variant of the online Newton step rule from Hazan et al. (2014). The online version presented as Algorithm 3 in Appendix F finds an approximate MLE solution only using the context vectors of the last assortment. We show that the modified algorithm enjoys the following regret rate.

Theorem 4.4. *Suppose Assumptions 2.1, 3.1, and 4.1 hold and we run CAP-ONS (Algorithm 3) with initialization length T_0 given in (28) and confidence width α_t given in (30). Then, the expected regret for a sufficiently large time horizon T satisfies*

$$\mathcal{R}_T = \tilde{\mathcal{O}}(dK\sqrt{T}/L_0).$$

4.5 Regret Lower Bound

In this section, we provide a regret lower bound of order $\Omega(d\sqrt{T}/L_0)$ in terms of key problem primitives N , d , and T for the problem of assortment selection and pricing under the contextual MNL choice model. This result demonstrates that CAP (Algorithm 2) and CAP-ONS (Algorithm 3) are optimal, up to logarithmic terms in d , T , and L_0 .

Theorem 4.5. *For any maximum assortment size K , any minimum price sensitivity $L_0 > 0$, any context dimension d divisible by 4, and any policy π , there exists a worst-case problem instance with $n = \Theta(K \cdot 2^d)$ items, bounded context vectors (i.e., $\|\mathbf{x}_{ti}\| \leq 1$ for all $i \in [n]$), and bounded feature vectors (i.e., $\|(\theta^*; \phi^*)\| \leq 1$) such that the regret of policy π is lower bounded as*

$$\mathcal{R}_T(\pi) \geq C_3 \cdot d\sqrt{T}/L_0$$

for some universal constant $C_3 > 0$.

Proof. (Sketch) We reduce the task of lower bounding the worst-case regret to lower bounding the Bayes risk over an adversarial parameter set. Then, we use a counting argument similar to the one used in Chen et al. (2020) to provide an explicit lower bound on the Bayes risk. See Appendix G for details. \square

5 NUMERICAL EXPERIMENTS

We demonstrate the efficacy of our proposed algorithms: CAP presented in Algorithm 2 and its online version CAP-ONS in Algorithm 3. We numerically evaluate our algorithms over independently generated problem instances and provide our results in Figure 3. In each instance, we generate problem parameters $(\psi^*; \phi^*)$ and context vectors \mathbf{x}_{ti} by sampling their entries from uniform distributions such that we satisfy Assumptions 3.1 and 4.1. See Appendix H for further details. The code for our experiments is available at https://github.com/basics-lab/assortment_selection_pricing.

We compare the performance of our proposed algorithms against state-of-the-art algorithms designed for the MNL choice model. Since the literature primarily focuses on either assortment selection or pricing, our baselines concentrate solely on either assortment selection or pricing. Figure 3 illustrates that our algorithms, which simultaneously address both assortment selection and pricing, outperform baseline methods.

Our baselines include two MNL pricing algorithms: M3P (Javanmard et al., 2020) and ONS-MPP (Perivier and Goyal, 2022). These algorithms are designed to optimize prices under the assumption that all N items can be offered without any need for assortment selection. To account for the requirements of our experimental setting, we consider that only top K items (based on their estimated utility value) are offered with chosen prices. These pricing-only algorithms perform comparably when (since there is no assortment decision to be made) but their performance deteriorates as $N \gg K$.

We also consider two MNL assortment selection algorithms as baselines: DBL-MNL (Oh and Iyengar, 2021) and TS-MNL (Oh and Iyengar, 2019). Since both of these algorithms are designed specifically for assortment selection under fixed prices, they cannot achieve

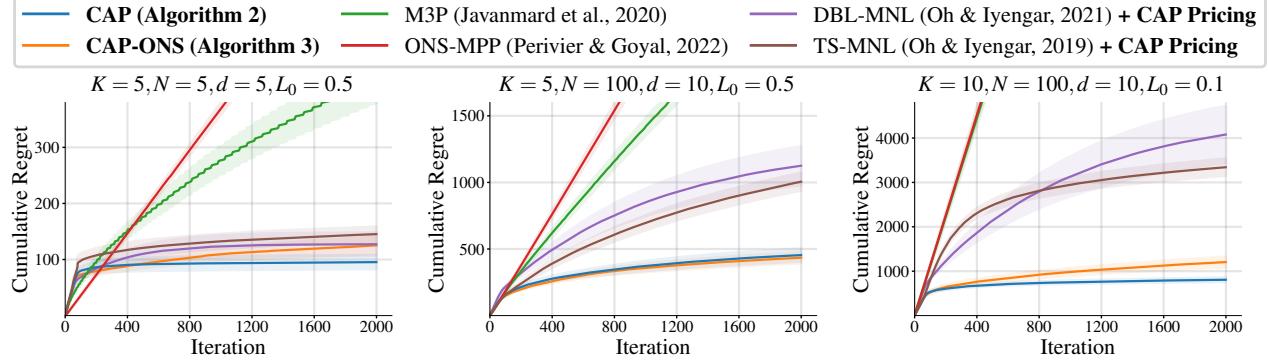


Figure 3: Cumulative regret for CAP (Algorithm 2), CAP-ONS (Algorithm 3), M3P (Javanmard et al., 2020), ONS-MPP (Perivier and Goyal, 2022), a version of DBL-MNL (Oh and Iyengar, 2021) extended with our dynamic pricing, and a version of TS-MNL (Oh and Iyengar, 2019) extended with our dynamic pricing. The center lines show the mean across the runs while the shaded regions indicate two standard deviations. Results demonstrate the efficacy of our algorithms in achieving diminishing regret per round as our theoretical results predict. Since M3P and ONS-MPP consider only dynamic pricing, they are not able to achieve diminishing regret. DBL-MNL and TS-MNL are designed solely for assortment selection, but their extensions using our pricing approach enable simultaneous assortment selection and pricing. However, even with dynamic pricing, their regret rates quickly deteriorate as K increases or L_0 decreases.

diminishing regret in their original form. Therefore, we use our pricing approach to implement heuristic extensions of these algorithms applicable for the joint assortment selection and pricing setting. These extensions utilize the respective frameworks to derive linear estimates for the utility functions and determine the optimal assortments and prices using our Algorithm 1. In our empirical studies, these heuristic extensions are able to achieve diminishing regret, but the regret gaps between these algorithms and CAP increase as K increases or L decreases.

6 CONCLUSION

We study the joint problem of contextual assortment selection and pricing in which a seller aims to maximize cumulative revenue over a horizon. The user’s choice behavior follows a multinomial logit model with unknown parameters, and the seller learns from sequential user feedback. We propose an algorithm that achieves $\tilde{O}(d\sqrt{KT}/L_0)$ regret in T rounds where d is the dimension of the context vectors, K is the assortment size, and L_0 is the minimum price sensitivity. We show that this regret rate is optimal up to logarithmic terms in d , T , N , and L_0 .

Acknowledgements

This work was supported in part by NSF grants CIF-2007669 and CIF-1937357.

References

- Abbasi-Yadkori, Y. (2011). Improved algorithms for linear stochastic bandits. In *Proceedings of Twenty-Fifth Conference on Neural Information Processing Systems (NeurIPS)*, pages 2312–2320.
- Agrawal, P., Tulabandhula, T., and Avadhanula, V. (2020). A tractable online learning algorithm for the multinomial logit contextual bandit. *arXiv e-prints*, pages arXiv–2011.
- Agrawal, S., Avadhanula, V., Goyal, V., and Zeevi, A. (2017). Thompson sampling for the MNL-bandit. In *Proceedings of the 2017 Conference on learning theory (COLT)*, page 76–78.
- Agrawal, S., Avadhanula, V., Goyal, V., and Zeevi, A. (2019). MNL-bandit: A dynamic learning approach to assortment selection. *Operations Research*, 67(5):1453–1485.
- Amani, S. and Thrampoulidis, C. (2021). Ucb-based algorithms for multinomial logistic regression bandits.
- Aouad, A., Levi, R., and Segev, D. (2018). Greedy-like algorithms for dynamic assortment planning under multinomial logit preferences. *Operations Research*, 66(5):1321–1345.
- Ban, G.-Y. and Keskin, N. B. (2021). Personalized dynamic pricing with machine learning: High-dimensional features and heterogeneous elasticity. *Management Science*, 67(9):5549–5568.
- Bartlett, P. L., Bousquet, O., and Mendelson, S. (2005).

- Local rademacher complexities. *The Annals of Statistics*, 33. arXiv:math/0508275.
- Besbes, O. and Zeevi, A. (2009). Dynamic pricing without knowing the demand function: Risk bounds and near-optimal algorithms. *Operations Research*, 57(6):1407–1420.
- Boyd, S. P. and Vandenberghe, L. (2004). *Convex Optimization*. Cambridge University Press.
- Bubeck, S., Devanur, N. R., Huang, Z., and Niazadeh, R. (2019). Multi-scale online learning: Theory and applications to online auctions and pricing. *Journal of Machine Learning Research*, 20(62):1–37.
- Caro, F. and Gallien, J. (2007). Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*, 53(2):276–292.
- Chen, K., Hu, I., and Ying, Z. (1999). Strong consistency of maximum quasi-likelihood estimators in generalized linear models with fixed and adaptive designs. *The Annals of Statistics*, 27(4):1155–1163.
- Chen, W., Wang, Y., and Yuan, Y. (2013). Combinatorial multi-armed bandit: General framework and applications. In *Proceedings of the Thirtieth International Conference on Machine Learning (ICML)*, volume 28, pages 151–159.
- Chen, X. and Wang, Y. (2018). A note on a tight lower bound for capacitated MNL-bandit assortment selection models. *Operations Research Letters*, 46(5):534–537.
- Chen, X., Wang, Y., and Zhou, Y. (2020). Dynamic assortment optimization with changing contextual information. *The Journal of Machine Learning Research*, 21(1):216:8918–216:8961.
- Cheung, W. C. and Simchi-Levi, D. (2017). Thompson sampling for online personalized assortment optimization problems with multinomial logit choice models. Available at SSRN.
- Chu, W., Li, L., Reyzin, L., and Schapire, R. (2011). Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, page 208–214.
- Faury, L., Abeille, M., Calauzènes, C., and Fercoq, O. (2020). Improved optimistic algorithms for logistic bandits. In *Proceedings of the Thirty-Seventh International Conference on Machine Learning (ICML)*.
- Freedman, D. A. (1975). On tail probabilities for martingales. *The Annals of Probability*, 3:100–118.
- Hazan, E., Koren, T., and Levy, K. Y. (2014). Logistic regression: Tight bounds for stochastic and online optimization. In *Proceedings of The Twenty-Seventh Conference on Learning Theory (COLT)*, page 197–209. PMLR.
- Javanmard, A., Nazerzadeh, H., and Shao, S. (2020). Multi-product dynamic pricing in high-dimensions with heterogeneous price sensitivity. In *Proceedings of The 2020 IEEE International Symposium on Information Theory (ISIT)*, page 2652–2657.
- Kleinberg, R. and Leighton, T. (2003). The value of knowing a demand curve: bounds on regret for online posted-price auctions. In *Proceedings of the Forty-Fourth Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 594–605.
- Kohler, J. M. and Lucchi, A. (2017). Sub-sampled cubic regularization for non-convex optimization. In *Proceedings of the Thirty-Fourth International Conference on Machine Learning (ICML)*.
- Kveton, B., Wen, Z., Ashkan, A., and Szepesvári, C. (2015). Tight Regret Bounds for Stochastic Combinatorial Semi-Bandits. In *Proceedings of the Eighteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, volume 38, pages 535–543.
- Lattimore, T. and Szepesvári, C. (2020). *Bandit Algorithms*. Cambridge University Press, 1 edition.
- Li, L., Lu, Y., and Zhou, D. (2017). Provably optimal algorithms for generalized linear contextual bandits. In *Proceedings of the 34th International Conference on Machine Learning*, page 2071–2080. PMLR.
- McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*.
- Miao, S. and Chao, X. (2018). Dynamic joint assortment and pricing optimization with demand learning. *Manufacturing & Service Operations Management*, 23(2):525–545.
- Oh, M. and Iyengar, G. (2021). Multinomial logit contextual bandits: Provable optimality and practicality. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 9205–9213.
- Oh, M.-h. and Iyengar, G. (2019). Thompson sampling for multinomial logit contextual bandits. *Advances in Neural Information Processing Systems*, 32.
- Paes Leme, R. and Schneider, J. (2018). Contextual search via intrinsic volumes. In *Proceedings of the Fifty-Ninth Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 268–282.
- Perivier, N. and Goyal, V. (2022). Dynamic pricing and assortment under a contextual mnl demand. In *Proceedings of Thirty-Fifth Conference on Neural Information Processing Systems (NeurIPS)*.
- Qin, L., Chen, S., and Zhu, X. (2014). Contextual combinatorial bandit and its application on diversified online recommendation. In *Proceedings of the 2014 SIAM International Conference on Data Mining (SDM)*, Proceedings, page 461–469.

- Wang, H., Talluri, K., and Li, X. (2025). On dynamic pricing with covariates. *Operations Research*.
- Wang, R. (2013). Capacitated assortment and price optimization under the multinomial logit model. *Operations Research Letters*.
- Xu, J. and Wang, Y.-X. (2021). Logarithmic regret in feature-based dynamic pricing. In *Proceedings of Thirty-Fifth Conference on Neural Information Processing Systems (NeurIPS)*, pages 13898–13910.
- Xu, J. and Wang, Y.-X. (2024). Pricing with contextual elasticity and heteroscedastic valuation. In *International Conference on Machine Learning*, pages 55286–55304. PMLR.
- Zhang, L., Yang, T., Jin, R., Xiao, Y., and Zhou, Z.-h. (2016). Online stochastic linear optimization under one-bit feedback. In *Proceedings of the Thirty-Third International Conference on Machine Learning (ICML)*, pages 392–401.
- Zong, S., Ni, H., Sung, K., Ke, N. R., Wen, Z., and Kveton, B. (2016). Cascading bandits for large-scale recommendation problems. In *Proceedings of the Thirty-Second Conference on Uncertainty in Artificial Intelligence (UAI)*, page 835–844.

Checklist

1. For all models and algorithms presented, check if you include:
 - (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. [Yes]
 - (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. [Yes]
 - (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. [Yes, URL is provided.]
2. For any theoretical claim, check if you include:
 - (a) Statements of the full set of assumptions of all theoretical results. [Yes]
 - (b) Complete proofs of all theoretical results. [Yes]
 - (c) Clear explanations of any assumptions. [Yes]
3. For all figures and tables that present empirical results, check if you include:
 - (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). [Yes, URL is provided.]
 - (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). [Yes, provided in Section 5 and Appendix H]
 - (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). [Yes, provided in the caption for Figure 3]
 - (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). [Not Applicable]
4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
 - (a) Citations of the creator If your work uses existing assets. [Not Applicable]
 - (b) The license information of the assets, if applicable. [Not Applicable]
 - (c) New assets either in the supplemental material or as a URL, if applicable. [Not Applicable]
 - (d) Information about consent from data providers/curators. [Not Applicable]
 - (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. [Not Applicable]

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
 - (a) The full text of instructions given to participants and screenshots. [Not Applicable]
 - (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. [Not Applicable]
 - (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. [Not Applicable]

A Properties of Maximum Likelihood Estimation

Proposition A.1. *The maximum likelihood estimator is any parameter $\hat{\boldsymbol{\theta}}_t$ that minimizes the negative log-likelihood function over the parameter space, that is*

$$\hat{\boldsymbol{\theta}}_t \in \operatorname{argmin}_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}). \quad (7)$$

The negative log-likelihood function $\ell_t(\boldsymbol{\theta})$ is convex over $\boldsymbol{\theta} \in \mathbb{R}^{2d}$. Furthermore, if the Fisher information matrix $\mathcal{I}_t(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta})$ is positive definite, then $\ell_t(\boldsymbol{\theta})$ is strongly convex and thus admits a unique minimizer.

For each item $i \in S_t \cup \{0\}$, we define the choice response variables $y_{ti} = \mathbf{1}\{i_t = i\} \in \{0, 1\}$. Then, the gradient of these probabilities with respect to $\boldsymbol{\theta}$ can be written as

$$\nabla_{\boldsymbol{\theta}} q_{ti}(\boldsymbol{\theta}) = q_{ti}(\boldsymbol{\theta}) \left(\tilde{\mathbf{x}}_{ti} - \sum_{j \in S_t} q_{tj}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{tj} \right).$$

On the other hand, we can write the negative log-likelihood function at time t as

$$\ell_t(\boldsymbol{\theta}) := - \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau \cup \{0\}} y_{ti} \log q_{\tau i}(\boldsymbol{\theta}).$$

Calculating the gradient of this negative log-likelihood with respect to $\boldsymbol{\theta}$ we obtain

$$\nabla_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) = \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau} (q_{\tau i}(\boldsymbol{\theta}) - y_{\tau i}) \tilde{\mathbf{x}}_{\tau i}$$

On the other hand, the Hessian of the negative log-likelihood is given by

$$\begin{aligned} \nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta}) &= \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau} q_{\tau i}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{\tau i} \left(\tilde{\mathbf{x}}_{\tau i} - \sum_{j \in S_t} q_{\tau j}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{\tau j} \right)^\top \\ &= \sum_{\tau=1}^{t-1} \left(\sum_{i \in S_\tau} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top - \sum_{i \in S_\tau} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{\tau j}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau j}^\top \right). \end{aligned}$$

Since this log-likelihood satisfies the necessary regularity conditions, the Hessian of the negative log-likelihood is also equal to the Fisher information matrix $\mathcal{I}_t(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta})$.

Now, let $\mathbf{q}_t(\boldsymbol{\theta})$ denote the vector of probabilities $q_{ti}(\boldsymbol{\theta})$ and let $\tilde{\mathbf{X}}_t$ be the matrix with columns $\tilde{\mathbf{x}}_{ti}$ for $i \in S_t$, we can write

$$\nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta}) = \sum_{\tau=1}^{t-1} \tilde{\mathbf{X}}_\tau \Sigma_\tau(\boldsymbol{\theta}) \tilde{\mathbf{X}}_\tau^\top.$$

where $\Sigma_t(\boldsymbol{\theta}) = \operatorname{diag}(\mathbf{q}_t(\boldsymbol{\theta})) - \mathbf{q}_t(\boldsymbol{\theta}) \mathbf{q}_t(\boldsymbol{\theta})^\top$. Since we have $q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) > 0$ for all $\boldsymbol{\theta} \in \mathbb{R}^{2d}$, we conclude that $\Sigma_t(\boldsymbol{\theta}) \succ \mathbf{0}$ for all $\boldsymbol{\theta} \in \mathbb{R}^{2d}$.

Therefore, $\nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta}) \succ \mathbf{0}$ for all $\boldsymbol{\theta} \in \mathbb{R}^{2d}$. Hence, the negative log-likelihood is convex with respect to $\boldsymbol{\theta}$. As a result, any $\boldsymbol{\theta}$ that satisfies the first-order optimality condition $\nabla_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) = \mathbf{0}$ is a minimizer.

Furthermore, if we are given that the Fisher Information Matrix $\nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta})$ is positive definite, i.e. $\nabla_{\boldsymbol{\theta}}^2 \ell_t(\boldsymbol{\theta}) \succ \mathbf{0}$, the negative log-likelihood function becomes strongly convex with respect to $\boldsymbol{\theta}$. Consequently, we have a unique MLE solution.

B Solving the Assortment and Price Optimization Problem

As stated in Proposition 3.2, we make the following regularity assumption for the utility functions.

Assumption B.1. For each item $i \in [N]$, the utility function $u_i(p_i)$ is differentiable, strictly decreasing in price p_i , and satisfies $\lim_{p \rightarrow \infty} p e^{u_i(p)} = 0$.

We first consider the price optimization for a given assortment S . That is,

$$\max_{\mathbf{p} \in \mathbb{R}^N} \sum_{i \in S} p_i \cdot q_i(\mathbf{p})$$

where $q_i(\mathbf{p})$ denotes the probability of choosing item i under prices $\mathbf{p} \in \mathbb{R}^N$.

Proposition B.2. Fix some assortment $S \subseteq [N]$. Under Assumption B.1, the quantity $p_i^* + 1/u'_i(p_i^*)$ is constant for all $i \in S$ at the optimal price vector \mathbf{p}^* . Moreover, $p_i^* + 1/u'_i(p_i^*)$ is equal to the total revenue obtained by pricing at \mathbf{p}^* .

Proof. The first-order condition for optimality is

$$\nabla_{\mathbf{p}} \left\{ \sum_{i \in S} p_i \cdot q_i(\mathbf{p}) \right\} = \mathbf{0}.$$

For any $i \in S$, it is straightforward to verify that

$$\begin{aligned} \frac{\partial q_i(\mathbf{p})}{\partial p_i} &= u'_i(p_i) \cdot q_i(\mathbf{p})(1 - q_i(\mathbf{p})) < 0, \\ \frac{\partial q_j(\mathbf{p})}{\partial p_i} &= -u'_i(p_i) \cdot q_j(\mathbf{p})q_i(\mathbf{p}) > 0, \forall j \in S, j \neq i. \end{aligned}$$

Therefore, for each $i \in S$, we need

$$q_i(\mathbf{p}) + p_i u'_i(p_i) q_i(\mathbf{p})(1 - q_i(\mathbf{p})) - \sum_{j \in S, j \neq i} p_j u'_i(p_i) q_i(\mathbf{p}) q_j(\mathbf{p}) = 0.$$

Rearranging the above equation results in

$$q_i(\mathbf{p}) u'_i(p_i) \left[p_i + \frac{1}{u'_i(p_i)} - \sum_{j \in S} p_j q_j(\mathbf{p}) \right] = 0.$$

Since $q_i(\mathbf{p}) > 0$ and $u'_i(p_i) < 0$ for all $i \in S$ and for all $p_i \in \mathbb{R}$, the above equation is equivalent to

$$p_i + \frac{1}{u'_i(p_i)} = \sum_{i \in S} p_i q_i(\mathbf{p}). \quad (8)$$

The right-hand side of this equation is independent of the item index i , so $p_i^* + 1/u'_i(p_i^*)$ is constant for all $i \in S$ at an optimal price vector \mathbf{p}^* . Moreover, this equality shows that the $p_i^* + 1/u'_i(p_i^*)$ quantity is equal to the total revenue at the optimal price vector \mathbf{p}^* .

□

Let us denote this constant quantity as $\theta = p_i + 1/u'_i(p_i)$.

Remark B.3. Under the concavity and twice-differentiability assumptions made in Wang (2013), it is possible to show that there exists a unique p_i value that satisfies $\theta = p_i + 1/u'_i(p_i)$ for all different values of θ . In other words, it is possible to show that, $\theta = p_i + 1/u'_i(p_i)$ is a one-to-one relation. However, it is not possible to show this property only under Assumption B.1.

For some values of θ , there might exist multiple p_i values that satisfy $\theta = p_i + 1/u'_i(p_i)$. Let $\mathcal{P}_i(\theta)$ denote this set of prices that satisfy $\theta = p_i + 1/u'_i(p_i)$.

With this notation, the necessary condition for optimality in (8) can be equivalently written as

$$\begin{aligned}\theta &= \sum_{i \in S} \left(\theta - \frac{1}{u'_i(p_i)} \right) q_i(\mathbf{p}) \quad \text{and} \\ p_i &\in \mathcal{P}_i(\theta), \quad \forall i \in S.\end{aligned}$$

Noting that $1 - \sum_{i \in S} q_i(\mathbf{p}) = q_0(\mathbf{p})$ and $q_i(\mathbf{p})/q_0(\mathbf{p}) = e^{u_i(p_i)}$, we have the equivalent conditions

$$\begin{aligned}\theta &= - \sum_{i \in S} \frac{e^{u_i(p_i)}}{u'_i(p_i)} \quad \text{and} \\ p_i &\in \mathcal{P}_i(\theta), \quad \forall i \in S.\end{aligned}$$

Under these conditions, Proposition B.2 states that the objective function (revenue) is equal to the value of θ . Therefore, we can cast the following problem to find the optimum values for θ and \mathbf{p} .

$$\begin{aligned}\theta^* &= \max_{\substack{\theta \in \mathbb{R} \\ p_i \in \mathbb{R}, \forall i \in S}} \theta \\ \text{s.t.} \quad \theta &= - \sum_{i \in S} \frac{e^{u_i(p_i)}}{u'_i(p_i)}, \\ p_i &\in \mathcal{P}_i(\theta), \quad \forall i \in S.\end{aligned}\tag{9}$$

Now, we define the following functions

$$\begin{aligned}g_i(p_i) &= - \frac{e^{u_i(p_i)}}{u'_i(p_i)} \\ f_i(\theta) &= \max \{g_i(p_i) : p_i \in \mathcal{P}_i(\theta)\}\end{aligned}$$

In Lemma B.5, we show that $f_i(\theta) > 0$ is a continuous and strictly decreasing function of θ . As a result, we have the following proposition.

Proposition B.4. *Under Assumption B.1, the optimum objective value θ^* of problem (9) satisfies*

$$\sum_{i \in S} f_i(\theta^*) = \theta^*.$$

Proof. Assume $\sum_{i \in S} f_i(\theta) < \theta$ for some θ . Then, for any $\theta' \geq \theta$ and for any \mathbf{p} that satisfies $p_i \in \mathcal{P}_i(\theta')$ for all $i \in S$, we have $\theta' \geq \theta > \sum_{i \in S} f_i(\theta) \geq \sum_{i \in S} f_i(\theta') \geq \sum_{i \in S} g_i(p_i)$. Therefore, all $\theta' \geq \theta$ is infeasible and hence $\theta^* < \theta$.

Now, assume $\sum_{i \in S} f_i(\theta) > \theta$ for some θ . Then, since each $f_i(\theta) > 0$ is continuous and strictly decreasing, there exists some $\theta' > \theta$ such that $\sum_{i \in S} f_i(\theta') = \theta'$. Since this θ' is a feasible solution to the problem together with prices $p'_i = \operatorname{argmax}_{p_i} \{g_i(p_i) : p_i \in \mathcal{P}_i(\theta)\}$, we have $\theta^* \geq \theta' > \theta$.

Therefore, any optimum point $\theta = \theta^*$ must satisfy $\sum_{i \in S} f_i(\theta) = \theta$. \square

Using Proposition B.4, we can reduce the multi-product price optimization problem for any given assortment S to a single-dimensional problem given by

$$\max_{\theta \in \mathbb{R}} \left\{ \theta : \theta = \sum_{i \in S} f_i(\theta) \right\}.$$

Furthermore, since $f_i(\theta) > 0$ is strictly decreasing in θ , there exists a unique solution to the condition $\sum_{i \in S} f_i(\theta) = \theta$ for any given assortment S . Let us denote this unique feasible (and hence optimal) solution by θ_S^* .

The next step is to incorporate the assortment selection into our optimization problem. We can achieve this by considering the problem

$$\max \{\theta_S^* : S \in \mathcal{S}_K\} = \max_{\theta \in \mathbb{R}} \left\{ \theta : S \in \mathcal{S}_K \text{ and } \theta = \sum_{i \in S} f_i(\theta) \right\}. \quad (10)$$

This assortment selection problem in the given form requires searching all possible assortments of size at most K and there are $\sum_{\ell=1}^K \binom{N}{\ell} = \sum_{\ell=1}^K N! / ((N - \ell)! \ell!)$ assortments to consider. However, this search space can be significantly reduced by noticing that finding the unique fixed point of the equation

$$\theta = \max_{S \in \mathcal{S}_K} \sum_{i \in S} f_i(\theta). \quad (11)$$

is equivalent to solving (10). Since each $f_i(\theta)$ is strictly decreasing in θ and the right-hand side is strictly increasing in θ , there exists a unique solution θ^* to this equation.

Denote an assortment $S^* \in \mathcal{S}_K$ such that $\theta^* = \sum_{i \in S^*} f_i(\theta^*)$. Then, S^* is an optimal assortment together with prices $p_i^* = \operatorname{argmax}_{p_i} \{g_i(p_i) : p_i \in \mathcal{P}_i(\theta^*)\}$ for all $i \in S^*$.

As a result, we obtain the following proposition given in the main paper.

Proposition 3.2 (Optimum assortment and prices). *Suppose utility functions $h_{ti}(p)$ are differentiable and strictly decreasing for all items $i \in [N]$. Let B_t be the unique solution of the fixed point equation*

$$B = \max_{S \in \mathcal{S}_K} \sum_{i \in S} v_{ti}(B) \quad (3)$$

where $v_{ti}(B) := \max_{p \in \mathbb{R}} \{f_{ti}(p) : p + 1/h'_{ti}(p) = B\}$ and $f_{ti}(p) := -e^{h_{ti}(p)} / h'_{ti}(p)$. Then, the optimum assortment S_t^* is the assortment S that achieves the maximum in the optimization problem (3), the optimum prices are

$$p_{ti}^* = \operatorname{argmax}_{p \in \mathbb{R}} \{f_{ti}(p) : p + 1/h'_{ti}(p) = B_t\},$$

and the optimum revenue achieved by (S_t^*, \mathbf{p}_t^*) is B_t .

B.1 Assortment Selection and Pricing under Estimated Utility Functions

In the following section, we describe how we can **efficiently** run Algorithm 1 for the estimated utility functions $h_{ti}(p)$ described in Section 4.2.

In Lemma C.7, we show that $h_{ti}(p)$ is differentiable and strictly decreasing with derivatives $h'_{ti}(p) \leq -L_0$. Therefore, $h_{ti}(p)$ satisfies the conditions of Proposition 3.2 and we can use Algorithm 4.2 to find an optimal assortment and pricing.

The main difficulty in running Algorithm 4.2 is to find the set of points $\mathcal{P}_{ti}(B) = \{p : p + 1/h'_{ti}(p) = B\}$ for any given $B > 0$. In the following, we will show how we can obtain this set for the specific structure of $h_{ti}(p)$.

In the proof of Lemma C.7, we establish that $\tilde{h}_{ti}(p)$ is smooth and strictly convex. Then, letting p_0 be the unique value such that $\tilde{h}'_{ti}(p_0) = -L_0$, we show that

$$h_{ti}(p) = \begin{cases} \tilde{h}_{ti}(p_0) - L_0(p - p_0) & \text{if } p \geq p_0, \\ \tilde{h}_{ti}(p) & \text{if } p < p_0. \end{cases}$$

To find all points that satisfy $p + 1/h'_{ti}(p) = B$, we search for p values to the left and right of p_0 separately.

To find all the points $p \in \mathcal{P}_{ti}(B)$ such that $p \leq p_0$, we will use the structure of $\tilde{h}_{ti}(p)$. Since $\tilde{h}_{ti}(p)$ is given as a sum of a linear function and a square root of a quadratic function, we can write it as

$$\tilde{h}_{ti}(p) = a_1 - a_2 \cdot p + \sqrt{a_3 - 2a_4 \cdot p + a_5 \cdot p^2}$$

for some $a_1, a_2, a_3, a_4, a_5 \in \mathbb{R}$. Since the square root part is a norm, the quadratic inside must have non-positive determinant, i.e. $a_4^2 - a_3 a_5 \leq 0$.

Therefore, for any $p \leq p_0$, we have

$$h'_{ti}(p) = \frac{a_5 \cdot p - a_4}{\sqrt{a_3 - 2a_4 \cdot p + a_5 \cdot p^2}} - a_2$$

$$h''_{ti}(p) = \frac{a_3 a_5 - a_4^2}{(a_3 - 2a_4 \cdot p + a_5 \cdot p^2)^{3/2}}$$

Our goal is to find the solutions for $p + \frac{1}{h'_{ti}(p)} - B = 0$, or equivalently we want to find the roots of $z(p) := h'_{ti}(p) - \frac{1}{B-p}$. Since $h'_{ti}(p) < 0$, this equation only has root on $p > B$. Since $z(p)$ is continuous on $p > B$, there exists at most one root between each local minima/maxima points of $z(p)$. Furthermore, since $z(p)$ is also differentiable function on $p > B$, we can find all points with $z'(p) = 0$ to identify local minima and maxima. Now, we observe that

$$z'(p) = h''_{ti}(p) - \frac{1}{(B-p)^2}.$$

Hence, at any point satisfying $z'(p) = 0$, we must have

$$\frac{a_3 a_5 - a_4^2}{(a_3 - 2a_4 \cdot p + a_5 \cdot p^2)^{3/2}} = \frac{1}{(B-p)^2}$$

Then, raising this equation two the second power, we obtain

$$\frac{1}{(a_3 a_5 - a_4^2)^2} (a_3 - 2a_4 \cdot p + a_5 \cdot p^2)^3 - (B-p)^4 = 0.$$

Since the left hand side is a 6th order polynomial in p , we can easily find all the roots for this equation. Since any local minima/maxima points of $z(p)$ must be one of these roots, this gives us a necessary condition for local minima/maxima points of $z(p)$. Then, check for each of these points and construct the set \mathcal{Z} that contains local minima/maxima points of $z(p)$.

Then, we can search for a root of $z(p)$ between each pair of consecutive points in \mathcal{Z} . Since there is at most one root between every pair of consecutive points, we can find all roots of $z(p)$ efficiently. After finding all the roots, we only add the ones that satisfy $p < p_0$ to the set $\mathcal{P}_{ti}(B)$.

Lastly, we check for any solutions $p + 1/h'_{ti}(p) - B = 0$ over $p > p_0$. Since $h_{ti}(p)$ is a linear function over $p > p_0$, the only possible solution is $p = 1/L_0 + B$. If this solution satisfies $p > p_0$, we also add it to the set $\mathcal{P}_{ti}(B)$.

B.2 Technical Lemmas for Assortment Selection and Pricing

Lemma B.5. *Under Assumption B.1, $f_i(\theta) > 0$ is a continuous and strictly decreasing function of $\theta \in \mathbb{R}$.*

Proof. Recall the definitions

$$g_i(p_i) = -\frac{e^{u_i(p_i)}}{u'_i(p_i)}$$

$$f_i(\theta) = \max \{g_i(p_i) : p_i \in \mathcal{P}_i(\theta)\}$$

where $\mathcal{P}_i(\theta) = \{p : \theta = p + 1/u'_i(p)\}$. Since $u_i(p)$ is a differentiable function, its derivative $u'_i(p)$ is continuous everywhere. We also have $u'_i(p) < L_0$ since $u_i(p)$ is decreasing.

Let $z(p) = p + 1/u'_i(p)$. By continuity of $u'_i(p)$, both $z(p)$ and $g_i(p)$ are continuous functions.

First, we show that any p that is a local minimum for $z(p)$ is a local maximum for $g_i(p)$. Similarly, any p that is a local maximum for $z(p)$ is a local minimum for $g_i(p)$.

Suppose p is a local minimum for $z(p)$. Then, there exists some $\delta > 0$ such that $z(p') \geq z(p)$ for all $|p - p'| \leq \delta$. That is, $p' - p + \frac{1}{u'_i(p')} - \frac{1}{u'_i(p)} \geq 0$.

Now, we use Taylor's expansion for $e^{u_i(p)}$ at p to write

$$e^{u_i(p')} = e^{u_i(p)} + u'_i(p)e^{u_i(p)}(p' - p) + o(\delta).$$

Then, dividing through by $e^{u_i(p)}u'_i(p')$, we obtain

$$\frac{e^{u_i(p')} - u_i(p)}{u'_i(p')} = \frac{1}{u'_i(p')} + \frac{u'_i(p)}{u'_i(p')}(p' - p) + e^{-u_i(p)}o(\delta).$$

Since $u'_i(p) < 0$ for all p , we have $\frac{u'_i(p)}{u'_i(p')} > 0$. Then, using $p' - p + \frac{1}{u_i(p')} - \frac{1}{u_i(p)} \geq 0$ for any $|p - p'| \leq \delta$, we can write

$$\begin{aligned} \frac{e^{u_i(p')} - u_i(p)}{u'_i(p')} &= \frac{1}{u'_i(p')} + \frac{u'_i(p)}{u'_i(p')}\left(\frac{1}{u'_i(p)} - \frac{1}{u'_i(p')}\right) + e^{-u_i(p)}o(\delta) \\ &= \frac{2}{u'_i(p')} - \frac{u'_i(p)}{(u'_i(p'))^2} + e^{-u_i(p)}o(\delta). \end{aligned}$$

It is possible to show that $f(x) = \frac{2}{x} - \frac{a}{x^2}$ has a local minimum at $x = a$ when $a < 0$. Therefore, using continuity of $u'_i(p)$, we can show that there exists $\eta > 0$ such that

$$\frac{e^{u_i(p')} - u_i(p)}{u'_i(p')} \geq \frac{1}{u'_i(p')}$$

This inequality is equivalent to $g_i(p') \leq g_i(p)$. Therefore, $g_i(p)$ has a local maximum at p . We can show the symmetric result using similar arguments.

The next step is to show the continuity of $f_i(p)$. Consider any θ and any $p \in \mathcal{P}_i(\theta)$ that is not a local maximum or minimum for $z(p)$. Then, for any $\eta > 0$, there exists a real $\delta > 0$ such that for any θ' , $0 < |\theta - \theta'| < \delta$ implies that there exists a $p' \in \mathcal{P}_i(\theta')$ such that $|p - p'| \leq \eta$. In other words, $\mathcal{P}_i(\theta)$ is a *continuous* function (that maps a real number to a set of real numbers) unless one of the prices in $\mathcal{P}_i(\theta)$ is a local maximum or minimum for $z(p)$.

As a result, $f_i(\theta)$ is a continuous function over θ values for which no price in $\mathcal{P}_i(\theta)$ is local maxima or minima for $z(p)$. Next, we show continuity on other θ values. Consider a θ value and a price point $p \in \mathcal{P}_i(\theta)$ that is a local maximum or minimum for $z(p)$.

If $p \in \mathcal{P}_i(\theta)$ is a local minimum for $z(p)$, there exists another $p_1 \in \mathcal{P}_i(\theta)$ such that $p_1 < p$ and $z(p) \geq \theta$ for all $p' \in (p_1, p)$. This is because $z(p)$ is continuous $\lim_{p \rightarrow -\infty} z(p) = -\infty$. Then, by Lemma B.6, we have $g_i(p) \leq g_i(p_1)$. Since $g_i(p) \leq g_i(p_1)$, θ is not *active* in f_i . As a result, the continuity of $f_i(\theta)$ is preserved at θ .

Similarly, if $p \in \mathcal{P}_i(\theta)$ is a local maximum for $z(p)$, there exists another $p_2 \in \mathcal{P}_i(\theta)$ such that $p < p_2$ and $z(p) \leq \theta$ for all $p' \in (p, p_2)$. This is because $z(p)$ is continuous $\lim_{p \rightarrow \infty} z(p) = \infty$. Then, by Lemma B.6, we have $g_i(p_2) \leq g_i(p)$. As a result, the continuity of $f_i(\theta)$ is preserved at θ .

We have shown that $f_i(\theta)$ is continuous. Next, we show that it is a decreasing function. As we showed in previous parts of this proof, the continuity of $f_i(\theta)$ is not affected at θ values with some $p \in \mathcal{P}_i(\theta)$ that is a local maximum or minimum for $z(p)$. Therefore, it is sufficient to show that $f_i(\theta)$ is decreasing over θ values such that no $p \in \mathcal{P}_i(\theta)$ is a local maximum or minimum for $z(p)$.

Let θ be such a value. Since $g_i(p)$ is strictly increasing on every interval in which $z(p)$ is strictly decreasing and $g_i(p)$ is strictly decreasing on every interval in which $z(p)$ is strictly increasing, there exists a real $\delta > 0$ such that for any $p \in \mathcal{P}_i(\theta)$, there exists a $p' \in \mathcal{P}_i(\theta')$ satisfying $g_i(p) > g_i(p')$ whenever $\theta < \theta' < \theta + \delta$. Therefore, there exists a real $\delta > 0$ such that $f_i(\theta) > f_i(\theta')$ whenever $\theta < \theta' < \theta + \delta$.

Since the function $f_i(\theta)$ is continuous and it is locally strictly decreasing almost everywhere in \mathbb{R} , it must be strictly decreasing.

□

Lemma B.6. Let $p_1 < p_2$ be two price points such that $p_1, p_2 \in \mathcal{P}_i(\theta)$ for some θ . If $p + 1/u'_i(p) \leq \theta$ for all $p \in (p_1, p_2)$, then $g_i(p_2) \geq g_i(p_1)$. If $p + 1/u'_i(p) \geq \theta$ for all $p \in (p_1, p_2)$, then $g_i(p_2) \leq g_i(p_1)$.

Proof. Using $p_1 + 1/u'_i(p_1) = p_2 + 1/u'_i(p_2) = \theta$, we have

$$\begin{aligned} g_i(p_1) &= -e^{u_i(p_1)}/u'_i(p_1) = e^{u_i(p_1)}(p_1 - \theta) \\ g_i(p_2) &= -e^{u_i(p_2)}/u'_i(p_2) = e^{u_i(p_2)}(p_2 - \theta). \end{aligned}$$

We let $w(p) = e^{u_i(p)}(p - \theta)$ and notice that $w(p_1) = g_i(p_1)$ and $w(p_2) = g_i(p_2)$. Now, we compute the derivative of $w(p)$ as

$$w'(p) = e^{u_i(p)}(1 + (p - \theta)u'_i(p)).$$

Since $u'_i(p) < 0$ for all p , we have $w'(p) \geq 0$ if and only if $p + 1/u'_i(p) \leq \theta$. Hence, if $p + 1/u'_i(p) \leq \theta$ for all $p \in (p_1, p_2)$, then $w'(p) \geq 0$ for all $p \in (p_1, p_2)$. Since $w(p)$ is continuous and differentiable, we conclude $g_i(p_2) \geq g_i(p_1)$. The symmetric result also follows similarly. \square

C Proof of Theorem 4.2 (Regret Upper Bound for Algorithm 2)

In the following section, we present our proof for Theorem 4.2. For better readability, we first present the overall proof using a series of technical lemmas. We provide the proofs for these technical lemmas later in Appendix C.1.

We start by recalling Proposition 3.2 which defines B_t as the unique solution of

$$B = \max_{S \in \mathcal{S}_K} \sum_{i \in S} v_{ti}(B) \tag{12}$$

where $v_{ti}(B) = \max_{p \in \mathbb{R}} \{-e^{h_{ti}(p)}/h'_{ti}(p) : p + 1/h'_{ti}(p) = B\}$. This proposition also asserts that the optimum prices are $p_{ti}^* = \operatorname{argmax}_{p \in \mathbb{R}} \{-e^{h_{ti}(p)}/h'_{ti}(p) : p + 1/h'_{ti}(p) = B_t\}$. Our first lemma shows that this fixed point B_t lies within $[0, P_0]$ for some P_0 under our assumptions, allowing us to constrain our search for the fixed point into a bounded interval. This result also implies that the optimum prices p_{ti}^* are bounded within $[0, P]$ for some P .

Lemma C.1 (Bounded optimum prices). *Consider that the utility function for each item $i \in [N]$ is given by a differentiable function $h_{ti}(p)$ such that $h_{ti}(0) \leq 1 + \mu$ and its first order derivative satisfies $h'_{ti}(p) \leq -L_0$ for all $p \in \mathbb{R}$. Then, the fixed point satisfies $B_t \in [0, P_0(\mu)]$ and the optimum prices satisfy $p_{ti}^* \in [0, P(\mu)]$ for constants*

$$P_0(\mu) = \frac{e^\mu \cdot (0.6 + \log K)}{L_0} \quad \text{and} \quad P(\mu) = P_0(\mu) + \frac{1}{L_0}.$$

For ease of notation, define $P_0 := P_0(1)$, $P := P(1)$, and $\bar{P} := 1 + P$.

Based on Assumption 4.1, the true utility functions satisfy $u_{ti}(0) \leq 1$ and $u'_{ti}(p) \leq -L_0$. Furthermore, as we show in the following proof, the estimated utility functions satisfy $h_{ti}(0) \leq 2$ and $h'_{ti}(p) \leq -L_0$. Therefore, both the true optimum prices under $u_{ti}(p)$ and the estimated optimum prices calculated under $h_{ti}(p)$ are bounded by P .

Recall that T_0 is the length of random initialization. At each round $t < T_0$, the algorithm chooses a subset S_t uniformly at random from $\{S \subseteq [N] : |S| \leq K\}$ and sets $p_{ti} \in [1, 2]$ uniformly at random for all $i \in S_t$. Then, we use the assumption that there exists a constant $\sigma_0 > 0$ such that $\mathbb{E}[\mathbf{x}_{ti}\mathbf{x}_{ti}^\top] \succcurlyeq \sigma_0 \mathbf{I}$ and show how many rounds of initialization are required to achieve a target minimum eigenvalue.

Lemma C.2 (Initialization). *Define our target minimum eigenvalue for $\mathbf{V}_{T_0} = \frac{1}{K^2} \sum_{t=1}^{T_0-1} \sum_{i \in S_t} \tilde{\mathbf{x}}_{ti}\tilde{\mathbf{x}}_{ti}^\top$ as*

$$\lambda_{\min}^0 = C_1 \frac{d\bar{P}^2 \log^3(T)}{\sigma_0^2}$$

for some universal constant C_1 . Then, there exist some positive, universal constant C_2 such that if the length of random initialization satisfies

$$T_0 \geq C_2 \frac{\lambda_{\min}^0 K}{\sigma_0},$$

then $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0$ with probability at least $1 - \frac{1}{T}$.

This condition is central in showing that the maximum likelihood estimator is consistent (Lemma C.5) and satisfies a finite-sample normality-type estimation error bound (Lemma C.6). Similar to Li et al. (2017) and Oh and Iyengar (2021), the independence assumption (Assumption 4.1) on the feature vectors \mathbf{x}_{ti} is only needed to ensure that $\hat{\boldsymbol{\theta}}_t$ sufficiently close to $\boldsymbol{\theta}^*$ at the end of the initialization phase. We do not require this stochasticity assumption in the rest of the regret analysis. Therefore, after the random initialization period of the first T_0 rounds, the context vectors \mathbf{x}_{ti} can even be chosen adversarially as long as their norms $\|\mathbf{x}_{ti}\|$ are bounded and they satisfy the minimum price sensitivity condition $\langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle \geq L_0$.

Next, we show that the probability of selection for any item $i \in S_t$ for the assortments S_t and prices \mathbf{p}_t offered by Algorithm 2 can be estimated well enough using $\hat{\boldsymbol{\theta}}_t$ sufficiently close to $\boldsymbol{\theta}^*$. Namely, we let $\gamma = \log 2/(8\bar{P}) < 1$ and define

$$\mathcal{B}_\gamma := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq \gamma\}.$$

Then, as we show in Lemma C.12, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{B}_\gamma$, we have

$$\frac{1}{\sqrt{2}} \leq \frac{q_{ti}(\boldsymbol{\theta}_1)}{q_{ti}(\boldsymbol{\theta}_2)} \leq \sqrt{2}.$$

As a result of this relation, we obtain the following estimation results. These results show that \mathbf{V}_t can estimate the Fisher Information Matrix $\mathcal{H}_t(\boldsymbol{\theta})$ within a constant factor in a small enough neighborhood around $\boldsymbol{\theta}^*$.

Lemma C.3 (Regularity of Fisher Information). *Let $\mathbf{H}_t : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d \times 2d}$ denote the function defined as*

$$\mathbf{H}_t(\boldsymbol{\theta}) = \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top.$$

Then, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{B}_\gamma$, and any $t \geq T_0$, we have

$$\frac{1}{4} \mathbf{H}_t(\boldsymbol{\theta}_1) \preceq \mathbf{H}_t(\boldsymbol{\theta}_2) \preceq 4 \mathbf{H}_t(\boldsymbol{\theta}_1).$$

Lemma C.4 (Fisher Information Estimation). *Let $t \geq T_0$ and assume that $\hat{\boldsymbol{\theta}}_\tau \in \mathcal{B}_\gamma$ for all $T_0 \leq \tau < t$. Then, for any $\boldsymbol{\theta} \in \mathcal{B}_\gamma$, we have*

$$\mathcal{H}_t(\boldsymbol{\theta}) := \sum_{\tau=1}^{t-1} \mathbf{H}_\tau(\boldsymbol{\theta}) \succcurlyeq C_3 \mathbf{V}_t$$

for some universal constant $C_3 > 0$.

The next result shows that our MLE estimates can reach and stay within the γ -neighborhood of the true parameter $\boldsymbol{\theta}^*$ with high probability as long as the initialization is successful.

Lemma C.5 (Consistency of MLE). *Let T_0 be any round such that $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0$. Then, we have*

$$\mathbb{P}(\exists t \geq T_0, \hat{\boldsymbol{\theta}}_t \notin \mathcal{B}_\gamma) \leq \frac{1}{T}.$$

Combining the results of Lemma C.2 and Lemma C.5, we can show that the conditions $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0$ and $\hat{\boldsymbol{\theta}}_t \in \mathcal{B}_\gamma$ for all $t \geq T_0$ are satisfied with probability $1 - \mathcal{O}(T^{-1})$ if we select

$$T_0 = \Theta\left(\frac{\lambda_{\min}^0 K}{\sigma_0}\right) = \Theta\left(\frac{d\bar{P}^2 K \log^3(T)}{\sigma_0^3}\right). \quad (13)$$

Thus, we can define a *good* event

$$\mathcal{E}_0 = \{\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0\} \cap \{\hat{\boldsymbol{\theta}}_t \in \mathcal{B}_\gamma, \forall t \geq T_0\}$$

that holds with probability $1 - \mathcal{O}(T^{-1})$.

Now, given that the initialization successfully identifies a point γ -neighborhood of the true parameter $\boldsymbol{\theta}^*$, the next step is to construct tight confidence regions that contain the true parameter with high probability. The next Lemma establishes that we can construct a confidence region using the estimated Fisher information matrix \mathbf{V}_t .

Lemma C.6 (Normality of MLE). *Suppose the event \mathcal{E}_0 holds. Then, for any $t \geq T_0$,*

$$\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq C_5 \left(\sqrt{d \log \left(1 + \frac{2t^3}{d} \right)} + \frac{1}{\sigma_0} \log(T) \right) \quad (14)$$

with probability at least $1 - \mathcal{O}(t^{-2})$.

For the selection of T_0 given in (28), we already showed that \mathcal{E}_0 holds with probability $1 - \mathcal{O}(T^{-1})$. Therefore, conditioned on \mathcal{E}_0 happens, we can further ensure that $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq \alpha_t$ holds with probability at least $1 - t^{-2}$ if we choose the confidence radius as

$$\alpha_t = C_5 \left(\sqrt{d \log \left(1 + \frac{2t^3}{d} \right)} + \frac{1}{\sigma_0} \log(T) \right). \quad (15)$$

Consequently, for each $t \geq T_0$, we can define another *good* event $\mathcal{E}_t = \{\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq \alpha_t\}$ that holds with probability at least $1 - \mathcal{O}(t^{-2})$ conditioned on \mathcal{E}_0 .

Now, we are ready to construct our optimistic utility functions using the confidence regions established in Lemma C.6. The following lemma establishes important properties for the optimistic utility functions constructed in Algorithm 2.

Lemma C.7. *Suppose \mathcal{E}_t holds for all $t \geq T_0$. Let $\tilde{h}_{ti} : \mathbb{R} \rightarrow \mathbb{R}$ be the function defined as*

$$\tilde{h}_{ti}(p) := \langle \hat{\boldsymbol{\psi}}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\boldsymbol{\phi}}_t, \mathbf{x}_{ti} \rangle \cdot p + g_{ti}(p),$$

where $\hat{\boldsymbol{\theta}}_t = (\hat{\boldsymbol{\psi}}_t, \hat{\boldsymbol{\phi}}_t)$ and $g_{ti}(p) := \alpha_t \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}}$. Furthermore, let $h_{ti} : \mathbb{R} \rightarrow \mathbb{R}$ be the function defined as

$$h_{ti}(p) := \min_{p' \leq p} \left\{ \tilde{h}_{ti}(p') - L_0(p - p') \right\}.$$

Then, the function $h_{ti}(p)$ is differentiable and satisfies

$$h_{ti}(p) \geq u_{ti}(p), \quad (16)$$

$$h_{ti}(p) - u_{ti}(p) \leq 2g_{ti}(p), \quad (17)$$

$$h'_{ti}(p) \leq -L_0, \quad (18)$$

$$h_{ti}(0) \leq 2 \quad (19)$$

for all $p \in \mathbb{R}$.

We recall that Algorithm 2 chooses the assortment S_t and prices \mathbf{p}_t by solving

$$(S_t, \mathbf{p}_t) \in \underset{\substack{S \in \mathcal{S}_K \\ \mathbf{p} \in \mathbb{R}_+^n}}{\operatorname{argmax}} \tilde{R}_t(S, \mathbf{p})$$

where $\tilde{R}_t(S, \mathbf{p})$ denotes the optimistic estimate of the revenue function as defined in (5). Then, using the properties of the optimistic estimate of the utility functions $h_{ti}(p)$, we can show the following lemma.

Lemma C.8. *Assume good event \mathcal{E}_t holds for some $t \geq T_0$. Then,*

- (a) $R_t(S_t^*, \mathbf{p}_t^*) \leq \tilde{R}_t(S_t, \mathbf{p}_t)$, and
- (b) $\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t) \leq 4P \cdot \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) g_{ti}(p_{ti})$.

Now, we break the regret \mathcal{R}_T into the initialization phase and the learning phase:

$$\begin{aligned}\mathcal{R}_T &= \mathbb{E} \left[\sum_{t=1}^{T_0-1} (R_t(S_t^*, \mathbf{p}_t^*) - R_t(S_t, \mathbf{p}_t)) \right] + \mathbb{E} \left[\sum_{t=T_0}^T (R_t(S_t^*, \mathbf{p}_t^*) - R_t(S_t, \mathbf{p}_t)) \right] \\ &\leq PT_0 + \mathbb{E} \left[\sum_{t=T_0}^T (R_t(S_t^*, \mathbf{p}_t^*) - R_t(S_t, \mathbf{p}_t)) \right] \\ &\leq PT_0 + \mathbb{E} \left[\sum_{t=T_0}^T (\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \right]\end{aligned}$$

where the first inequality follows from $R_t(S_t^*, \mathbf{p}_t^*) \leq P$ and the second inequality follows from property (a) in Lemma C.8. Now, we decompose the remaining expectation term into two parts where the high probability event \mathcal{E}_0 holds and it does not.

$$\begin{aligned}\mathcal{R}_T &\leq PT_0 + \mathbb{E} \left[\sum_{t=T_0}^T (\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \mathbb{1}(\mathcal{E}_0) \right] + \mathbb{E} \left[\sum_{t=T_0}^T (\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \mathbb{1}(\neg\mathcal{E}_0) \right] \\ &\leq PT_0 + \sum_{t=T_0}^T \mathbb{E} \left[(\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \mathbb{1}(\mathcal{E}_0) \right] + \mathcal{O}(P).\end{aligned}$$

where the last inequality uses $\tilde{R}_t(S_t, \mathbf{p}_t) \leq P$ and $\Pr(\neg\mathcal{E}_0) \leq \mathcal{O}(T^{-1})$. For each expectation term in the remaining summation, we can split it into two parts where the high probability event \mathcal{E}_t holds and it does not:

$$\begin{aligned}&\mathbb{E} \left[(\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \mathbb{1}(\mathcal{E}_0) \right] \\ &= \mathbb{E} \left[(\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \mathbb{1}(\mathcal{E}_0) \mathbb{1}(\mathcal{E}_t) \right] + \mathbb{E} \left[(\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t)) \mathbb{1}(\mathcal{E}_0) \mathbb{1}(\neg\mathcal{E}_t) \right] \\ &\leq 4P\alpha_t \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}} + \mathcal{O}(P \cdot t^{-2}).\end{aligned}$$

where the last inequality follows from property (b) in Lemma C.8 as well as $\Pr(\neg\mathcal{E}_t) \leq \mathcal{O}(t^{-2})$ and $\tilde{R}_t(S_t, \mathbf{p}_t) \leq P$. As a result,

$$\begin{aligned}\mathcal{R}_T &\leq PT_0 + 4P \sum_{t=T_0}^T \alpha_t \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}} + \sum_{t=1}^T \mathcal{O}(P \cdot t^{-2}) + \mathcal{O}(P) \\ &\leq PT_0 + 4P \sum_{t=T_0}^T \alpha_t \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}} + \mathcal{O}(P)\end{aligned}$$

Applying Cauchy-Schwarz and Jensen's inequalities in the second term, it follows that

$$\begin{aligned}\mathcal{R}_T &\leq PT_0 + 4P\alpha_T \sqrt{T \sum_{t=T_0}^T \left(\sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}} \right)^2} + \mathcal{O}(P) \\ &\leq PT_0 + 4P\alpha_T \sqrt{T \sum_{t=T_0}^T \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2} + \mathcal{O}(P).\end{aligned}$$

Applying Lemma C.11 and Lemma C.12, we obtain

$$\mathcal{R}_T \leq PT_0 + 29P\alpha_T \sqrt{dKT \log(T/d)} + \mathcal{O}(P).$$

C.1 Proofs for Technical Lemmas

Lemma C.1 (Bounded optimum prices). *Consider that the utility function for each item $i \in [N]$ is given by a differentiable function $h_{ti}(p)$ such that $h_{ti}(0) \leq 1 + \mu$ and its first order derivative satisfies $h'_{ti}(p) \leq -L_0$ for all*

$p \in \mathbb{R}$. Then, the fixed point satisfies $B_t \in [0, P_0(\mu)]$ and the optimum prices satisfy $p_{ti}^* \in [0, P(\mu)]$ for constants

$$P_0(\mu) = \frac{e^\mu \cdot (0.6 + \log K)}{L_0} \quad \text{and} \quad P(\mu) = P_0(\mu) + \frac{1}{L_0}.$$

For ease of notation, define $P_0 := P_0(1)$, $P := P(1)$, and $\bar{P} := 1 + P$.

Proof. Given the conditions on $h_{ti}(p)$, we have $h_{ti}(p) \leq 1 + \mu - L_0 p$ for all $p \geq 0$. Therefore, for any $B \geq 0$, we have

$$\begin{aligned} v_{ti}(B) &= \max_{p \in \mathbb{R}} \{-\exp\{h_{ti}(p)\}/h'_{ti}(p) : p + 1/h'_{ti}(p) = B\} \\ &\leq \max_{p \in \mathbb{R}} \left\{ \frac{\exp\{h_{ti}(B - 1/h'_{ti}(p))\}}{L_0} \right\} \\ &\leq \frac{\exp\{h_{ti}(B + 1/L_0)\}}{L_0} \\ &\leq \frac{1}{L_0} e^{\mu - L_0 B}. \end{aligned}$$

The first equality is the definition of $v_{ti}(B)$ and the second inequality uses the condition $h'_{ti}(p) \leq -L_0$ for all p , and the last inequality uses the result $h_{ti}(p) \leq 1 + \mu - L_0 p$ for all $p \geq 0$.

As a result, for any $S \in \mathcal{S}_K$, we have $\sum_{i \in S} v_{ti}(B) \leq \frac{K}{L_0} e^{\mu - L_0 B}$ for all $B \geq 0$.

Now, we let B_u be the unique solution of the fixed point equation

$$B = \frac{K}{L_0} e^{\mu - L_0 B}. \tag{20}$$

Since the right-hand sides of (20) and (3) are both positive for all $B \in \mathbb{R}$, the fixed points B_u and B_t are both positive. Furthermore, since the right-hand side of (20) is an upper bound for the right-hand side of (3) for all $B \geq 0$, we must have $B_t \leq B_u$.

In (20), the left-hand side is increasing and the right-hand side is decreasing in B . Additionally, for $B = e^\mu(0.6 + \log(K))/L_0$, the left-hand side of (20) is greater than the right-hand side. Hence, the fixed point satisfies $0 \leq B_t \leq B_u \leq P_0 = e^\mu(0.6 + \log(K))/L_0$.

Furthermore, the optimum prices satisfy $p_{ti}^* + 1/h'_{ti}(p_{ti}^*) = B_t$. Hence, $0 \leq p_{ti}^* + 1/h'_{ti}(p_{ti}^*) \leq B_u$. Using that $h'_{ti}(p) \leq -L_0$, we have $0 \leq p_{ti}^* \leq P_0 + 1/L_0$.

□

Lemma C.2 (Initialization). Define our target minimum eigenvalue for $\mathbf{V}_{T_0} = \frac{1}{K^2} \sum_{t=1}^{T_0-1} \sum_{i \in S_t} \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top$ as

$$\lambda_{\min}^0 = C_1 \frac{d\bar{P}^2 \log^3(T)}{\sigma_0^2}$$

for some universal constant C_1 . Then, there exist some positive, universal constant C_2 such that if the length of random initialization satisfies

$$T_0 \geq C_2 \frac{\lambda_{\min}^0 K}{\sigma_0},$$

then $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0$ with probability at least $1 - \frac{1}{T}$.

Proof. Let $\Sigma = \mathbb{E}[\mathbf{x}_{ti} \mathbf{x}_{ti}^\top]$ and $\tilde{\Sigma} = \mathbb{E}[\tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top]$. Then, noting that p_{ti} is uniformly and independently distributed over $[1, 2]$ for all $t \leq T_0$, we can write

$$\tilde{\Sigma} = \begin{bmatrix} \Sigma & -\frac{3}{2}\Sigma \\ -\frac{3}{2}\Sigma & \frac{7}{3}\Sigma \end{bmatrix}.$$

Then, using Schur's formula, each eigenvalue $\tilde{\lambda}$ of $\tilde{\Sigma}$ are given by solutions of the equation

$$\begin{aligned} 0 &= \det(\tilde{\Sigma} - \tilde{\lambda}\mathbf{I}) \\ &= \det(\Sigma - \tilde{\lambda}\mathbf{I}) \det\left(\frac{7}{3}\Sigma - \tilde{\lambda}\mathbf{I} - \frac{9}{4}\Sigma(\Sigma - \tilde{\lambda}\mathbf{I})^{-1}\Sigma\right). \end{aligned}$$

Since the inverse of the matrix $\Sigma - \tilde{\lambda}\mathbf{I}$ appears on the right-hand side, we must have $\det(\Sigma - \tilde{\lambda}\mathbf{I}) \neq 0$. Hence, all eigenvalues must satisfy

$$\det\left(\frac{7}{3}\Sigma - \tilde{\lambda}\mathbf{I} - \frac{9}{4}\Sigma(\Sigma - \tilde{\lambda}\mathbf{I})^{-1}\Sigma\right) = 0.$$

Letting $\Sigma = \mathbf{V}\Lambda\mathbf{V}^\top$ be the eigen-decomposition of Σ with $\{\lambda_j\}_{j=1}^d$ denoting the eigenvalues. Then, we can write

$$\begin{aligned} 0 &= \det\left(\frac{7}{3}\mathbf{V}\Lambda\mathbf{V}^\top - \tilde{\lambda}\mathbf{I} - \frac{9}{4}\mathbf{V}\Lambda\mathbf{V}^\top(\mathbf{V}\Lambda\mathbf{V}^\top - \tilde{\lambda}\mathbf{I})^{-1}\mathbf{V}\Lambda\mathbf{V}^\top\right) \\ &= \det(\mathbf{V})^2 \det\left(\frac{7}{3}\Lambda - \tilde{\lambda}\mathbf{I} - \frac{9}{4}\Lambda(\Lambda - \tilde{\lambda}\mathbf{I})^{-1}\Lambda\right) \\ &= \prod_{j=1}^d \left(\frac{7}{3}\lambda_j - \tilde{\lambda} - \frac{9}{4}\frac{\lambda_j^2}{\lambda_j - \tilde{\lambda}}\right). \end{aligned}$$

Consequently, the eigenvalues of $\tilde{\Sigma}$ are given by

$$\tilde{\lambda}_{j,1} = (20 + 2\sqrt{97})\lambda_j \text{ and } \tilde{\lambda}_{j,2} = (20 - 2\sqrt{97})\lambda_j, \forall j \in [d].$$

Since $\lambda_j \geq \sigma_0$ for all j by Assumption 4.1, $\lambda_{\min}(\tilde{\Sigma}) \geq C\sigma_0$ for some positive, universal constant C . Then, using Proposition 1 from Li et al. (2017), we establish that there exist some positive, universal constants C_1 and C_2 such that if the length of random initialization satisfies

$$T_0 \geq \left(\frac{C_2\sqrt{d} + C_3\sqrt{\log T}}{\sigma_0}\right)^2 + \frac{2BK}{\sigma_0},$$

then $\lambda_{\min}(\mathbf{V}_{T_0}) \geq B$ with probability at least $1 - \frac{1}{T}$. Lastly, we set $B = \lambda_{\min}^0$ and observe that

$$\lambda_{\min}^0 = C_1 \left(\frac{\bar{P}d\log(T)}{\sigma_0}\right)^2 \geq C_4 \left(\frac{C_1\sqrt{d} + C_2\sqrt{\log T}}{\sigma_0}\right)^2$$

for some universal constant $C_4 > 0$ since $\bar{P} \geq 1$.

□

Lemma C.3 (Regularity of Fisher Information). *Let $\mathbf{H}_t : \mathbb{R}^{2d} \rightarrow \mathbb{R}^{2d \times 2d}$ denote the function defined as*

$$\mathbf{H}_t(\boldsymbol{\theta}) = \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top.$$

Then, for any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{B}_\gamma$, and any $t \geq T_0$, we have

$$\frac{1}{4}\mathbf{H}_t(\boldsymbol{\theta}_1) \preccurlyeq \mathbf{H}_t(\boldsymbol{\theta}_2) \preccurlyeq 4\mathbf{H}_t(\boldsymbol{\theta}_1).$$

Proof. We start with some definitions that will be useful in the following proof.

First, we let $\tilde{\mathbf{x}}_{t0} = \mathbf{0}_{2d}$ and $\bar{S}_t = S_t \cup \{0\}$ denote the extended assortment that includes the null item. Then, we can write

$$\mathbf{H}_t(\boldsymbol{\theta}) = \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in \bar{S}_t} \sum_{j \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top.$$

Then, we let $\mathbf{H}_1 = \mathbf{H}_t(\boldsymbol{\theta}_1)$ and $\mathbf{H}_2 = \mathbf{H}_t(\boldsymbol{\theta}_2)$ denote the matrices of interest.

Define the differences in the probabilities as $\delta_{ti} = q_{ti}(\boldsymbol{\theta}_1) - q_{ti}(\boldsymbol{\theta}_2)$.

Define the expected value of the context selection as $\bar{\mathbf{x}}_1 = \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) \tilde{\mathbf{x}}_{ti}$ and $\bar{\mathbf{x}}_2 = \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \tilde{\mathbf{x}}_{ti}$.

Lastly, we define the mean-centered contexts as $\mathbf{w}_{ti} = \tilde{\mathbf{x}}_{ti} - \bar{\mathbf{x}}_1$ and $\tilde{\mathbf{w}}_{ti} = \tilde{\mathbf{x}}_{ti} - \bar{\mathbf{x}}_2$.

Using these definitions and noting that $\sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) \mathbf{w}_{ti} = \mathbf{0}$ and $\sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \tilde{\mathbf{w}}_{ti} = \mathbf{0}$, we can write

$$\begin{aligned}\mathbf{H}_1 &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) \mathbf{w}_{ti} \mathbf{w}_{ti}^\top \\ \mathbf{H}_2 &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \tilde{\mathbf{w}}_{ti} \tilde{\mathbf{w}}_{ti}^\top.\end{aligned}$$

Our initial goal is to show $\frac{1}{2} \tilde{\mathbf{H}} \preccurlyeq \mathbf{H}_2 \preccurlyeq 2\tilde{\mathbf{H}}$ for an intermediate matrix defined as

$$\tilde{\mathbf{H}} = \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) \tilde{\mathbf{w}}_{ti} \tilde{\mathbf{w}}_{ti}^\top.$$

To achieve this goal, it is sufficient to show that $-\frac{1}{2} \tilde{\mathbf{H}} \preccurlyeq \tilde{\mathbf{H}} - \mathbf{H}_2 \preccurlyeq \frac{1}{2} \tilde{\mathbf{H}}$. We notice that this difference can be written as

$$\tilde{\mathbf{H}} - \mathbf{H}_2 = \sum_{i \in \bar{S}_t} \delta_{ti} \tilde{\mathbf{w}}_{ti} \tilde{\mathbf{w}}_{ti}^\top.$$

To prove that this inequality holds, it is sufficient to show that $|\delta_{ti}| \leq \frac{1}{2} q_{ti}(\boldsymbol{\theta}_1)$. Using Lemma C.12, we can show that

$$\frac{\delta_{ti}}{q_{ti}(\boldsymbol{\theta}_1)} = \frac{q_{ti}(\boldsymbol{\theta}_1) - q_{ti}(\boldsymbol{\theta}_2)}{q_{ti}(\boldsymbol{\theta}_1)} = 1 - \frac{q_{ti}(\boldsymbol{\theta}_2)}{q_{ti}(\boldsymbol{\theta}_1)} \leq 1 - \frac{1}{\sqrt{2}} < \frac{1}{2}$$

and similarly

$$\frac{-\delta_{ti}}{q_{ti}(\boldsymbol{\theta}_1)} = \frac{q_{ti}(\boldsymbol{\theta}_2) - q_{ti}(\boldsymbol{\theta}_1)}{q_{ti}(\boldsymbol{\theta}_1)} = \frac{q_{ti}(\boldsymbol{\theta}_2)}{q_{ti}(\boldsymbol{\theta}_1)} - 1 \leq \sqrt{2} - 1 < \frac{1}{2}.$$

With this, we showed that $\frac{1}{2} \tilde{\mathbf{H}} \preccurlyeq \mathbf{H}_2 \preccurlyeq 2\tilde{\mathbf{H}}$. To show the final intended result, the next step is to show $\frac{1}{2} \mathbf{H}_1 \preccurlyeq \tilde{\mathbf{H}} \preccurlyeq 2\mathbf{H}_1$. Similar to the previous part, it is sufficient to show that $-\frac{1}{2} \mathbf{H}_1 \preccurlyeq \tilde{\mathbf{H}} - \mathbf{H}_1 \preccurlyeq \frac{1}{2} \mathbf{H}_1$. We can write this difference as

$$\begin{aligned}\tilde{\mathbf{H}} - \mathbf{H}_1 &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) [\tilde{\mathbf{w}}_{ti} \tilde{\mathbf{w}}_{ti}^\top - \mathbf{w}_{ti} \mathbf{w}_{ti}^\top] \\ &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) [(\mathbf{w}_{ti} + \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\mathbf{w}_{ti} + \bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top - \mathbf{w}_{ti} \mathbf{w}_{ti}^\top] \\ &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) [(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top + (\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)\mathbf{w}_{ti}^\top + \mathbf{w}_{ti}(\bar{\mathbf{x}}_1 - \bar{\mathbf{x}}_2)^\top] \\ &= (\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)(\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1)^\top.\end{aligned}$$

Now, we note that $\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1 = -\sum_{i \in \bar{S}_t} \delta_{ti} \tilde{\mathbf{x}}_{ti}$. On the other hand, we also have

$$\begin{aligned}\sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \mathbf{w}_{ti} &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) (\tilde{\mathbf{x}}_{ti} - \bar{\mathbf{x}}_1) \\ &= \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \tilde{\mathbf{x}}_{ti} - \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) \tilde{\mathbf{x}}_{ti} \\ &= -\sum_{i \in \bar{S}_t} \delta_{ti} \tilde{\mathbf{x}}_{ti}.\end{aligned}$$

Hence, we can write $\bar{\mathbf{x}}_2 - \bar{\mathbf{x}}_1 = \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \mathbf{w}_{ti} = \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_2) \mathbf{w}_{ti} - \sum_{i \in \bar{S}_t} q_{ti}(\boldsymbol{\theta}_1) \mathbf{w}_{ti} = -\sum_{i \in \bar{S}_t} \delta_{ti} \mathbf{w}_{ti}$. Putting our results together, we have

$$\begin{aligned}\tilde{\mathbf{H}} - \mathbf{H}_1 &= \left(\sum_{i \in \bar{S}_t} \delta_{ti} \mathbf{w}_{ti} \right) \left(\sum_{i \in \bar{S}_t} \delta_{ti} \mathbf{w}_{ti} \right)^\top \\ &= \sum_{i \in \bar{S}_t} \sum_{j \in \bar{S}_t} \delta_{ti} \delta_{tj} \mathbf{w}_{ti} \mathbf{w}_{tj}^\top \\ &= \frac{1}{2} \sum_{i \in \bar{S}_t} \sum_{j \in \bar{S}_t} \delta_{ti} \delta_{tj} (\mathbf{w}_{ti} \mathbf{w}_{tj}^\top + \mathbf{w}_{tj} \mathbf{w}_{ti}^\top)\end{aligned}$$

Using the inequality $-2\mathbf{w}_{ti} \mathbf{w}_{ti}^\top \preceq (\mathbf{w}_{ti} \mathbf{w}_{tj}^\top + \mathbf{w}_{tj} \mathbf{w}_{ti}^\top) \preceq 2\mathbf{w}_{ti} \mathbf{w}_{ti}^\top$ and the fact that $\sum_{j \in \bar{S}_t} |\delta_{tj}| \leq 2$, we can show

$$-2 \sum_{i \in \bar{S}_t} |\delta_{ti}| \mathbf{w}_{ti} \mathbf{w}_{ti}^\top \preceq \tilde{\mathbf{H}} - \mathbf{H}_1 \preceq 2 \sum_{i \in \bar{S}_t} |\delta_{ti}| \mathbf{w}_{ti} \mathbf{w}_{ti}^\top$$

Now, we use our result $|\delta_{ti}| \leq \frac{1}{2} q_{ti}(\boldsymbol{\theta}_1)$ to conclude $-\frac{1}{2} \mathbf{H}_1 \preceq \tilde{\mathbf{H}} - \mathbf{H}_1 \preceq \frac{1}{2} \mathbf{H}_1$. This inequality implies $\frac{1}{2} \mathbf{H}_1 \preceq \tilde{\mathbf{H}} \preceq 2\mathbf{H}_1$. Combining with the previous result $\frac{1}{2} \tilde{\mathbf{H}} \preceq \mathbf{H}_2 \preceq 2\tilde{\mathbf{H}}$, we show the final result. \square

Lemma C.4 (Fisher Information Estimation). *Let $t \geq T_0$ and assume that $\hat{\boldsymbol{\theta}}_\tau \in \mathcal{B}_\gamma$ for all $T_0 \leq \tau < t$. Then, for any $\boldsymbol{\theta} \in \mathcal{B}_\gamma$, we have*

$$\mathcal{H}_t(\boldsymbol{\theta}) := \sum_{\tau=1}^{t-1} \mathbf{H}_\tau(\boldsymbol{\theta}) \succcurlyeq C_3 \mathbf{V}_t$$

for some universal constant $C_3 > 0$.

Proof. For any $t \geq T_0$, \mathbf{V}_t is given by

$$\begin{aligned}\mathbf{V}_t &= \frac{1}{K^2} \sum_{\tau=1}^{T_0-1} \sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top + \sum_{\tau=T_0}^{t-1} \left[\sum_{i \in S_\tau} q_{\tau i}(\hat{\boldsymbol{\theta}}_\tau) \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top - \sum_{i \in S_t} \sum_{j \in S_\tau} q_{\tau i}(\hat{\boldsymbol{\theta}}_\tau) q_{\tau j}(\hat{\boldsymbol{\theta}}_\tau) \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau j}^\top \right] \\ &= \frac{1}{K^2} \sum_{\tau=1}^{T_0-1} \sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top + \sum_{\tau=T_0}^{t-1} \mathbf{H}_\tau(\hat{\boldsymbol{\theta}}_\tau).\end{aligned}$$

Now, we will upper bound these two terms separately.

To upper bound the terms for $\tau \geq T_0$, we use Lemma C.3 which states that $\mathbf{H}_\tau(\boldsymbol{\theta}) \succcurlyeq \frac{1}{4} \mathbf{H}_\tau(\hat{\boldsymbol{\theta}}_\tau)$.

For $\tau < T_0$, we use

$$\begin{aligned}\mathbf{H}_\tau(\boldsymbol{\theta}) &= \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top \\ &= \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \frac{1}{2} \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) (\tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top + \tilde{\mathbf{x}}_{tj} \tilde{\mathbf{x}}_{ti}^\top) \\ &\succcurlyeq \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \frac{1}{2} \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) (\tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top + \tilde{\mathbf{x}}_{tj} \tilde{\mathbf{x}}_{tj}^\top) \\ &= \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}) q_{tj}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top \\ &= \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top \\ &\succcurlyeq \frac{\nu}{K^2} \sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top\end{aligned}$$

where we define the constant

$$\nu := K^2 \cdot \min_{t < T_0} \inf_{\boldsymbol{\theta} \in \mathcal{B}_\gamma} q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) > 0.$$

Next, we will show that $\nu > C'_3$ for some universal constant $C'_3 > 0$. For any $\boldsymbol{\theta}$, let $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\phi})$. If $\boldsymbol{\theta} \in \mathcal{B}_\gamma$, we can show that $\|\boldsymbol{\psi} - \boldsymbol{\psi}^*\| \leq \gamma < 1$ and $\|\boldsymbol{\phi} - \boldsymbol{\phi}^*\| \leq \gamma < 1$,

$$\begin{aligned} |\langle \boldsymbol{\psi}, \mathbf{x}_{ti} \rangle| &\leq |\langle \boldsymbol{\psi}^*, \mathbf{x}_{ti} \rangle| + |\langle \boldsymbol{\psi} - \boldsymbol{\psi}^*, \mathbf{x}_{ti} \rangle| \leq 1 + \gamma < 2 \quad \text{and} \\ |\langle \boldsymbol{\phi}, \mathbf{x}_{ti} \rangle| &\leq |\langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle| + |\langle \boldsymbol{\phi} - \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle| \leq 1 + \gamma < 2. \end{aligned}$$

Note that for all $t < T_0$, we have $1 \leq p_{ti} \leq 2$ for all $i \in S_t$. Therefore, for any $\boldsymbol{\theta} \in \mathcal{B}_\gamma$ and $t < T_0$, we have

$$\begin{aligned} q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) &= \frac{\exp(\langle \boldsymbol{\psi}, \mathbf{x}_{ti} \rangle - \langle \boldsymbol{\phi}, \mathbf{x}_{ti} \rangle p_{ti})}{\left(1 + \sum_{j \in S_t} \exp(\langle \boldsymbol{\psi}, \mathbf{x}_{tj} \rangle - \langle \boldsymbol{\phi}, \mathbf{x}_{tj} \rangle p_{tj})\right)^2} \\ &> \frac{e^{-6}}{(1 + Ke^6)^2}, \end{aligned}$$

showing that $\nu > C'_3$ for some constant $C'_3 > 0$.

Letting $C_3 = \min\{C'_3, 1/4\}$, we can show that

$$\begin{aligned} C_3 \mathbf{V}_t &= \frac{C_3}{K^2} \sum_{\tau=1}^{T_0-1} \sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top + C_3 \sum_{\tau=T_0}^{t-1} \mathbf{H}_\tau(\hat{\boldsymbol{\theta}}_\tau) \\ &\preccurlyeq \sum_{\tau=1}^{t-1} \mathbf{H}_\tau(\boldsymbol{\theta}). \end{aligned}$$

□

Lemma C.5 (Consistency of MLE). *Let T_0 be any round such that $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0$. Then, we have*

$$\mathbb{P}(\exists t \geq T_0, \hat{\boldsymbol{\theta}}_t \notin \mathcal{B}_\gamma) \leq \frac{1}{T}.$$

Proof. Recall that the gradient of the negative log-likelihood is given by

$$\nabla_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta}) = \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau} (q_{\tau i}(\boldsymbol{\theta}) - y_{\tau i}) \tilde{\mathbf{x}}_{\tau i}$$

and we have $\nabla_{\boldsymbol{\theta}} \ell_t(\hat{\boldsymbol{\theta}}_t) = 0$ by definition of $\hat{\boldsymbol{\theta}}_t$.

We can write the expectation of $\nabla_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta})$ over the user choices $y_{\tau i}$ as

$$G_t(\boldsymbol{\theta}) := \mathbb{E}[\nabla_{\boldsymbol{\theta}} \ell_t(\boldsymbol{\theta})] = \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau} (q_{\tau i}(\boldsymbol{\theta}) - q_{\tau i}(\boldsymbol{\theta}^*)) \tilde{\mathbf{x}}_{\tau i}.$$

We can show that

$$G_t(\boldsymbol{\theta}^*) = 0 \quad \text{and} \quad G_t(\hat{\boldsymbol{\theta}}_t) = \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau} \epsilon_{\tau i} \tilde{\mathbf{x}}_{\tau i},$$

where $\epsilon_{ti} = y_{ti} - q_{ti}(\boldsymbol{\theta}^*)$ are sub-Gaussian random variables with parameter 1. Note that collections of variables $\{\epsilon_{ti}\}_{i \in S_t}$ are independent over t , but the variables within each collection are not independent.

For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathbb{R}^{2d}$ and any $\mathbf{z} \in \mathbb{R}^{2d}$, the mean value theorem implies that there exists some $\bar{\boldsymbol{\theta}} = \lambda\boldsymbol{\theta}_1 + (1 - \lambda)\boldsymbol{\theta}_2$ with $0 < \lambda < 1$, such that

$$\mathbf{z}^\top (G_t(\boldsymbol{\theta}_1) - G_t(\boldsymbol{\theta}_2)) = \mathbf{z}^\top \mathcal{H}_t(\bar{\boldsymbol{\theta}})(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)$$

where we defined

$$\begin{aligned} \mathcal{H}_t(\boldsymbol{\theta}) &:= \nabla_{\boldsymbol{\theta}} G_t(\boldsymbol{\theta}) \\ &= \sum_{\tau=1}^{t-1} \sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \nabla_{\boldsymbol{\theta}} q_{\tau i}(\boldsymbol{\theta}). \end{aligned}$$

Recalling the definition

$$\begin{aligned} \mathbf{H}_\tau(\boldsymbol{\theta}) &= \sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \nabla_{\boldsymbol{\theta}} q_{\tau i}(\boldsymbol{\theta}) \\ &= \sum_{i \in S_\tau} q_{\tau i}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top - \sum_{i \in S_\tau} \sum_{j \in S_\tau} q_{\tau i}(\boldsymbol{\theta}) q_{\tau j}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau j}^\top, \end{aligned}$$

we also see that $\mathcal{H}_t(\boldsymbol{\theta}) = \sum_{\tau=1}^{t-1} \mathbf{H}_\tau(\boldsymbol{\theta})$.

Now, we're ready to complete the proof with strong induction. The base case is $t = T_0$ and we proceed with inductive steps for each $t \in \{T_0 + 1, T_0 + 2, \dots, T\}$.

We start with proving the inductive steps. Assuming that $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \leq \gamma$ for all $T_0 \leq \tau < t$, we have $\mathcal{H}_t(\boldsymbol{\theta}) \succcurlyeq C_3 \mathbf{V}_t$ for any $\boldsymbol{\theta} \in \mathcal{B}_\gamma$ using Lemma C.4. Therefore, we can write

$$(\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^\top (G_t(\boldsymbol{\theta}_1) - G_t(\boldsymbol{\theta}_2)) \geq C_3 (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2)^\top \mathbf{V}_t (\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2) > 0$$

for any $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ and therefore $G_t(\boldsymbol{\theta})$ is an injection from \mathbb{R}^{2d} to \mathbb{R}^{2d} . This allows us to use Lemma A of Chen et al. (1999) which implies that

$$\left\{ \boldsymbol{\theta} : \|G_t(\boldsymbol{\theta})\|_{\mathbf{V}_t^{-1}} \leq C_3 \gamma \sqrt{\lambda_{\min}(\mathbf{V}_t)} \right\} \subseteq \mathcal{B}_\gamma.$$

In addition, Lemma 15 of Oh and Iyengar (2021) shows that the event

$$\mathcal{E}_G := \left\{ \|G_t(\hat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_t^{-1}} \leq 4\sqrt{4d + \log(1/\delta)} \right\}$$

holds with probability at least $1 - \delta$. Thus, $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \leq \gamma$ holds with probability at least $1 - \delta$ when $\lambda_{\min}(\mathbf{V}_t) \geq \lambda_{\min}(\mathbf{V}_{T_0}) \geq \frac{16}{C_3^2 \gamma^2} (4d + \log(1/\delta))$. Since we have $\lambda_{\min}^0 = C_1 \left(\frac{\bar{P} d \log(T)}{\sigma_0} \right)^2 \geq \frac{16}{C_3^2 \gamma^2} (4d + 2 \log(T))$ for some constant C_1 , the minimum eigenvalue condition is satisfied when $\lambda_{\min}(\mathbf{V}_t) \geq \lambda_{\min}^0$.

For the base case $t = T_0$, we similarly have $\mathcal{H}_{T_0}(\boldsymbol{\theta}) \succcurlyeq C_3 \mathbf{V}_{T_0}$ for any $\boldsymbol{\theta} \in \mathcal{B}_\gamma$ by Lemma C.4. Therefore, we can follow similar steps for $t = T_0$ to argue that $\|\hat{\boldsymbol{\theta}}_{T_0} - \boldsymbol{\theta}^*\| \leq \gamma$ holds true with probability at least $1 - T^{-2}$ when $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \frac{16}{C_3^2 \gamma^2} (4d + 2 \log(T))$. Similarly, the minimum eigenvalue condition is satisfied when $\lambda_{\min}(\mathbf{V}_t) \geq \lambda_{\min}^0$.

Taking a union bound over the base case and the inductive steps of the proof, we complete the proof of the theorem. \square

Lemma C.6 (Normality of MLE). *Suppose the event \mathcal{E}_0 holds. Then, for any $t \geq T_0$,*

$$\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq C_5 \left(\sqrt{d \log \left(1 + \frac{2t^3}{d} \right)} + \frac{1}{\sigma_0} \log(T) \right) \quad (14)$$

with probability at least $1 - \mathcal{O}(t^{-2})$.

Proof. Following the proof of Lemma C.5, we use $\mathcal{H}_t(\bar{\boldsymbol{\theta}}) \succcurlyeq C_3 \mathbf{V}_t$ to obtain

$$\begin{aligned} \|G(\hat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_t^{-1}}^2 &= \|G(\hat{\boldsymbol{\theta}}_t) - G(\boldsymbol{\theta}^*)\|_{\mathbf{V}_t^{-1}}^2 \\ &\geq (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^\top \mathcal{H}_t(\bar{\boldsymbol{\theta}}) \mathbf{V}_t^{-1} \mathcal{H}_t(\bar{\boldsymbol{\theta}}) (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \\ &\geq C_3^2 \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t}^2 \end{aligned} \quad (21)$$

for any $\hat{\boldsymbol{\theta}}_t \in \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq \gamma\}$.

The next step is to upper bound $\|G(\hat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_t^{-1}}$. We first separate it into two terms that correspond to the initialization rounds and the remaining rounds respectively. That is, we write

$$\begin{aligned} \|G(\hat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_t^{-1}} &= \left\| G(\hat{\boldsymbol{\theta}}_{T_0}) + G(\hat{\boldsymbol{\theta}}_t) - G(\hat{\boldsymbol{\theta}}_{T_0}) \right\|_{\mathbf{V}_t^{-1}} \\ &\leq \|G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_t^{-1}} + \|G(\hat{\boldsymbol{\theta}}_t) - G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_t^{-1}} \\ &\leq \|G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_{T_0}^{-1}} + \|G(\hat{\boldsymbol{\theta}}_t) - G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_t^{-1}} \end{aligned}$$

where the last inequality follows from $\mathbf{V}_t \succcurlyeq \mathbf{V}_{T_0}$ for any $t \geq T_0$.

We upper bound the first term using Lemma C.13 which states that

$$\|G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_{T_0}^{-1}} \leq \frac{C_5}{\sigma_0} \log(T) \quad (22)$$

with probability $1 - \mathcal{O}(T^{-2})$.

To upper bound the second term, we use an improved self-normalized bound for vector-valued martingales as given in Theorem D.1. In using this result, we let $\boldsymbol{\epsilon}_t$ denote the random vector with entries $\epsilon_{ti} = y_{ti} - q_{ti}(\boldsymbol{\theta}^*)$ and we let $\tilde{\mathbf{X}}_t \in R^{2d \times K}$ denote the matrix with columns $\tilde{\mathbf{x}}_{ti}$. We note that we have $\|\boldsymbol{\epsilon}_t\|_1 \leq 2$ and

$$\tilde{\mathbf{X}}_t \Sigma_t \tilde{\mathbf{X}}_t^\top = H_t(\boldsymbol{\theta}^*) = \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\boldsymbol{\theta}^*) q_{tj}(\boldsymbol{\theta}^*) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top$$

where Σ_t is the covariance matrix $\mathbb{E}[\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_t^\top]$. As a result, Theorem D.1 shows that

$$\|G(\hat{\boldsymbol{\theta}}_t) - G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_t^{-1}} \leq \frac{\sqrt{\lambda}}{4} + \frac{4}{\sqrt{\lambda}} \log\left(\frac{\det(\mathbf{V}_t)^{1/2}}{\delta \lambda^d}\right) + \frac{8}{\sqrt{\lambda}} d \log(2)$$

with probability at least $1 - \delta$ for any $0 < \lambda < \lambda_{\min}(\mathbf{V}_{T_0})$.

Then we combine with Lemma C.9 to obtain

$$\begin{aligned} \|G(\hat{\boldsymbol{\theta}}_t) - G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_t^{-1}} &\leq \frac{\sqrt{\lambda}}{4} + \frac{4d}{\sqrt{\lambda}} \log\left(\frac{t \bar{P}^2}{d \lambda}\right) + \frac{4}{\sqrt{\lambda}} \log\left(\frac{1}{\delta}\right) + \frac{8}{\sqrt{\lambda}} d \log(2) \\ &\leq \frac{\sqrt{\lambda}}{4} + \frac{4d}{\sqrt{\lambda}} \left(\log\left(\frac{t}{d}\right) + \log\left(\frac{1}{\delta}\right) + 2 \log(2) \right) \\ &\leq \frac{\sqrt{\lambda}}{4} + \frac{4d}{\sqrt{\lambda}} \log\left(1 + \frac{2t}{d\delta}\right) \end{aligned}$$

for any $\lambda \geq \bar{P}^2$. Accordingly, we set $\lambda = \max\left\{\bar{P}^2, 16d \log\left(1 + \frac{2t}{d\delta}\right)\right\}$ to obtain

$$\|G(\hat{\boldsymbol{\theta}}_t) - G(\hat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_t^{-1}} \leq 2 \sqrt{d \log\left(1 + \frac{2t}{d\delta}\right)}. \quad (23)$$

Now, we set $\delta = t^{-2}$ and obtain $\lambda = 16d \log\left(1 + \frac{2t}{d\delta}\right) \geq 16d \log\left(1 + \frac{2}{d}\right) \geq 1$. Lastly, we confirm that $1 < \lambda < \lambda_{\min}(\mathbf{V}_{T_0})$ is satisfied for our selection of λ . On the other hand, we can verify that $\lambda < \lambda_{\min}(\mathbf{V}_{T_0})$ is satisfied under good event \mathcal{E}_0 because $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0 = C_1 \frac{d \bar{P}^2 \log^3(T)}{\sigma_0^2} > \lambda$ for some constant C_1 .

Combining (22) and (23) gives the stated result in the lemma. \square

Lemma C.7. Suppose \mathcal{E}_t holds for all $t \geq T_0$. Let $\tilde{h}_{ti} : \mathbb{R} \rightarrow \mathbb{R}$ be the function defined as

$$\tilde{h}_{ti}(p) := \langle \hat{\psi}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\phi}_t, \mathbf{x}_{ti} \rangle \cdot p + g_{ti}(p),$$

where $\hat{\theta}_t = (\hat{\psi}_t, \hat{\phi}_t)$ and $g_{ti}(p) := \alpha_t \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}}$. Furthermore, let $h_{ti} : \mathbb{R} \rightarrow \mathbb{R}$ be the function defined as

$$h_{ti}(p) := \min_{p' \leq p} \left\{ \tilde{h}_{ti}(p') - L_0(p - p') \right\}.$$

Then, the function $h_{ti}(p)$ is differentiable and satisfies

$$h_{ti}(p) \geq u_{ti}(p), \tag{16}$$

$$h_{ti}(p) - u_{ti}(p) \leq 2g_{ti}(p), \tag{17}$$

$$h'_{ti}(p) \leq -L_0, \tag{18}$$

$$h_{ti}(0) \leq 2 \tag{19}$$

for all $p \in \mathbb{R}$.

Proof. Recall the definition of the utility function

$$u_{ti}(p) = \langle \psi^*, \mathbf{x}_{ti} \rangle - \langle \phi^*, \mathbf{x}_{ti} \rangle \cdot p.$$

and recall the definition $\tilde{\mathbf{x}}_{ti} = (\mathbf{x}_{ti}, -p\mathbf{x}_{ti})$ to write

$$\begin{aligned} |\langle \hat{\theta}_t, \tilde{\mathbf{x}}_{ti} \rangle - \langle \theta^*, \tilde{\mathbf{x}}_{ti} \rangle| &= \left| \langle \mathbf{V}_t^{1/2}(\hat{\theta}_t - \theta^*), \mathbf{V}_t^{-1/2}\tilde{\mathbf{x}}_{ti} \rangle \right| \\ &\leq \|\mathbf{V}_t^{1/2}(\hat{\theta}_t - \theta^*)\| \|\mathbf{V}_t^{-1/2}\tilde{\mathbf{x}}_{ti}\| \\ &\leq \|\hat{\theta}_t - \theta^*\|_{\mathbf{V}_t} \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}} \\ &\leq \|\hat{\theta}_t - \theta^*\|_{\mathbf{V}_t} \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}} \\ &\leq \alpha_t \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}}. \end{aligned}$$

where $\|\hat{\theta}_t - \theta^*\|_{\mathbf{V}_t} \leq \alpha_t$ follows from the assumption that \mathcal{E}_t holds. Hence, we obtain

$$\langle \hat{\psi}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\phi}_t, \mathbf{x}_{ti} \rangle \cdot p - g_{ti}(p) \leq u_{ti}(p) \leq \langle \hat{\psi}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\phi}_t, \mathbf{x}_{ti} \rangle \cdot p + g_{ti}(p),$$

showing that $u_{ti}(p) \leq \tilde{h}_{ti}(p) \leq u_{ti}(p) + 2g_{ti}(p)$ for all $p \in \mathbb{R}$.

Since $u'_{ti}(p) \leq -L_0$ for all $p \in \mathbb{R}$, we also have $u_{ti}(p) \leq u_{ti}(p') - L_0(p - p') \leq \tilde{h}_{ti}(p') - L_0(p - p')$ for any $p' \leq p$. Therefore, $u_{ti}(p) \leq h_{ti}(p)$ for all $p \in \mathbb{R}$ proving condition (16).

On the other hand, we have $h_{ti}(p) \leq \tilde{h}_{ti}(p) \leq u_{ti}(p) + 2g_{ti}(p)$ for all $p \in \mathbb{R}$ proving (17).

Furthermore, $h_{ti}(0) \leq \tilde{h}_{ti}(0) \leq u_{ti}(0) + 2g_{ti}(0) \leq 1 + \alpha_t \|(\mathbf{x}_{ti}, \mathbf{0})\|_{\mathbf{V}_t^{-1}} \leq 1 + \alpha_t / \sqrt{\lambda_{\min}^0} \leq 1 + \mathcal{O}\left(\frac{1}{P\sqrt{\log T}}\right)$. As a result, $h_{ti}(0) \leq 2$ for sufficiently large T .

Next, we show that $h_{ti}(p)$ is a differentiable function and its derivative is at most $-L_0$ for all $p \in \mathbb{R}$. Notice that the function $\tilde{h}_{ti}(p)$ can be written as $\tilde{h}_{ti}(p) = y(p) + c\sqrt{z(p)}$ for a linear function $y : \mathbb{R} \rightarrow \mathbb{R}$ and a positive quadratic function $z : \mathbb{R} \rightarrow \mathbb{R}_+$ of the form $z(p) = a + bp + p^2$ satisfying $4a - b^2 > 0$. With this notation, the second derivative of $\tilde{h}_{ti}(p)$ is given as

$$\tilde{h}_{ti}''(p) = \frac{4a - b^2}{4(a + p(b + p))^{3/2}} > 0.$$

Therefore, $\tilde{h}_{ti}(p)$ is smooth and strictly convex. Let p_0 be the unique value such that $\tilde{h}'_{ti}(p_0) = -L_0$.

We let p^\dagger denote the value of p' that minimizes the function $\tilde{h}_{ti}(p') - L_0(p - p')$ over $(-\infty, p]$. As a result, we obtain $h_{ti}(p) = \tilde{h}_{ti}(p^\dagger) - L_0(p - p^\dagger)$. Using that the function $\tilde{h}_{ti}(p') - L_0(p - p')$ is convex, we can write

$$p^\dagger = \begin{cases} p_0 & \text{if } p_0 \leq p, \\ p & \text{if } p < p_0. \end{cases}$$

Consequently, we obtain

$$h_{ti}(p) = \begin{cases} \tilde{h}_{ti}(p_0) - L_0(p - p_0) & \text{if } p \geq p_0, \\ \tilde{h}_{ti}(p) & \text{if } p < p_0. \end{cases}$$

The function $h_{ti}(p)$ is differentiable everywhere including $p = p_0$ since $\tilde{h}'_{ti}(p_0) = -L_0$. Furthermore, $h'_{ti}(p) \leq -L_0$ for all $p \geq 0$ since $\tilde{h}'_{ti}(p) \leq -L_0$ for $p < p_0$. Consequently, we prove property (18). \square

Lemma C.8. Assume good event \mathcal{E}_t holds for some $t \geq T_0$. Then,

- (a) $R_t(S_t^*, \mathbf{p}_t^*) \leq \tilde{R}_t(S_t, \mathbf{p}_t)$, and
- (b) $\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t) \leq 4P \cdot \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) g_{ti}(p_{ti})$.

Proof. Inequality (a): Fix some t and define revenue functions $R^A : 2^{[N]} \times \mathbb{R}_+^N \rightarrow \mathbb{R}$ given by

$$R^A(S, \mathbf{p}) = \frac{\sum_{i \in S \setminus A} p_i \exp(u_{ti}(p_i)) + \sum_{i \in S \cap A} p_i \exp(h_{ti}(p_i))}{1 + \sum_{i \in S \setminus A} \exp(u_{ti}(p_i)) + \sum_{i \in S \cap A} \exp(h_{ti}(p_i))}$$

for any $A \subseteq [N]$. Note that this definition leads to $R^\emptyset(S, \mathbf{p}) = R_t(S, \mathbf{p})$ and $R^S(S, \mathbf{p}) = \tilde{R}_t(S, \mathbf{p})$. We also define

$$(S^A, \mathbf{p}^A) = \underset{\substack{S \subseteq [N]: |S| \leq K \\ \mathbf{p} \in \mathbb{R}_+^N}}{\operatorname{argmax}} R^A(S, \mathbf{p}).$$

which satisfies $(S^\emptyset, \mathbf{p}^\emptyset) = (S_t^*, \mathbf{p}_t^*)$ and $(S^{[N]}, \mathbf{p}^{[N]}) = (S_t, \mathbf{p}_t)$.

By the optimality of (S^A, \mathbf{p}^A) for any revenue function R^A , we have $p_j^A \geq R^A(S^A, \mathbf{p}^A)$ for all $j \in S^A$. We can write this inequality as $p_j^A \geq a/b$ where

$$\begin{aligned} a &= \sum_{i \in S^A \setminus A} p_i^A \exp(u_{ti}(p_i^A)) + \sum_{i \in S^A \cap A} p_i^A \exp(h_{ti}(p_i^A)) \quad \text{and} \\ b &= 1 + \sum_{i \in S^A \setminus A} \exp(u_{ti}(p_i^A)) + \sum_{i \in S^A \cap A} \exp(h_{ti}(p_i^A)). \end{aligned}$$

Letting $\delta = \exp(h_{tj}(p_j^A)) - \exp(u_{tj}(p_j^A))$, we have $ab + b\delta p_j^A \geq ab + a\delta$ which implies

$$\frac{a + p_j^A \delta}{b + \delta} \geq \frac{a}{b}.$$

Hence, we have $R^{A \cup \{j\}}(S^A, \mathbf{p}^A) \geq R^A(S^A, \mathbf{p}^A)$ for all $j \in S^A$.

We also have $R^{A \cup \{j\}}(S^A, \mathbf{p}^A) = R^A(S^A, \mathbf{p}^A)$ for any $j \notin S^A$. Therefore, $R^{A \cup \{j\}}(S^A, \mathbf{p}^A) \geq R^A(S^A, \mathbf{p}^A)$ for any $j \in [N]$. Using the optimality of $(S^{A \cup \{j\}}, \mathbf{p}^{A \cup \{j\}})$ for function $R^{A \cup \{j\}}$, we can write

$$R^{A \cup \{j\}}(S^{A \cup \{j\}}, \mathbf{p}^{A \cup \{j\}}) \geq R^A(S^A, \mathbf{p}^A)$$

for any $j \in [N]$. Therefore, by induction, we can show that

$$\tilde{R}_t(S_t, \mathbf{p}_t) = R^{[N]}(S^{[N]}, \mathbf{p}^{[N]}) \geq R^\emptyset(S^\emptyset, \mathbf{p}^\emptyset) = R_t(S_t^*, \mathbf{p}_t^*).$$

Inequality (b): Let $u_{ti} := u_{ti}(p_{ti})$ and $h_{ti} := h_{ti}(p_{ti})$ with $2g_{ti}(p_{ti}) \geq h_{ti} - u_{ti} \geq 0$. By the mean value theorem, for any i , there exists $z_{ti} := (1 - c)u_{ti} + ch_{ti}$ for some $c \in (0, 1)$ such that

$$\begin{aligned} \tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t) &= \frac{\sum_{i \in S_t} p_{ti} \exp(h_{ti})}{1 + \sum_{j \in S_t} \exp(h_{tj})} - \frac{\sum_{i \in S_t} p_{ti} \exp(u_{ti})}{1 + \sum_{j \in S_t} \exp(u_{tj})} \\ &= \frac{(\sum_{i \in S_t} p_{ti} \exp(z_{ti})(h_{ti} - u_{ti}))(1 + \sum_{i \in S_t} \exp(z_{ti}))}{(1 + \sum_{i \in S_t} \exp(z_{ti}))^2} \\ &\quad - \frac{(\sum_{i \in S_t} p_{ti} \exp(z_{ti}))(\sum_{i \in S_t} \exp(z_{ti})(h_{ti} - u_{ti}))}{(1 + \sum_{i \in S_t} \exp(z_{ti}))^2} \\ &= \sum_{i \in S_t} p_{ti} q_t(i | \mathbf{z}_t)(h_{ti} - u_{ti}) \\ &\quad - \left(\sum_{i \in S_t} p_{ti} q_t(i | \mathbf{z}_t) \right) \left(\sum_{i \in S_t} q_t(i | \mathbf{z}_t)(h_{ti} - u_{ti}) \right) \\ &= \sum_{i \in S_t} \left(p_{ti} - \sum_{i \in S_t} p_{ti} q_t(i | \mathbf{z}_t) \right) q_t(i | \mathbf{z}_t)(h_{ti} - u_{ti}) \\ &\leq P \cdot \sum_{i \in S_t} q_t(i | \mathbf{z}_t)(h_{ti} - u_{ti}) \\ &\leq 2P \cdot \sum_{i \in S_t} q_t(i | \mathbf{z}_t) g_{ti}(p_{ti}) \end{aligned}$$

where the first inequality follows from $|p_{ti}| \leq P$ and $q_t(i | \mathbf{z}_t)$ is a categorical distribution given by

$$q_t(i | \mathbf{z}_t) = \frac{\exp(z_{ti})}{1 + \sum_{j \in S_t} \exp(z_{tj})}$$

for $i \in S_t$. Then, noting that

$$\begin{aligned} g_{ti}(p_{ti}) &= \alpha_t \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}} \\ &\leq \alpha_t \frac{1}{\sqrt{\lambda_{\min}(\mathbf{V}_{T_0})}} \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_2 \\ &\leq \alpha_t \frac{1}{\sqrt{\lambda_{\min}(\mathbf{V}_{T_0})}} (1 + P) \\ &\leq \frac{\log 2}{2}, \end{aligned}$$

we have $u_{ti} \leq z_{ti} \leq u_{ti} + \log 2$. Hence,

$$\frac{1}{2} \leq \frac{q_t(i | \mathbf{z}_t)}{q_{ti}(\boldsymbol{\theta}^*)} \leq 2$$

for all $i \in S_t$. Consequently, we obtain

$$\tilde{R}_t(S_t, \mathbf{p}_t) - R_t(S_t, \mathbf{p}_t) \leq 4P \cdot \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) g_{ti}(p_{ti}),$$

completing the proof. \square

Lemma C.9. For $t > T_0$, $\det(\mathbf{V}_t)$ is increasing with respect to t and $\det(\mathbf{V}_t) \leq (t\bar{P}^2/d)^{2d}$.

Proof. Let $\lambda_1, \dots, \lambda_{2d}$ be the eigenvalues of \mathbf{V}_t . Then, using the AM-GM inequality we can write

$$\begin{aligned} \det(\mathbf{V}_t) &= \prod_{i=1}^{2d} \lambda_i \\ &\leq \left(\frac{\sum_{i=1}^{2d} \lambda_i}{2d} \right)^{2d} \\ &= \left(\frac{\text{trace}(\mathbf{V}_t)}{2d} \right)^{2d} \\ &\leq \left(\frac{\sum_{s=1}^{T_0} \sum_{i \in S_s} \frac{1}{K^2} \|\tilde{\mathbf{x}}_{si}\|_2^2 + 2 \sum_{s=T_0+1}^t \sum_{i \in S_s} q_{si}(\hat{\boldsymbol{\theta}}_s) \|\tilde{\mathbf{x}}_{si}\|_2^2}{2d} \right)^{2d} \\ &\leq \left(\frac{t \bar{P}^2}{d} \right)^{2d}. \end{aligned}$$

□

Lemma C.10. *If good events \mathcal{E}_0 and \mathcal{E}_t hold for all $t \geq T_0$, then*

$$\sum_{t=T_0}^T \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \leq 18K \log \left(\frac{\det(\mathbf{V}_{T+1})}{\det(\mathbf{V}_{T_0})} \right).$$

Proof. Let $\lambda_1, \dots, \lambda_{2d}$ be the eigenvalues of $H_t(\hat{\boldsymbol{\theta}}_t) = \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{tj}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top$. Since $H_t(\hat{\boldsymbol{\theta}}_t)$ is positive semi-definite, $\lambda_j \geq 0$ for all j . Then, we have

$$\begin{aligned} \det \left(\mathbf{I} + \mathbf{V}_t^{-1/2} H_t(\hat{\boldsymbol{\theta}}_t) \mathbf{V}_t^{-1/2} \right) &= \prod_{i=1}^{2d} (1 + \lambda_i) \\ &\geq 1 + \sum_{i=1}^{2d} \lambda_i \\ &= 1 - 2d + \sum_{i=1}^{2d} (1 + \lambda_i) \\ &= 1 - 2d + \text{trace} \left(\mathbf{I} + \mathbf{V}_t^{-1/2} H_t(\hat{\boldsymbol{\theta}}_t) \mathbf{V}_t^{-1/2} \right) \\ &= 1 + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{tj}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{tj}^\top \mathbf{V}_t^{-1} \tilde{\mathbf{x}}_{ti} \\ &\geq 1 + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \end{aligned}$$

using the inequality $\tilde{\mathbf{x}}_{tj}^\top \mathbf{V}_t^{-1} \tilde{\mathbf{x}}_{ti} + \tilde{\mathbf{x}}_{ti}^\top \mathbf{V}_t^{-1} \tilde{\mathbf{x}}_{tj} \leq \tilde{\mathbf{x}}_{ti}^\top \mathbf{V}_t^{-1} \tilde{\mathbf{x}}_{ti} + \tilde{\mathbf{x}}_{tj}^\top \mathbf{V}_t^{-1} \tilde{\mathbf{x}}_{tj} = \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 + \|\tilde{\mathbf{x}}_{tj}\|_{\mathbf{V}_t^{-1}}^2$.

Now, to lower bound $\det(\mathbf{V}_{T+1})$, we write

$$\begin{aligned} \det(\mathbf{V}_{T+1}) &= \det \left(\mathbf{V}_T + H_T(\hat{\boldsymbol{\theta}}_T) \right) \\ &= \det(\mathbf{V}_T) \det \left(\mathbf{I} + \mathbf{V}_T^{-1/2} H_T(\hat{\boldsymbol{\theta}}_T) \mathbf{V}_T^{-1/2} \right) \\ &\geq \det(\mathbf{V}_{T_0}) \prod_{t=T_0}^T \left(1 + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \right). \end{aligned}$$

Now, using that $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \bar{P}^2$ which is satisfied under event \mathcal{E}_0 , we have

$$\|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \leq \frac{\|\tilde{\mathbf{x}}_{ti}\|^2}{\lambda_{\min}(\mathbf{V}_t)} \leq \frac{(1 + P^2)}{\lambda_{\min}(\mathbf{V}_t)} \leq \frac{\bar{P}^2}{\bar{P}^2} = 1.$$

Hence, $\sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \leq 1$ for all $t \geq T_0$. Then, using the fact that $z \leq 2 \log(1+z)$ for any $z \in [0, 1]$,

$$\begin{aligned} \sum_{t=T_0}^T \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 &\leq 2 \sum_{t=T_0}^T \log \left(1 + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \right) \\ &= 2 \log \prod_{t=T_0}^T \left(1 + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \right) \\ &\leq 2 \log \left(\frac{\det(\mathbf{V}_{T+1})}{\det(\mathbf{V}_{T_0})} \right) \end{aligned}$$

Note that we have $p_{ti} \geq 0$ for all $i \in S_t$ and all $t \geq T_0$. Furthermore, for any $\hat{\boldsymbol{\theta}}_t \in \mathcal{B}_\gamma$, we have $\langle \hat{\boldsymbol{\phi}}_t, \mathbf{x}_{tj} \rangle \geq L_0 - \gamma \geq 0$ and $\langle \hat{\boldsymbol{\psi}}_t, \mathbf{x}_{tj} \rangle \leq 1 + \gamma \leq 2$. Hence, we can lower bound $q_{t0}(\hat{\boldsymbol{\theta}}_t)$ as

$$q_{t0}(\hat{\boldsymbol{\theta}}_t) = \frac{1}{1 + \sum_{j \in S_t} \exp(\langle \hat{\boldsymbol{\psi}}_t, \mathbf{x}_{tj} \rangle - \langle \hat{\boldsymbol{\phi}}_t, \mathbf{x}_{tj} \rangle p_{tj})} > \frac{1}{(1 + K e^2)} \geq \frac{1}{9K}.$$

Combining this result with the previous inequality, we show the intended result. \square

Lemma C.11. *If good events \mathcal{E}_0 and \mathcal{E}_t hold for all $t \geq T_0$, then*

$$\sum_{t=T_0}^T \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 \leq 36dK \log(T/d).$$

Proof. Combining Lemma C.9 and Lemma C.10, we obtain

$$\begin{aligned} \sum_{t=T_0}^T \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2 &\leq 18K \log \left(\frac{\det(\mathbf{V}_T)}{\det(\mathbf{V}_{T_0})} \right) \leq 18K \log \left(\frac{T \bar{P}^2}{d \lambda_{\min}(\mathbf{V}_{T_0})} \right)^{2d} \\ &\leq 36dK \log(T/d) \end{aligned}$$

where the last inequality is by $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \bar{P}^2$ which is satisfied under event \mathcal{E}_0 . \square

Lemma C.12. *For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{B}_\gamma$, we have*

$$\frac{1}{\sqrt{2}} \leq \frac{q_{ti}(\boldsymbol{\theta}_1)}{q_{ti}(\boldsymbol{\theta}_2)} \leq \sqrt{2}$$

for all $i \in S_t$.

Proof. Let $z_{ti}^1 = \exp(\langle \boldsymbol{\psi}_1, \mathbf{x}_{ti} \rangle - \langle \boldsymbol{\phi}_1, \mathbf{x}_{ti} \rangle p_{ti})$ and $z_{ti}^2 = \exp(\langle \boldsymbol{\psi}_2, \mathbf{x}_{ti} \rangle - \langle \boldsymbol{\phi}_2, \mathbf{x}_{ti} \rangle p_{ti})$ for all $i \in S_t$. Then, we have

$$\begin{aligned} \frac{z_{ti}^1}{z_{ti}^2} &= \exp(\langle \boldsymbol{\psi}_1 - \boldsymbol{\psi}_2, \mathbf{x}_{ti} \rangle - \langle \boldsymbol{\phi}_1 - \boldsymbol{\phi}_2, \mathbf{x}_{ti} \rangle p_{ti}) \\ &= \exp(\langle \boldsymbol{\theta}_1 - \boldsymbol{\theta}_2, \tilde{\mathbf{x}}_{ti} \rangle). \end{aligned}$$

Therefore, we have $e^{-2\gamma\bar{P}} = 1/\sqrt[4]{2} \leq z_{ti}^1/z_{ti}^2 \leq \sqrt[4]{2} = e^{2\gamma\bar{P}}$. On the other hand, we have

$$\frac{q_{ti}(\boldsymbol{\theta}_1)}{q_{ti}(\boldsymbol{\theta}_2)} = \frac{z_{ti}^1}{1 + \sum_{j \in S_t} z_{tj}^1} \cdot \frac{1 + \sum_{j \in S_t} z_{tj}^2}{z_{ti}^2}$$

Now, we note that for any two sets of positive numbers $\{a_i\}_{i \in S}$ and $\{b_i\}_{i \in S}$ such that $1/c \leq a_i/b_i \leq c$ for some $c > 1$, we have $1/c \leq (\sum_{i \in S} a_i)/(\sum_{i \in S} b_i) \leq c$. Using this result, we complete the proof. \square

Lemma C.13. *If the number of initialization rounds satisfies*

$$T_0 \geq \left(\frac{C_3\sqrt{d} + C_4\sqrt{\log T}}{\sigma_0} \right)^2$$

for some universal constants $C_3 = \sqrt{2}C_1$ and $C_4 = \max\{C_2, 10\}$, then

$$\|G_{T_0}(\widehat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_{T_0}^{-1}} = \left\| \sum_{t=1}^{T_0} \sum_{i \in S_t} \epsilon_{ti} \tilde{\mathbf{x}}_{ti} \right\|_{\mathbf{V}_{T_0}^{-1}} \leq \frac{C_5}{\sigma_0} \log(T).$$

with probability at least $1 - \mathcal{O}(T^{-2})$ for some universal constant $C_5 = 192$.

Proof. We have

$$\|G_{T_0}(\widehat{\boldsymbol{\theta}}_{T_0})\|_{\mathbf{V}_{T_0}^{-1}}^2 \leq \frac{1}{\lambda_{\min}(\mathbf{V}_{T_0})} \|G_{T_0}(\widehat{\boldsymbol{\theta}}_{T_0})\|_2^2.$$

Let $\mathbf{z}_t := \sum_{i \in S_t} \epsilon_{ti} \tilde{\mathbf{x}}_{ti}$ and recall the definition $\epsilon_{ti} = y_{ti} - q_{ti}(\boldsymbol{\theta}^*)$, to write

$$\mathbf{z}_t = \sum_{i \in S_t} \epsilon_{ti} \tilde{\mathbf{x}}_{ti} = \sum_{i \in S_t} y_{ti} \tilde{\mathbf{x}}_{ti} - \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \tilde{\mathbf{x}}_{ti}.$$

Furthermore, note that $\|\tilde{\mathbf{x}}_{ti}\| \leq 3$ for all $t \leq T_0$ since $1 \leq p_{ti} \leq 2$ for all $t \leq T_0$. Therefore, $\|\mathbf{z}_t\| \leq \|\sum_{i \in S_t} y_{ti} \tilde{\mathbf{x}}_{ti}\| + \|\sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \tilde{\mathbf{x}}_{ti}\| \leq 6$.

Then, using the vector Bernstein inequality from Kohler and Lucchi (2017), we have

$$\|G_{T_0}(\widehat{\boldsymbol{\theta}}_{T_0})\|_2^2 \leq 48T_0 \log(2T^2)$$

with probability at least $1 - T^{-2}$ given that $T_0 > 24 \log(2T^2)$. Note that we satisfy $T_0 > 24 \log(2T^2)$ with the condition given for T_0 in the statement of the lemma because $\sigma_0 \leq 1$.

On the other hand, by Lemma C.2, we have

$$\begin{aligned} \lambda_{\min}(\mathbf{V}_{T_0}) &\geq \frac{\sigma_0}{2} \left(T_0 - \left(\frac{C_1\sqrt{d} + C_2\sqrt{\log T}}{\sigma_0} \right)^2 \right) \\ &\geq \frac{\sigma_0 T_0}{4}. \end{aligned}$$

with probability at least $1 - T^{-2}$. After combining these inequalities, we obtain the intended result. \square

C.2 Estimating Minimum Price Sensitivity

Assume the time horizon is large enough so that $\frac{1}{L_0} < \frac{1}{4}T^{1/4}\sqrt{\frac{\sigma_0}{K}}$. Then, instead of setting the parameters for the initialization round using the true value of \bar{P} (which requires knowing L_0), we set the target minimum eigenvalue

as $\lambda_{\min}^0 = \Theta(T^{1/2})$. As a result, we can use the initialization result established in Lemma C.2 to translate this target to an initialization period of length

$$T_0 = \Theta\left(\frac{\lambda_{\min}^0 K}{\sigma_0}\right) = \Theta\left(\sqrt{T}\right).$$

As the consistency results established in Lemma C.5 shows, the MLE estimate at time T_0 satisfies

$$\|\hat{\boldsymbol{\theta}}_{T_0} - \boldsymbol{\theta}^*\|_2 \leq T^{-1/4} \sqrt{\frac{K}{\sigma_0}} < \frac{L_0}{4}.$$

with probability $1 - \mathcal{O}(T^{-1})$.

Consequently, we have $\|\hat{\boldsymbol{\phi}}_{T_0} - \boldsymbol{\phi}^*\| \leq L_0/4$. Then, we can write

$$\begin{aligned} \langle \hat{\boldsymbol{\phi}}_{T_0}, \mathbf{x}_{ti} \rangle &= \langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle + \langle \hat{\boldsymbol{\phi}}_{T_0} - \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle \\ &\geq L_0 - \|\hat{\boldsymbol{\phi}}_{T_0} - \boldsymbol{\phi}^*\| \|\mathbf{x}_{ti}\| \\ &\geq L_0 - \frac{L_0}{4} \\ &\geq \frac{3L_0}{4}. \end{aligned}$$

Taking the minimum over all previous iterations, we have

$$\begin{aligned} \min_{t \in [T_0], i \in [N]} \langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle &= \min_{t \in [T_0], i \in [N]} \left\{ \langle \hat{\boldsymbol{\phi}}_{T_0} - \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle + \langle \hat{\boldsymbol{\phi}}_{T_0}, \mathbf{x}_{ti} \rangle \right\} \\ &= \min_{t \in [T_0], i \in [N]} \langle \hat{\boldsymbol{\phi}}_{T_0} - \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle + \min_{t \in [T_0], i \in [N]} \langle \hat{\boldsymbol{\phi}}_{T_0}, \mathbf{x}_{ti} \rangle \\ &\geq -T^{-1/4} \sqrt{\frac{K}{\sigma_0}} + \min_{t \in [T_0], i \in [N]} \langle \hat{\boldsymbol{\phi}}_{T_0}, \mathbf{x}_{ti} \rangle. \end{aligned}$$

Then, we estimate the minimum sensitivity parameter as

$$\widehat{L}_0 = \min_{t \in [T_0], i \in [N]} \langle \hat{\boldsymbol{\phi}}_{T_0}, \mathbf{x}_{ti} \rangle - T^{-1/4} \sqrt{\frac{K}{\sigma_0}}.$$

Note that \widehat{L}_0 satisfies $\min_{t \in [T_0], i \in [N]} \langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle \geq \widehat{L}_0$. Furthermore, $\widehat{L}_0 \geq -\frac{L_0}{4} + \frac{3L_0}{4} = \frac{L_0}{2} > 0$.

The next step is to upper bound the expected number of rounds in which there is a context vector \mathbf{x}_{ti} such that $\langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle \leq \widehat{L}_0$. Note that the contexts are sampled independently from an identical distribution by our assumption. Therefore, each ordering among $\langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle$ values for $t \in [T]$ and $i \in [N]$ is equally likely. As a result, the expected number of time-item index pairs $(t, i) \geq T_0 \times [N]$ for which $\langle \boldsymbol{\phi}^*, \mathbf{x}_{ti} \rangle \leq \widehat{L}_0$ can be upper bounded as

$$\sum_{k=0}^{\infty} k \left(1 - \frac{N\sqrt{T}}{NT}\right)^k \left(\frac{N\sqrt{T}}{NT}\right) = \sum_{k=0}^{\infty} k \left(1 - \frac{1}{\sqrt{T}}\right)^k \left(\frac{1}{\sqrt{T}}\right) \leq \sqrt{T}.$$

Consequently, the minimum sensitivity parameter estimate \widehat{L}_0 fails only in \sqrt{T} rounds and causes additional $P\sqrt{T} = \mathcal{O}(\log K \sqrt{T}/L_0)$ regret. In total, this algorithm still manages to achieve an asymptotic regret rate of $\mathcal{O}(d\sqrt{KT}/L_0)$.

D Self-Normalized Bounds for Vector-Valued Martingales

Theorem D.1. Let $\{\mathcal{F}_t\}_{t=1}^\infty$ be a filtration. Let $\{\mathbf{X}_t\}_{t=1}^\infty$ be a stochastic process such that $\mathbf{X}_t \in \mathbb{R}^{d \times K}$ is \mathcal{F}_t measurable and the columns of \mathbf{X}_t denoted by \mathbf{x}_{ti} satisfy $\|\mathbf{x}_{ti}\| \leq B$ almost surely for some $B > 0$. Let $\{\epsilon_t\}_{t=1}^\infty$ be

a martingale difference process such that $\epsilon_t \in \mathbb{R}^K$ is \mathcal{F}_{t+1} -measurable. Let $\mathbf{H}_0 \in \mathbb{R}^{d \times d}$ such that $\lambda_{\min}(\mathbf{H}_0) > \lambda$ for some $\lambda > 0$. Furthermore, assume that we have $\|\epsilon_t\|_1 \leq 2$ almost surely conditional on \mathcal{F}_t and the conditional covariance is given by $\Sigma_t := \mathbb{E}[\epsilon_t \epsilon_t^\top | \mathcal{F}_t]$. For any $t \geq 1$ define

$$\mathbf{S}_t = \sum_{s=1}^{t-1} \mathbf{X}_s \epsilon_s \quad \text{and} \quad \mathbf{H}_t = \mathbf{H}_0 + \sum_{s=1}^{t-1} \mathbf{X}_s \Sigma_s \mathbf{X}_s^\top.$$

Then, for any $\delta \in (0, 1]$, with probability at least $1 - \delta$, for all $t \geq 1$, we have

$$\|\mathbf{S}_t\|_{\mathbf{H}_t^{-1}} \leq \frac{\sqrt{\lambda}}{4} + \frac{4}{\sqrt{\lambda}} \log \left(\frac{\det(\mathbf{H}_t)^{1/2}}{\delta \lambda^{d/2}} \right) + \frac{4}{\sqrt{\lambda}} d \log(2).$$

Proof. Let $\tilde{\mathbf{H}}_t = \sum_{s=1}^{t-1} \mathbf{X}_s \Sigma_s \mathbf{X}_s^\top$ and define the function

$$M_t(\boldsymbol{\xi}) = \exp(\boldsymbol{\xi}^\top \mathbf{S}_t - \|\boldsymbol{\xi}\|_{\tilde{\mathbf{H}}_t}^2),$$

for any $t \geq 1$ and $\boldsymbol{\xi} \in \mathbb{R}^d$. For $t = 0$, let $M_0(\boldsymbol{\xi}) = 0$.

By Lemma D.2, we can show that $\{M_t(\boldsymbol{\xi})\}_{t=1}^\infty$ is a non-negative super-martingale for any $\|\boldsymbol{\xi}\|_2 \in \frac{1}{2B} \mathcal{B}_2(d)$. Then, we let $h(\boldsymbol{\xi})$ be a probability density with support on $\frac{1}{2B} \mathcal{B}_2(d)$ and define

$$\widetilde{M}_t = \int_{\boldsymbol{\xi}} M_t(\boldsymbol{\xi}) dh(\boldsymbol{\xi})$$

for all $t \geq 1$. Lemma 20.3 of Lattimore and Szepesvári (2020) shows that \widetilde{M}_t is also a non-negative super-martingale and $\mathbb{E}[\widetilde{M}_0] = 1$. Then, the maximal inequality (Theorem 3.9 of Lattimore and Szepesvári (2020)) shows that

$$\Pr \left[\sup_{t \geq 0} \log(\widetilde{M}_t) \geq \log(1/\delta) \right] \leq \delta. \quad (24)$$

Next, let $h(\boldsymbol{\xi})$ be the density of a normal distribution with the precision matrix $2\mathbf{H}_0$ truncated on $\frac{1}{2B} \mathcal{B}_2(d)$ and $N(h)$ be its normalization constant. Then, we can show that

$$\widetilde{M}_t = \frac{1}{N(h)} \int_{\frac{1}{2B} \mathcal{B}_2(d)} \exp(\boldsymbol{\xi}^\top \mathbf{S}_t - \|\boldsymbol{\xi}\|_{\tilde{\mathbf{H}}_t}^2) d\boldsymbol{\xi}.$$

Additionally, let $g(\boldsymbol{\xi})$ be the density of a normal distribution with the precision matrix $2\mathbf{H}_t$ truncated on $\frac{1}{4B} \mathcal{B}_2(d)$ and $N(g)$ be its normalization constant. Following the arguments in the proof of Theorem 1 in Faury et al. (2020), for any $t \geq 1$, one can show that

$$\widetilde{M}_t \geq \exp(\boldsymbol{\xi}^\top \mathbf{S}_t - \|\boldsymbol{\xi}\|_{\tilde{\mathbf{H}}_t}^2) \cdot \frac{N(g)}{N(h)}$$

for any $\boldsymbol{\xi} \in \frac{1}{4B} \mathcal{B}_2(d)$. Let $\boldsymbol{\xi}_0 = \frac{\mathbf{H}_t^{-1} \mathbf{S}_t}{\|\mathbf{S}_t\|_{\mathbf{H}_t^{-1}}} \cdot \frac{\sqrt{\lambda}}{4}$ which satisfies $\|\boldsymbol{\xi}_0\| \leq 1/4$. Then, we can write

$$\log(\widetilde{M}_t) \geq \boldsymbol{\xi}_0^\top \mathbf{S}_t - \|\boldsymbol{\xi}_0\|_{\tilde{\mathbf{H}}_t}^2 + \log \left(\frac{N(g)}{N(h)} \right) = \frac{\sqrt{\lambda}}{4} \|\mathbf{S}_t\|_{\mathbf{H}_t^{-1}} - \frac{\lambda}{16} + \log \left(\frac{N(g)}{N(h)} \right). \quad (25)$$

Combining (24) and (25), for any $t \geq 1$, we have

$$\Pr \left[\|\mathbf{S}_t\|_{\mathbf{H}_t^{-1}} \leq \frac{\sqrt{\lambda}}{4} + \frac{4}{\sqrt{\lambda}} \log \left(\frac{N(h)}{\delta N(g)} \right) \right] \geq 1 - \delta.$$

Using Lemma 6 of Faury et al. (2020), we can write

$$\log \left(\frac{N(h)}{N(g)} \right) \leq \log \left(\frac{\det(\mathbf{H}_t)^{1/2}}{\lambda^{d/2}} \right) + d \log(2)$$

□

Lemma D.2. For all $\xi \in \frac{1}{2B}\mathcal{B}(d)$, the process $\{M_t(\xi)\}_{t=1}^\infty$ is a non-negative super-martingale.

Proof. To show that $\{M_t(\xi)\}_{t=1}^\infty$ is a non-negative super-martingale, it is sufficient to show that $\mathbb{E}[M_{t+1}(\xi)|\mathcal{F}_t] \leq M_t(\xi)$ for all $t \geq 1$ and $\xi \in \frac{1}{2B}\mathcal{B}(d)$. We have

$$\begin{aligned} \mathbb{E}[M_{t+1}(\xi)|\mathcal{F}_t] &= \mathbb{E}[\exp(\xi^\top \mathbf{S}_{t+1} - \|\xi\|_{\mathbf{H}_{t+1}}^2)|\mathcal{F}_t] \\ &= \mathbb{E}[\exp(\xi^\top \mathbf{X}_t \boldsymbol{\epsilon}_t - \xi^\top \mathbf{X}_t \Sigma_t \mathbf{X}_t^\top \xi)|\mathcal{F}_t] M_t(\xi) \\ &= \mathbb{E}[\exp(\xi^\top \mathbf{X}_t \boldsymbol{\epsilon}_t)|\mathcal{F}_t] \exp(-\xi^\top \mathbf{X}_t \Sigma_t \mathbf{X}_t^\top \xi) M_t(\xi) \end{aligned}$$

Using Hölder's inequality, we can check that

$$|\xi^\top \mathbf{X}_t \boldsymbol{\epsilon}_t| \leq \|\boldsymbol{\epsilon}_t\|_1 \|\mathbf{X}_t^\top \xi\|_\infty = \|\boldsymbol{\epsilon}_t\|_1 \max_{i \in [K]} |\mathbf{x}_{ti}^\top \xi| \leq 1.$$

Therefore, we can use Lemma 6 from Amani and Thrampoulidis (2021), to write

$$\mathbb{E}[\exp(\xi^\top \mathbf{X}_t \boldsymbol{\epsilon}_t)|\mathcal{F}_t] \leq \exp(\xi^\top \mathbf{X}_t \Sigma_t \mathbf{X}_t^\top \xi).$$

Consequently, we can show that $\mathbb{E}[M_{t+1}(\xi)|\mathcal{F}_t] \leq M_t(\xi)$ and complete the proof. □

E Importance of Estimating Fisher Information Matrix

Oh and Iyengar (2021) has the best regret rate among efficient contextual MNL bandit algorithms in the literature. Their analysis shows that their algorithm has a regret rate of $\tilde{\mathcal{O}}(\kappa d \sqrt{T})$ where

$$\kappa := \left(\min_{t,i} \inf_{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq 1} q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) \right)^{-1}.$$

As we will show in the following proof, computation of this parameter for our setting results in $\kappa = \mathcal{O}(K^{2+1/L_0})$ which translates into a $\tilde{\mathcal{O}}(K^{2+1/L_0} d \sqrt{T})$ regret bound.

Proof. We can write $q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta})$ as

$$q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) = \frac{\exp(\langle \boldsymbol{\psi}_t, \mathbf{x}_{ti} \rangle - \langle \boldsymbol{\phi}_t, \mathbf{x}_{ti} \rangle p_{ti})}{\left(1 + \sum_{j \in S_t} \exp(\langle \boldsymbol{\psi}_t, \mathbf{x}_{tj} \rangle - \langle \boldsymbol{\phi}_t, \mathbf{x}_{tj} \rangle p_{tj}) \right)^2}$$

Since optimum prices lie in the interval $[0, P]$, and we have $\|\boldsymbol{\theta}_t\| \leq 2$ and $\|\mathbf{x}_{ti}\| \leq 1$, we can show that

$$q_{ti}(\boldsymbol{\theta}) q_{t0}(\boldsymbol{\theta}) > \frac{e^{-6-6P}}{(1 + K e^{2+2P})^2} > \frac{e^{-2-2P}}{(K+1)^2}.$$

Therefore, we can conclude that $\kappa = \mathcal{O}(K^2 e^P) = \mathcal{O}(K^{2+\frac{1}{L_0}})$.

□

Furthermore, we can show that $\kappa > (K-1)^{\frac{1}{(1+\epsilon)L_0}}$ for any $\epsilon > 0$. Therefore, the K^{1/L_0} dependency cannot be avoided for the regret of the algorithm provided in Oh and Iyengar (2021).

Proof. Assume that $N = K \geq 2$. Consider $u_{t1}(p) = 1 - p$ and $u_{ti}(p) = 1 - L_0 p$ for $i \in [K] \setminus \{1\}$. Under this construction, we will show a lower bound for κ .

For $i \in [K] \setminus \{1\}$, we have

$$v_{ti}(B) = \frac{1}{L_0} e^{-L_0 B}.$$

and for $i = 1$, we have

$$v_{t1}(B) = e^{-B}.$$

Then, following Proposition 3.2, if we let B_t be the unique solution of the fixed point equation

$$B = \frac{K-1}{L_0} e^{-L_0 B} + e^{-B}, \quad (26)$$

the optimum prices are given by $p_{t1}^* = B_t + 1$ and $p_{ti}^* = B_t + 1/L_0$ for $i \in [K] \setminus \{1\}$.

From Lemma C.1, we have the upper bound $B_t \leq P_0$. Next, we'll show a lower bound for B_t .

The right hand side (RHS) of (26) is decreasing and its left hand side (LHS) is increasing in B . Therefore, if we let B_ℓ be the solution of the fixed point equation

$$B = \frac{K-1}{L_0} e^{-L_0 B}, \quad (27)$$

then we have $B_\ell \leq B_t$. In (27), the LHS is increasing and the RHS is decreasing in B . Additionally, for $B = \log(K-1)/((1+\epsilon)L_0)$ with any $\epsilon > 0$ and large enough K , the LHS of (27) is smaller than its RHS. Hence, the fixed point satisfies

$$P_\ell := \frac{\log(K-1)}{(1+\epsilon)L_0} \leq B_\ell \leq B_t.$$

Now, we can write $q_{t1}(\boldsymbol{\theta}^*) q_{t0}(\boldsymbol{\theta}^*)$ as

$$\begin{aligned} q_{t1}(\boldsymbol{\theta}^*) q_{t0}(\boldsymbol{\theta}^*) &= \frac{\exp(1 - (B_t + 1))}{\left(1 + (K-1) \exp\left(1 - L_0(B_t + \frac{1}{L_0})\right) + \exp(1 - (B_t + 1))\right)^2} \\ &= \frac{e^{-B_t}}{\left(1 + (K-1)e^{-L_0 B_t} + e^{-B_t}\right)^2}. \end{aligned}$$

From this expression, we can show that

$$q_{t1}(\boldsymbol{\theta}^*) q_{t0}(\boldsymbol{\theta}^*) < e^{-P_\ell}$$

Therefore, we can conclude that $\kappa > (K-1)^{\frac{1}{(1+\epsilon)L_0}}$.

□

F Proof of Theorem 4.4 (Regret Upper Bound for Algorithm 3)

Similar to Algorithm 2, we run T_0 initialization rounds with random assortment and price selections to obtain an initial pilot estimate $\boldsymbol{\theta}_0 := \widehat{\boldsymbol{\theta}}_{T_0}$. Using the results of Lemma C.2 and Lemma C.5, we can show that the conditions $\lambda_{\min}(\mathbf{V}_{T_0}) \geq \lambda_{\min}^0$ and $\widehat{\boldsymbol{\theta}}_t \in \mathcal{B}_{\gamma/2}$ for all $t \geq T_0$ are satisfied with probability $1 - \mathcal{O}(T^{-1})$ if we select

$$T_0 = \Theta\left(\frac{\lambda_{\min}^0 K}{\sigma_0}\right) = \Theta\left(\frac{d\bar{P}^2 K \log^3(T)}{\sigma_0^3}\right). \quad (28)$$

Then, we apply the following parameter update at each time step t :

$$\hat{\boldsymbol{\theta}}_t = \underset{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \gamma/2}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{t-1}\|_{\mathbf{V}_t}^2 + 4(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{t-1})^\top g_t(\hat{\boldsymbol{\theta}}_{t-1}) \right\} \quad (29)$$

which directly ensures that $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \leq \gamma$ for all $t \geq T_0$ with probability $1 - \mathcal{O}(T^{-2})$.

Then, using this update rule, we modify our algorithm and present it in the algorithm block for Algorithm 3. Algorithm 2 requires $\Theta(tK)$ computational complexity to compute the MLE estimate in each round t . Since this cost grows linearly with each round t , the overall amortized computational cost turns out to be $\Theta(TK)$. On the other hand, the parameter update in Algorithm 3, only the $\Theta(K)$ context vector in the last offered assortment is needed per each round.

Algorithm 3 CAP-ONS: CAP with online Newton steps

- 1: **Input:** initialization rounds T_0 , confidence parameters $\{\alpha_t\}_{t \in [T]}$, minimum price sensitivity L_0
 - 2: $\mathbf{V}_0 \leftarrow \mathbf{0} \in \mathbb{R}^{2d \times 2d}$
 - 3: **for** $t = 1, 2, \dots, T_0$ **do** ▷ initialization rounds
 - 4: Choose S_t uniformly at random from $\{S \subseteq [N] : |S| \leq K\}$
 - 5: Choose p_{ti} independently and uniformly at random from $[1, 2]$ for all $i \in S_t$
 - 6: Offer assortment S_t at price \mathbf{p}_t and observe i_t
 - 7: $\mathbf{V}_{t+1} \leftarrow \mathbf{V}_t + \frac{1}{K^2} \sum_{i \in S_t} \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top$
 - 8: Compute MLE $\hat{\boldsymbol{\theta}}_{T_0}$ by solving (7) and set $\boldsymbol{\theta}_0 = \hat{\boldsymbol{\theta}}_{T_0}$.
 - 9: **for** $t = T_0 + 1, T_0 + 2, \dots, T$ **do**
 - 10: Compute $\hat{\boldsymbol{\theta}}_t$ by solving (29)
 - 11: Let $g_{ti}(p) := \alpha_t^{\text{OL}} \|(\mathbf{x}_{ti}, -p\mathbf{x}_{ti})\|_{\mathbf{V}_t^{-1}}$ for all $i \in [n]$ ▷ Price-dependent confidence function
 - 12: Let $\tilde{h}_{ti}^{\text{OL}}(p) := \langle \hat{\boldsymbol{\psi}}_t, \mathbf{x}_{ti} \rangle - \langle \hat{\boldsymbol{\phi}}_t, \mathbf{x}_{ti} \rangle \cdot p + g_{ti}(p)$ for all $i \in [n]$
 - 13: Let $h_{ti}^{\text{OL}}(p) := \min_{p' \leq p} \left\{ \tilde{h}_{ti}^{\text{OL}}(p') - L_0(p - p') \right\}$ for all $i \in [n]$ ▷ Utility function estimate
 - 14: Choose (S_t, \mathbf{p}_t) using Algorithm 1 with estimated utility functions $h_{ti}^{\text{OL}}(p)$
 - 15: Offer assortment S_t at price \mathbf{p}_t and observe i_t
 - 16: $\mathbf{V}_{t+1} \leftarrow \mathbf{V}_t + \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{tj}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top$ ▷ Information estimate
-

To analyze the regret of Algorithm 3, we first define a per-round negative log-likelihood function $f_t(\boldsymbol{\theta})$ and its gradient $\nabla_{\boldsymbol{\theta}} f_t(\boldsymbol{\theta})$ as

$$f_t(\boldsymbol{\theta}) = -q_{ti_t}(\boldsymbol{\theta})$$

$$g_t(\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} f_t(\boldsymbol{\theta}) = \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}) \tilde{\mathbf{x}}_{ti} - \tilde{\mathbf{x}}_{ti_t}.$$

We note that negative log-likelihood $f_t(\boldsymbol{\theta})$ for MNL model at each round t is a strongly convex function over a bounded domain, which enables us to apply a variant of online Newton updates (Hazan et al., 2014) that was also used in Hazan et al. (2014); Zhang et al. (2016); Oh and Iyengar (2021) which proposed online algorithms for logistic models.

To prove the regret rate for our algorithm with online parameter updates, we construct a new confidence region using a new confidence radius α_t^{OL} specified in the following lemma. Then, the utility function upper-bound estimate $h_{ti}^{\text{OL}}(p)$ is also modified accordingly.

Lemma F.1. *Let T_0 be any round such that $\lambda_{\min}(\mathbf{V}_{T_0}) \geq KP^2$. Then, for any $t > T_0$, we have $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq \alpha_t^{\text{OL}}$ with probability at least $1 - t^{-2}$ for confidence radius*

$$\alpha_t^{\text{OL}} = \sqrt{\gamma^2 T_0 + 576dK \log(T/d) + 16 \log \left(\frac{\lceil \log(K\gamma t^2/\omega) \rceil t^2}{\delta} \right)} + 8. \quad (30)$$

where $\omega := \min_{\boldsymbol{\theta} \in \mathcal{B}_\gamma} q_{t0}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t)$ and satisfies $1/\omega = \mathcal{O}(K^{2+1/L_0})$.

Then, similar to the proof of Theorem 4.2, we define a *good* event $\tilde{\mathcal{E}}_t = \{\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t} \leq \alpha_t^{\text{OL}}\}$ for $t \geq T_0$ that holds with probability at least $1 - t^{-2}$. Consequently, following steps similar to the proof of Theorem 4.2, we can write the regret as

$$\mathcal{R}_T \leq PT_0 + 4P\alpha_T^{\text{OL}} \sqrt{T \sum_{t=T_0}^T \sum_{i \in S_t} q_{ti}(\boldsymbol{\theta}^*) \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_t^{-1}}^2} + \mathcal{O}(P).$$

for α_T^{OL} given in Lemma F.1. Finally, using Lemma C.11 and Lemma C.12, we show that

$$\mathcal{R}_T \leq PT_0 + 29P\alpha_T^{\text{OL}} \sqrt{dKT \log(T/d)} + \mathcal{O}(P).$$

Note that $\alpha_T^{\text{OL}} = \tilde{\mathcal{O}}(\sqrt{dK})$ for the selection of T_0 given in (28).

F.1 Proof of Lemma F.1

The proof of Lemma F.1 depends on a few technical results we present next. First, we define the matrix

$$\mathbf{W}_t = \mathbf{H}_t(\hat{\boldsymbol{\theta}}_t) = \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{tj}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top.$$

We start by showing that following bound holds true over $\mathcal{B}_\gamma := \{\boldsymbol{\theta} : \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq \gamma\}$.

Lemma F.2. *For any $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \mathcal{B}_\gamma$, we have*

$$f_t(\boldsymbol{\theta}_2) \geq f_t(\boldsymbol{\theta}_1) + g_t(\boldsymbol{\theta}_1)^\top (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) + \frac{1}{4} (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^\top \mathbf{W}_t (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1).$$

Proof. Using the Taylor's expansion, there exists some $c \in (0, 1)$ such that

$$f_t(\boldsymbol{\theta}_2) = f_t(\boldsymbol{\theta}_1) + g_t(\boldsymbol{\theta}_1)^\top (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1) + (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)^\top \mathbf{H}_t(\bar{\boldsymbol{\theta}}) (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)$$

where $\bar{\boldsymbol{\theta}} = c\boldsymbol{\theta}_2 + (1 - c)\boldsymbol{\theta}_1$ and $\mathbf{H}_t(\bar{\boldsymbol{\theta}})$ is the Hessian of f_t at $\bar{\boldsymbol{\theta}}$. Furthermore, by Lemma C.3, we have $\mathbf{H}_t(\bar{\boldsymbol{\theta}}) \succcurlyeq \frac{1}{4}\mathbf{H}_t(\boldsymbol{\theta}_t) = \frac{1}{4}\mathbf{W}_t$. Consequently, the result follows. \square

Next, we prove the following lemma that shows the dependency between the error $(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)$ at time t and the error $(\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*)$ at time $t + 1$.

Lemma F.3. *For any t ,*

$$2g_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \leq 4\|g_t(\hat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_{t+1}^{-1}}^2 + \frac{1}{4}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 - \frac{1}{4}\|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2.$$

Proof. Note that

$$\hat{\boldsymbol{\theta}}_{t+1} = \underset{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \gamma/2}{\operatorname{argmin}} \left\{ \frac{1}{2} \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_{t+1}}^2 + 4(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_t)^\top g_t(\hat{\boldsymbol{\theta}}_t) \right\}.$$

From the first-order optimality condition, we have

$$(4g_t(\hat{\boldsymbol{\theta}}_t) + \mathbf{V}_{t+1}(\hat{\boldsymbol{\theta}}_{t+1} - \hat{\boldsymbol{\theta}}_t))^\top (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{t+1}) \geq 0$$

for any $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \gamma/2$. We can rewrite this inequality as

$$\boldsymbol{\theta}^\top \mathbf{V}_{t+1}(\hat{\boldsymbol{\theta}}_{t+1} - \hat{\boldsymbol{\theta}}_t) \geq \hat{\boldsymbol{\theta}}_{t+1}^\top \mathbf{V}_{t+1}(\hat{\boldsymbol{\theta}}_{t+1} - \hat{\boldsymbol{\theta}}_t) - 4g_t(\hat{\boldsymbol{\theta}}_t)^\top (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{t+1}).$$

Then, we can write

$$\begin{aligned}
 & \|\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 - \|\widehat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 \\
 &= \widehat{\boldsymbol{\theta}}_t^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_t - \widehat{\boldsymbol{\theta}}_{t+1}^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_{t+1} + 2\boldsymbol{\theta}^{*\top} \mathbf{V}_{t+1} (\widehat{\boldsymbol{\theta}}_{t+1} - \widehat{\boldsymbol{\theta}}_t) \\
 &\geq \widehat{\boldsymbol{\theta}}_t^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_t - \widehat{\boldsymbol{\theta}}_{t+1}^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_{t+1} + 2\widehat{\boldsymbol{\theta}}_{t+1}^\top \mathbf{V}_{t+1} (\widehat{\boldsymbol{\theta}}_{t+1} - \widehat{\boldsymbol{\theta}}_t) - 8g_t(\widehat{\boldsymbol{\theta}}_t)^\top (\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{t+1}) \\
 &= \widehat{\boldsymbol{\theta}}_t^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_t + \widehat{\boldsymbol{\theta}}_{t+1}^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_{t+1} - 2\widehat{\boldsymbol{\theta}}_{t+1}^\top \mathbf{V}_{t+1} \widehat{\boldsymbol{\theta}}_t - 8g_t(\widehat{\boldsymbol{\theta}}_t)^\top (\boldsymbol{\theta}^* - \widehat{\boldsymbol{\theta}}_{t+1}) \\
 &= \|\widehat{\boldsymbol{\theta}}_{t+1} - \widehat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_{t+1}}^2 + 8g_t(\widehat{\boldsymbol{\theta}}_t)^\top (\widehat{\boldsymbol{\theta}}_{t+1} - \widehat{\boldsymbol{\theta}}_t) + 8g_t(\widehat{\boldsymbol{\theta}}_t)^\top (\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \\
 &\geq -16\|g_t(\widehat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_{t+1}^{-1}}^2 + 8g_t(\widehat{\boldsymbol{\theta}}_t)^\top (\widehat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)
 \end{aligned}$$

where the last inequality follows from

$$\begin{aligned}
 \|\widehat{\boldsymbol{\theta}}_{t+1} - \widehat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_{t+1}}^2 + 8g_t(\widehat{\boldsymbol{\theta}}_t)^\top (\widehat{\boldsymbol{\theta}}_{t+1} - \widehat{\boldsymbol{\theta}}_t) &\geq \min_{\boldsymbol{\theta}: \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\| \leq \gamma/2} \left\{ \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_{t+1}}^2 + 8(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t)^\top g_t(\widehat{\boldsymbol{\theta}}_t) \right\} \\
 &\geq \min_{\boldsymbol{\theta}} \left\{ \|\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t\|_{\mathbf{V}_{t+1}}^2 + 8(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}_t)^\top g_t(\widehat{\boldsymbol{\theta}}_t) \right\} \\
 &= \min_{\boldsymbol{\theta}} \left\{ \|\boldsymbol{\theta}\|_{\mathbf{V}_{t+1}}^2 + 8\boldsymbol{\theta}^\top g_t(\widehat{\boldsymbol{\theta}}_t) \right\} \\
 &= -16\|g_t(\widehat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_{t+1}^{-1}}^2
 \end{aligned}$$

□

Next, we let \mathcal{F}_t denote the filtration up to time t and define the conditional expected values for the per-round negative log-likelihood $f_t(\boldsymbol{\theta})$ and its gradient $g_t(\boldsymbol{\theta})$ as follows.

$$\begin{aligned}
 \bar{f}_t(\boldsymbol{\theta}) &= \mathbb{E}_{i_t}[f_t(\boldsymbol{\theta})|\mathcal{F}_t] \\
 \bar{g}_t(\boldsymbol{\theta}) &= \mathbb{E}_{i_t}[g_t(\boldsymbol{\theta})|\mathcal{F}_t].
 \end{aligned}$$

Lemma F.4. *For any positive definite matrix \mathbf{V} ,*

$$\|g_t(\boldsymbol{\theta})\|_{\mathbf{V}}^2 \leq 4 \max_{i \in S_t} \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}}^2.$$

Proof. Recall that y_{ti} is a binary variable such that $y_{ti} = 1$ if $i_t = i$ and $y_{ti} = 0$ otherwise. For convenience also denote $q_{ti} = q_t(i|S_t, \mathbf{p}_t; \boldsymbol{\theta})$. Then, we note that $\sum_{i \in S_t} q_{ti} \leq 1$ and $\sum_{i \in S_t} y_{ti} \leq 1$. Consequently, we can write

$$\begin{aligned}
 \|g_t(\boldsymbol{\theta})\|_{\mathbf{V}}^2 &= \sum_{i \in S_t} \sum_{j \in S_t} (q_{ti} - y_{ti})(q_{tj} - y_{tj}) \tilde{\mathbf{x}}_{ti}^\top \mathbf{V} \tilde{\mathbf{x}}_{tj} \\
 &\leq \sum_{i \in S_t} \sum_{j \in S_t} (q_{ti}q_{tj} + y_{ti}y_{tj} + q_{ti}y_{tj} + q_{tj}y_{ti}) |\tilde{\mathbf{x}}_{ti}^\top \mathbf{V} \tilde{\mathbf{x}}_{tj}| \\
 &\leq 4 \max_{i, j \in S_t} |\tilde{\mathbf{x}}_{ti}^\top \mathbf{V} \tilde{\mathbf{x}}_{tj}| \\
 &\leq 4 \max_{i \in S_t} |\tilde{\mathbf{x}}_{ti}^\top \mathbf{V} \tilde{\mathbf{x}}_{ti}| \\
 &= 4 \max_{i \in S_t} \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}}^2.
 \end{aligned}$$

□

Then, we show that $\bar{f}_t(\boldsymbol{\theta})$ is minimized at $\boldsymbol{\theta}^*$. Formally, we prove the following lemma.

Lemma F.5. *For any $\boldsymbol{\theta} \in \mathbb{R}^{2d}$, we have $\bar{f}_t(\boldsymbol{\theta}) \geq \bar{f}_t(\boldsymbol{\theta}^*)$.*

Proof. For any $\boldsymbol{\theta} \in \mathbb{R}^{2d}$,

$$\begin{aligned}\bar{f}_t(\boldsymbol{\theta}) - \bar{f}_t(\boldsymbol{\theta}^*) &= \sum_{i \in S_t} q_t(i|S_t, \mathbf{p}_t; \boldsymbol{\theta}^*) [\log q_t(i|S_t, \mathbf{p}_t; \boldsymbol{\theta}^*) - \log q_t(i|S_t, \mathbf{p}_t; \boldsymbol{\theta})] \\ &\geq 0\end{aligned}$$

since it is equal to the Kullback-Leibler (KL) divergence between distributions $q_t(i|S_t, \mathbf{p}_t; \boldsymbol{\theta}^*)$ and $q_t(i|S_t, \mathbf{p}_t; \boldsymbol{\theta})$. \square

Lemma F.6. Suppose $\hat{\boldsymbol{\theta}}_t \in \mathbb{B}_\gamma$ for all $t \geq T_0$. Then, with probability at least $1 - \delta$,

$$\sum_{\tau=T_0}^t \left(\bar{g}_\tau(\hat{\boldsymbol{\theta}}_\tau) - g_\tau(\hat{\boldsymbol{\theta}}_\tau) \right)^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*) \leq 2 \log \left(\frac{\lceil \log(K\gamma t^2/\omega) \rceil t^2}{\delta} \right) + \frac{1}{8} \sum_{\tau=T_0}^t \|\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*\|_{\mathbf{W}_\tau}^2 + 1.$$

where $\omega := \min_{\boldsymbol{\theta} \in \mathcal{B}_\gamma} q_{t0}(\hat{\boldsymbol{\theta}}_t)$ and satisfies $1/\omega = \mathcal{O}(K^{2+1/L_0})$.

Proof. The following proof is adapted from Lemma 14 of Oh and Iyengar (2021). Note that $\xi_t = (\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)$ is a martingale difference sequence and it satisfies

$$\begin{aligned} |(\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)| &\leq |\bar{g}_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)| + |g_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)| \\ &\leq 2\sqrt{2\gamma P} \\ &\leq 2\sqrt{2}.\end{aligned}$$

using the fact that $\|g_t(\hat{\boldsymbol{\theta}}_t)\| = \|\sum_{i \in S_t} (q_{ti}(\boldsymbol{\theta}) - y_{ti}) \tilde{\mathbf{x}}_{ti}\| \leq \sqrt{2P}$ for any $\boldsymbol{\theta}$. Therefore,

$$M_t := \sum_{\tau=T_0}^t \left(\bar{g}_\tau(\hat{\boldsymbol{\theta}}_\tau) - g_\tau(\hat{\boldsymbol{\theta}}_\tau) \right)^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)$$

is a martingale. Now, we notice that $\mathbb{E}_{i_t}[\xi_t^2 | \mathcal{F}_t] = (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^\top \mathbf{H}_t(\hat{\boldsymbol{\theta}}_t)(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) = \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2$ and define the random variable

$$B_t := \sum_{\tau=T_0}^t \mathbb{E}_{i_\tau}[\xi_\tau^2 | \mathcal{F}_\tau] = \sum_{\tau=T_0}^t \|\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*\|_{\mathbf{W}_\tau}^2.$$

In the following, we will show how we can use this quantity to upper-bound M_t . Since B_t is a random variable, it is not possible to apply Freedman's inequality Freedman (1975) directly to M_t . Instead, we consider two cases with (i) $B_t \leq \frac{\omega}{tK}$ and (ii) $B_t > \frac{\omega}{tK}$ where $\omega := \min_{\boldsymbol{\theta} \in \mathcal{B}_\gamma} q_{t0}(\hat{\boldsymbol{\theta}}_t)$ as introduced in Sect E.

Case (i): When $B_t \leq \frac{\omega}{tK}$, we have

$$\begin{aligned}
 M_t &= \sum_{\tau=T_0}^t \left(\bar{g}_\tau(\hat{\boldsymbol{\theta}}_\tau) - g_\tau(\hat{\boldsymbol{\theta}}_\tau) \right)^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*) \\
 &= \sum_{\tau=T_0}^t \sum_{i \in S_\tau} (y_{\tau i} - q_{ti}(\boldsymbol{\theta}^*)) \tilde{\mathbf{x}}_{\tau i}^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*) \\
 &\leq \sum_{\tau=T_0}^t \sum_{i \in S_\tau} |y_{\tau i} - q_{ti}(\boldsymbol{\theta}^*)| |\tilde{\mathbf{x}}_{\tau i}^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)| \\
 &\leq \sum_{\tau=T_0}^t \sum_{i \in S_\tau} |\tilde{\mathbf{x}}_{\tau i}^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)| \\
 &\leq \sqrt{tK \sum_{\tau=T_0}^t \sum_{i \in S_\tau} (\tilde{\mathbf{x}}_{\tau i}^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*))^2} \\
 &= \sqrt{tK \sum_{\tau=T_0}^t (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)^\top \left(\sum_{i \in S_\tau} \tilde{\mathbf{x}}_{\tau i} \tilde{\mathbf{x}}_{\tau i}^\top \right) (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)} \\
 &\leq \sqrt{\frac{tK}{\omega} \sum_{\tau=T_0}^t (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)^\top \mathbf{W}_\tau (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*)} \\
 &= \sqrt{\frac{tK}{\omega} B_t} \\
 &\leq 1
 \end{aligned}$$

where we defined and used the result that

$$\begin{aligned}
 \mathbf{W}_t &= \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top - \sum_{i \in S_t} \sum_{j \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{tj}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{tj}^\top \\
 &\succcurlyeq \sum_{i \in S_t} q_{ti}(\hat{\boldsymbol{\theta}}_t) q_{t0}(\hat{\boldsymbol{\theta}}_t) \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top \\
 &\succcurlyeq \omega \sum_{i \in S_t} \tilde{\mathbf{x}}_{ti} \tilde{\mathbf{x}}_{ti}^\top.
 \end{aligned}$$

Case (ii): When $B_t > \frac{\omega}{tK}$, we have both a lower and upper bound for B_t , i.e., $\frac{\omega}{tK} < B_t \leq \gamma t$ since $\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\| \leq \gamma$ and $\mathbf{W}_t \preccurlyeq 1$ for all t . Then, we let η_t to denote a constant and apply the peeling technique from Bartlett et al. (2005) to obtain

$$\begin{aligned}
 \Pr(M_t \geq \sqrt{\eta_t B_t}) &= \Pr(M_t \geq \sqrt{\eta_t B_t}, \frac{\omega}{tK} < B_t < \gamma t) \\
 &= \Pr(M_t \geq \sqrt{\eta_t B_t}, \frac{\omega}{tK} < B_t < \gamma t) \\
 &= \sum_{j=1}^m \Pr(M_t \geq \sqrt{\eta_t B_t}, \frac{2^{j-1}\omega}{tK} < B_t < \frac{2^j\omega}{tK}) \\
 &\leq \sum_{j=1}^m \Pr\left(M_t \geq \sqrt{\eta_t \frac{2 \cdot 2^j \omega}{tK}}, B_t < \frac{2^j \omega}{tK}\right) \\
 &\leq 2m \exp(-\eta_t)
 \end{aligned}$$

where we set $m = \lceil \log(K\gamma t^2/\omega) \rceil$ and use the Freedman's inequality Freedman (1975) for the last inequality.

Combining the results from both cases, letting $\eta_t = \log \frac{mt^2}{\delta}$, and taking a union bound over t , we have

$$\begin{aligned} M_t &\leq \sqrt{\eta_t B_t} + 1 \\ &\leq 2\eta_t + \frac{1}{8}B_t + 1 \end{aligned}$$

where the last step uses the inequality $uv \leq cu^2 + v^2/(4c)$. \square

Now, we prove Lemma F.1 by using the previous results. First, we note that $\hat{\boldsymbol{\theta}}_t, \boldsymbol{\theta}^* \in \mathcal{B}_\gamma$ for $t \geq T_0$ by construction. Then, we use Lemma F.2 to write

$$f_t(\hat{\boldsymbol{\theta}}_t) \leq f_t(\boldsymbol{\theta}^*) + g_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) - \frac{1}{4}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^\top \mathbf{W}_t(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*).$$

Then, by taking the expectation over i_t on both sides, we obtain

$$\bar{f}_t(\hat{\boldsymbol{\theta}}_t) \leq \bar{f}_t(\boldsymbol{\theta}^*) + \bar{g}_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) - \frac{1}{4}(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*)^\top \mathbf{W}_t(\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*).$$

Since $\bar{f}_t(\boldsymbol{\theta}) \geq \bar{f}_t(\boldsymbol{\theta}^*)$ by Lemma F.5, we have

$$\begin{aligned} 0 &\leq \bar{f}_t(\hat{\boldsymbol{\theta}}_t) - \bar{f}_t(\boldsymbol{\theta}^*) \\ &\leq \bar{g}_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) - \frac{1}{4}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2 \\ &= g_t(\hat{\boldsymbol{\theta}}_t)^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) - \frac{1}{4}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2 + (\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*). \end{aligned}$$

Then, using Lemma F.3 and Lemma F.4, we have

$$\begin{aligned} 0 &\leq 2\|g_t(\hat{\boldsymbol{\theta}}_t)\|_{\mathbf{V}_{t+1}^{-1}}^2 + \frac{1}{8}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 - \frac{1}{8}\|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 \\ &\quad - \frac{1}{4}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2 + (\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \\ &\leq 2 \max_{i \in S_t} \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_{t+1}^{-1}}^2 + \frac{1}{8}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 - \frac{1}{8}\|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 \\ &\quad - \frac{1}{4}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2 + (\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \\ &= 2 \max_{i \in S_t} \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_{t+1}^{-1}}^2 + \frac{1}{8}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t}^2 - \frac{1}{8}\|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 \\ &\quad - \frac{1}{8}\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2 + (\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*) \end{aligned}$$

where the last equality follows by noting that we have

$$\|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 = \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t}^2 + \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2$$

since $\mathbf{V}_{t+1} = \mathbf{V}_t + \mathbf{W}_t$.

Hence, we have

$$\begin{aligned} \|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 &\leq \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_t}^2 + 16 \max_{i \in S_t} \|\tilde{\mathbf{x}}_{ti}\|_{\mathbf{V}_{t+1}^{-1}}^2 - \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{W}_t}^2 \\ &\quad + 8(\bar{g}_t(\hat{\boldsymbol{\theta}}_t) - g_t(\hat{\boldsymbol{\theta}}_t))^\top (\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*). \end{aligned}$$

Summing over $\{T_0, \dots, t\}$, we obtain

$$\begin{aligned} \|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 &\leq \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{T_0}}^2 + 16 \sum_{\tau=T_0}^t \max_{i \in S_\tau} \|\tilde{\mathbf{x}}_{\tau i}\|_{\mathbf{V}_{\tau+1}^{-1}}^2 - \sum_{\tau=T_0}^t \|\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*\|_{\mathbf{W}_\tau}^2 \\ &\quad + 8 \sum_{\tau=T_0}^t (\bar{g}_\tau(\hat{\boldsymbol{\theta}}_\tau) - g_\tau(\hat{\boldsymbol{\theta}}_\tau))^\top (\hat{\boldsymbol{\theta}}_\tau - \boldsymbol{\theta}^*). \end{aligned}$$

Then, Lemma F.6 shows with a probability at least $1 - \delta$,

$$\begin{aligned} & \|\hat{\boldsymbol{\theta}}_{t+1} - \boldsymbol{\theta}^*\|_{\mathbf{V}_{t+1}}^2 \\ & \leq \|\hat{\boldsymbol{\theta}}_t - \boldsymbol{\theta}^*\|_{\mathbf{V}_{T_0}}^2 + 16 \sum_{\tau=T_0}^t \max_{i \in S_\tau} \|\tilde{\mathbf{x}}_{\tau i}\|_{\mathbf{V}_{\tau+1}^{-1}}^2 + 16 \log \left(\frac{\lceil \log(K\gamma t^2/\omega) \rceil t^2}{\delta} \right) + 8 \\ & \leq \gamma^2 \lambda_{\max}(\mathbf{V}_{T_0}) + 576dK \log(T/d) + 16 \log \left(\frac{\lceil \log(K\gamma t^2/\omega) \rceil t^2}{\delta} \right) + 8 \\ & \leq \gamma^2 T_0 + 576dK \log(T/d) + 16 \log \left(\frac{\lceil \log(K\gamma t^2/\omega) \rceil t^2}{\delta} \right) + 8 \end{aligned}$$

where we apply Lemma C.11 for the last step.

G Proof of Theorem 4.5

At a high level, we prove Theorem 4.5 in three steps. In the first step, we construct an adversarial set of parameters and reduce the task of lower bounding the worst-case regret of any policy to lower bounding the Bayes risk over the constructed parameter set. In the second step, we use a counting argument similar to the one used in Chen and Wang (2018) and Chen et al. (2020) to provide an explicit lower bound on the Bayes risk of the constructed adversarial parameter set. Finally, we apply Pinsker's inequality to complete the proof. The following sections provide the details for each of these steps.

G.1 Adversarial construction and the Bayes risk

Let $\epsilon \in (0, (1 - L_0^2)/d\sqrt{d})$ be a small positive parameter to be specified later. For every subset $W \subseteq [d]$, define the corresponding parameter $\boldsymbol{\psi}_W \in \mathbb{R}^d$ as $[\boldsymbol{\psi}_W]_i = \epsilon$ for all $i \in W$, and $[\boldsymbol{\psi}_W]_i = 0$ for all $i \notin W$. Next, define $\boldsymbol{\phi}^* \in \mathbb{R}^d$ as $[\boldsymbol{\phi}^*]_i = L_0 \sqrt{1/d}$ for all $i \in [d]$. Finally, for any $W \subseteq [d]$, define the concatenated parameter vectors $\boldsymbol{\theta}_W \in \mathbb{R}^{2d}$ as $\boldsymbol{\theta}_W = (\boldsymbol{\psi}_W, \boldsymbol{\phi}^*)$. The parameter set that we consider is

$$\Theta := \{\boldsymbol{\theta}_W : W \in \mathcal{W}_{d/4}\}$$

where $\mathcal{W}_{d/4} := \{W \subseteq [d] : |W| = d/4\}$ denotes the set of all subsets of $[d]$ whose size is $d/4$. Note that $d/4$ is a positive integer because d is divisible by 4. It is also easy to check that with the condition $\epsilon \in (0, (1 - L_0^2)/\sqrt{d})$, we satisfy $\|\boldsymbol{\theta}\| \leq 1$ for any $\boldsymbol{\theta} \in \Theta$.

The feature vectors $\{\mathbf{x}_{ti}\}$ are constructed to be invariant over time iterations t . For each t and $U \in \mathcal{W}_{d/4}$, K identical feature vectors \mathbf{x}_U are constructed as $[\mathbf{x}_U]_i = 2/\sqrt{d}$ for all $i \in U$, and $[\mathbf{x}_U]_i = 0$ for all $i \notin U$. Furthermore, it is straightforward to verify that $\|\mathbf{x}_U\| \leq 1$ for any $U \in \mathcal{W}_{d/4}$.

Hence, the worst-case regret of any policy π can be lower bounded by the worst-case regret of parameters belonging to Θ , which can be further lower bounded by the average regret over a uniform prior over Θ . Formally,

$$\sup_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}}^{\pi} \sum_{t=1}^T R(S_{\boldsymbol{\theta}}^*, \mathbf{p}_{\boldsymbol{\theta}}^*) - R(S_t, \mathbf{p}_t) = \max_{\boldsymbol{\theta} \in \Theta} \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}}^{\pi} \sum_{t=1}^T R(S_{\boldsymbol{\theta}}^*, \mathbf{p}_{\boldsymbol{\theta}}^*) - R(S_t, \mathbf{p}_t) \quad (31)$$

$$= \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}_W}^{\pi} R(S_{\boldsymbol{\theta}_W}^*, \mathbf{p}_{\boldsymbol{\theta}_W}^*) - R(S_t, \mathbf{p}_t) \quad (32)$$

Here, the $R(\cdot)$ function refers to the expected revenue function $R_t(\cdot)$ defined in (1). Since both the context vectors and the feature vectors are invariant over time by construction, we drop the time subscript t to simplify the notation. Additionally, $S_{\boldsymbol{\theta}_W}^*$ and $\mathbf{p}_{\boldsymbol{\theta}_W}^*$ refer to the optimal size- K assortment and pricing that maximizes expected revenue under the feature parameter $\boldsymbol{\theta}_W$. By construction, it is easy to verify that $S_{\boldsymbol{\theta}_W}^*$ consists of all K items corresponding to feature \mathbf{x}_W .

For any fixed assortment $S \in \mathcal{S}_K$, let $\mathbf{p}^*(S)$ denote the revenue-maximizing price vector to offer with assortment S . That is,

$$\mathbf{p}^*(S) \in \max_{\mathbf{p} \in \mathbb{R}_+^n} R(S, \mathbf{p})$$

with entries $p_i^*(S)$. Then, the optimum prices $\mathbf{p}_{\theta_W}^* = \mathbf{p}^*(S_{\theta_W}^*)$ can be characterized using the following proposition which is a special case of the Proposition 3.2.

Proposition G.1. *Consider that items in an assortment S of size K have utility functions $u_i(p) = \alpha_i - \beta_i \cdot p$. Then, the revenue-maximizing prices for offering assortment S are given by*

$$p_i^*(S) = \frac{1}{\beta_i} + B^0(S)$$

where $B^0(S)$ is the unique fixed point solution B of the equation

$$B = \sum_{i \in S} \frac{1}{\beta} e^{\alpha_i - \beta_i B - 1}.$$

Furthermore, the revenue achieved by offering $(S, \mathbf{p}^*(S))$ is equal to $B^0(S)$.

In particular, if all items in an assortment S have the same utility function $u_i(p) = \alpha - \beta \cdot p$, then we can write $B^0(S)$ as the fixed point solution of

$$B = \frac{K}{\beta} e^{\alpha - \beta B - 1}.$$

G.2 The counting argument

In this section, we derive an explicit lower bound on the Bayes risk in (32). For any sequence $\{(S_t, \mathbf{p}_t)\}_{t=1}^T$ produced by the policy π , we first describe an alternative sequence $\{(\tilde{S}_t, \tilde{\mathbf{p}}_t)\}_{t=1}^T$ that provably enjoys less regret under the feature parameter θ_W .

Let $\{\mathbf{x}_{U_1}, \dots, \mathbf{x}_{U_M}\}$ be the set of context vectors of items contained in assortment S_t (if $S_t = \emptyset$, then choose an arbitrary feature vector \mathbf{x}_U). Let \tilde{U}_t be the subset among U_1, \dots, U_M that maximizes $\langle \mathbf{x}_{\tilde{U}_t}, \psi_W \rangle$, where $\theta_W = (\psi_W, \phi^*)$ is the underlying parameter. Let \tilde{S}_t be the assortment consisting of all K items corresponding to the feature $\mathbf{x}_{\tilde{U}_t}$ and let $\tilde{\mathbf{p}}_t = \mathbf{p}^*(\tilde{S}_t)$ be the optimum prices for assortment \tilde{S}_t according to Proposition G.1. Then, the following lemma holds true.

Lemma G.2. $R(S_t, \mathbf{p}_t) \leq R(\tilde{S}_t, \tilde{\mathbf{p}}_t)$ for feature parameter $\theta_W = (\psi_W, \phi^*)$.

Proof. First, from the optimality of prices $\mathbf{p}^*(S_t)$ under S_t , we have $R(S_t, \mathbf{p}_t) \leq R(S_t, \mathbf{p}^*(S_t))$. Then, by Proposition G.1, $R(S_t, \mathbf{p}^*(S_t))$ is equal to the unique fixed point solution for

$$B = \sum_{i \in S} \frac{1}{\beta} e^{\alpha_i - \beta_i B - 1}.$$

Note that the expression on the right-hand side of this equation is monotonically increasing in each α_i . Therefore, by replacing all $i \in S_t$ with $i \in \tilde{S}_t$, the α_i values do not decrease and therefore the fixed point does not increase. That is, the fixed-point solution for

$$B = \sum_{i \in \tilde{S}_t} \frac{1}{\beta} e^{\alpha_i - \beta_i B - 1}. \quad (33)$$

is greater than or equal to $R(S_t, \mathbf{p}^*(S_t))$. Since the unique fixed point solution of (33) is equal to $R(\tilde{S}_t, \tilde{\mathbf{p}}_t)$, we have $R(S_t, \mathbf{p}^*(S_t)) \leq R(\tilde{S}_t, \tilde{\mathbf{p}}_t)$, completing the proof. \square

To simplify notation, we use \mathbb{E}_W to denote the expectations under parameter θ_W and policy π . The following lemma gives a lower bound for $R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(S_t, \tilde{\mathbf{p}}_t)$.

Lemma G.3. *Suppose $\epsilon \in (0, 1/d\sqrt{d})$ and define $\delta := d/4 - |\tilde{U}_t \cap W|$. Then,*

$$R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) \geq \frac{\delta\epsilon}{15L_0\sqrt{d}}$$

Define random variables $\tilde{N}_i := \sum_{t=1}^T \mathbb{1}\{i \in \tilde{U}_t\}$. Lemma G.3 immediately implies

$$\mathbb{E}_W \left[R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) \right] \geq \frac{\epsilon}{15L_0\sqrt{d}} \left(\frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i] \right), \forall W \in \mathcal{W}_{d/4}.$$

Summing both sides of this equation over all $W \in \mathcal{W}_{d/4}$ gives

$$\sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \left[R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) \right] \geq \frac{\epsilon}{15L_0\sqrt{d}} \sum_{W \in \mathcal{W}_{d/4}} \left(\frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i] \right).$$

Next, we will upper-bound the term $\sum_{W \in \mathcal{W}_{d/4}} \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i]$. First, define

$$\mathcal{W}_{d/4}^{(i)} := \{W \in \mathcal{W}_{d/4} : i \in W\}.$$

Then, we swap the order of summation to write

$$\begin{aligned} \sum_{W \in \mathcal{W}_{d/4}} \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i] &= \sum_{i \in [d]} \sum_{W \in \mathcal{W}_{d/4}^{(i)}} \mathbb{E}_W[\tilde{N}_i] \\ &= \sum_{i \in [d]} \sum_{W \in \mathcal{W}_{d/4-1}} \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] \\ &\leq |\mathcal{W}_{d/4-1}| \max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \in [d]} \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] \\ &= |\mathcal{W}_{d/4-1}| \max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \in [d]} (\mathbb{E}_W[\tilde{N}_i] + \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i]) \\ &\leq |\mathcal{W}_{d/4-1}| \left[\max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \in [d]} (\mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i]) + \frac{dT}{4} \right] \end{aligned}$$

where the last step follows from the fact that $\sum_{i \in [d]} \mathbb{E}_W[\tilde{N}_i] \leq dT/4$ for any fixed $W \in \mathcal{W}_{d/4-1}$.

Next, we note that

$$\frac{|\mathcal{W}_{d/4-1}|}{|\mathcal{W}_{d/4}|} = \frac{\binom{d}{d/4-1}}{\binom{d}{d/4}} = \frac{d/4}{3d/4+1} \leq \frac{1}{3}$$

to write

$$\begin{aligned} \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \left[R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) \right] &\geq \frac{1}{|\mathcal{W}_{d/4}|} \frac{\epsilon}{15L_0\sqrt{d}} \sum_{W \in \mathcal{W}_{d/4}} \left(\frac{dT}{4} - \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i] \right) \\ &\geq \frac{\epsilon}{15L_0\sqrt{d}} \left(\frac{dT}{4} - \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \sum_{i \in W} \mathbb{E}_W[\tilde{N}_i] \right) \\ &\geq \frac{\epsilon}{45L_0\sqrt{d}} \left(\frac{dT}{2} - \max_{W \in \mathcal{W}_{d/4-1}} \sum_{i \in [d]} |\mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i]| \right) \end{aligned}$$

G.3 Pinsker's inequality

In this section, we upper bound $\left| \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i] \right|$ for any fixed $W \in \mathcal{W}_{d/4-1}$. Let \mathbb{P}_W and $\mathbb{P}_{W \cup \{i\}}$ to denote the probability law under parameter θ_W and $\theta_{W \cup \{i\}}$, respectively. Then,

$$\begin{aligned} \left| \mathbb{E}_{W \cup \{i\}}[\tilde{N}_i] - \mathbb{E}_W[\tilde{N}_i] \right| &\leq \sum_{n=0}^T n \cdot \left| \mathbb{P}_W[\tilde{N}_i = n] - \mathbb{P}_{W \cup \{i\}}[\tilde{N}_i = n] \right| \\ &\leq T \cdot \sum_{n=0}^T \left| \mathbb{P}_W[\tilde{N}_i = n] - \mathbb{P}_{W \cup \{i\}}[\tilde{N}_i = n] \right| \\ &\leq 2T \cdot \|\mathbb{P}_W - \mathbb{P}_{W \cup \{i\}}\|_{\text{TV}} \\ &\leq T \sqrt{2 \cdot \text{KL}(\mathbb{P}_W || \mathbb{P}_{W \cup \{i\}})} \end{aligned}$$

where $\|P - Q\|_{\text{TV}} = \sup_A |P(A) - Q(A)|$ is the total variation distance between laws P and Q ; $\text{KL}(P || Q) = \int (\log dP/dQ)dP$ is the Kullback-Leibler (KL) divergence between P and Q ; and the inequality $\|P - Q\|_{\text{TV}} \leq \sqrt{\frac{1}{2} \text{KL}(P || Q)}$ is the Pinsker's inequality.

Recall that $\{\mathbf{x}_{U_1}, \dots, \mathbf{x}_{U_M}\}$ denotes the set of context vectors of items contained in assortment S_t . Then, for every $i \in [d]$, define a new random variable $N_i := \frac{1}{K} \sum_{t=1}^T \sum_{j=1}^M \mathbf{1}\{i \in U_j\}$. The next lemma is used to upper bound the KL divergence term $\text{KL}(\mathbb{P}_W || \mathbb{P}_{W \cup \{i\}})$.

Lemma G.4 (Lemma 6 in Chen et al. (2020)). *For any $W \in \mathcal{W}_{d/4-1}$ and $i \in [d]$,*

$$\text{KL}(\mathbb{P}_W || \mathbb{P}_{W \cup \{i\}}) \leq C_{\text{KL}} \cdot \mathbb{E}_W[N_i] \cdot \epsilon^2/d$$

for some universal constant $C_{\text{KL}} > 0$.

Combining Lemma G.4 with the final result of the previous subsection, we obtain

$$\begin{aligned} \frac{1}{|\mathcal{W}_{d/4}|} \sum_{W \in \mathcal{W}_{d/4}} \mathbb{E}_W \left[R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) \right] \\ \geq \frac{\epsilon}{45L_0\sqrt{d}} \left(\frac{dT}{2} - T \sum_{i \in [d]} \sqrt{2C_{\text{KL}} \cdot \mathbb{E}_W[N_i] \cdot \epsilon^2/d} \right) \\ \geq \frac{\epsilon}{45L_0\sqrt{d}} \left(\frac{dT}{2} - T\epsilon \sqrt{2C_{\text{KL}} \sum_{i \in [d]} \mathbb{E}_W[N_i]} \right) \\ \geq \frac{\epsilon}{45L_0\sqrt{d}} \left(\frac{dT}{2} - T\epsilon \sqrt{C'_{\text{KL}}dT} \right) \end{aligned}$$

where $C'_{\text{KL}} = C_{\text{KL}}/2$. Setting $\epsilon = \sqrt{d/16C'_{\text{KL}}T} \in (0, (1 - L_0^2)/d\sqrt{d})$ for sufficiently large T , we obtain

$$\sup_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{x}, \boldsymbol{\theta}}^{\pi} \sum_{t=1}^T R(S_{\boldsymbol{\theta}}^*, \mathbf{p}_{\boldsymbol{\theta}}^*) - R(S_t, \mathbf{p}_t) \geq C_0 d\sqrt{T}/L_0$$

for some universal constant C_0 , completing the proof of the theorem.

G.4 Proofs for Technical Lemmas

Lemma G.3. *Suppose $\epsilon \in (0, 1/d\sqrt{d})$ and define $\delta := d/4 - |\tilde{U}_t \cap W|$. Then,*

$$R(S_{\theta_W}^*, \mathbf{p}_{\theta_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) \geq \frac{\delta\epsilon}{15L_0\sqrt{d}}$$

Proof. The optimum revenue from offering K identical items with utility functions $u(p) = \alpha - \beta p$ is equal to the unique fixed point solution B of the equation

$$B = \frac{K}{\beta} e^{\alpha - \beta B - 1}. \quad (34)$$

Using the product logarithm function $W(\cdot)$, we can express the optimum revenue as

$$\frac{W(e^{\alpha-1}K)}{\beta} \quad (35)$$

Let $f_K(x) := W(e^{x-1}K)$ and denote its first derivative with $f'_K(x)$ for any $K \geq 1$. Then, by Lemma G.5, there exists a constant $C_K < \frac{2}{3}f'_K(0)$ such that

$$f_K(0) + f'_K(0) \cdot x \leq f_K(x) \leq f_K(0) + f'_K(0) \cdot x + C_K \cdot x^2$$

for all $0 \leq x \leq 1$. For the remainder of this proof, let $\mathbf{x} = \mathbf{x}_W$, $\tilde{\mathbf{x}} = \mathbf{x}_{\tilde{U}_t}$, and $\boldsymbol{\theta} = \boldsymbol{\theta}_W$. Then, we can write

$$R(S_{\boldsymbol{\theta}_W}^*, \mathbf{p}_{\boldsymbol{\theta}_W}^*) = f_K(\mathbf{x}^\top \boldsymbol{\theta}) \quad \text{and} \quad R(\tilde{S}_t, \tilde{\mathbf{p}}_t) = f_K(\tilde{\mathbf{x}}^\top \boldsymbol{\theta}).$$

Putting it all together, we can show that

$$\begin{aligned} R(S_{\boldsymbol{\theta}_W}^*, \mathbf{p}_{\boldsymbol{\theta}_W}^*) - R(\tilde{S}_t, \tilde{\mathbf{p}}_t) &\geq \frac{1}{L_0} \left[(f_K(0) + f'_K(0)\mathbf{x}^\top \boldsymbol{\theta}) - (f_K(0) + f'_K(0)\tilde{\mathbf{x}}^\top \boldsymbol{\theta} + C_K(\tilde{\mathbf{x}}^\top \boldsymbol{\theta})^2) \right] \\ &= \frac{1}{L_0} \left[f'_K(0)(\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\theta} - C_K(\tilde{\mathbf{x}}^\top \boldsymbol{\theta})^2 \right] \\ &\geq \frac{f'_K(0)}{L_0} \left[(\mathbf{x} - \tilde{\mathbf{x}})^\top \boldsymbol{\theta} - \frac{2}{3}(\tilde{\mathbf{x}}^\top \boldsymbol{\theta})^2 \right] \\ &\geq \frac{f'_K(0)}{L_0} \left[\frac{\delta\epsilon}{\sqrt{d}} - \frac{2d\epsilon^2}{3} \right] \\ &\geq \frac{f'_K(0)\delta\epsilon}{3L_0\sqrt{d}} \end{aligned}$$

where the last three inequalities use the inequality $0 < f''(0) < f'_K(0)$, the definition of δ , and the inequality $d\epsilon^2 \leq \delta\epsilon/\sqrt{d}$ provided that $\epsilon \in (0, 1/d\sqrt{d})$. Lastly, noting that $f'_K(0) > 1/5$ by Lemma G.5 for any $K \geq 1$, we conclude the proof. \square

Lemma G.5. Let $f_K(x) := W(e^{x-1}K)$ and denote its first derivative with $f'_K(x)$. Then, for any $K \geq 1$,

- (a) $f'_K(x) > 1/5$ for all $0 \leq x \leq 1$, and
- (b) there exists a constant $C_K < \frac{2}{3}f'_K(0)$ such that

$$f_K(0) + f'_K(0) \cdot x \leq f_K(x) \leq f_K(0) + f'_K(0) \cdot x + C_K \cdot x^2$$

for all $0 \leq x \leq 1$.

Proof. Let $f''_K(x)$ and $f^{(3)}_K(x)$ denote the second and third derivatives of $f_K(x)$ respectively. Using the properties of the product logarithm function, it is easy to show that

$$f'_K(x) = \frac{f_K(x)}{1 + f_K(x)}, \quad f''_K(x) = \frac{f_K(x)}{(1 + f_K(x))^3}, \quad f^{(3)}_K(x) = \frac{(1 - 2f_K(x))f_K(x)}{(1 + f_K(x))^5}.$$

For any $K \geq 1$, $f_K(x)$ is a positive and increasing function of x . Hence, $\min_{0 \leq x \leq 1} f'_K(x) = f'_K(0)$. Furthermore, we can show that

$$\min_{K \geq 1} f'_K(0) = \min_{K \geq 1} \frac{W(K/e)}{1 + W(K/e)} = \frac{W(1/e)}{1 + W(1/e)} > 1/5$$

proving the first part of the lemma.

To prove the second part of the lemma, we use Taylor's Theorem to write

$$f_K(x) = f_K(0) + f'_K(0) \cdot x + \frac{f''_K(0)}{2} \cdot x^2 + R_K(\zeta; x)$$

$$R_K(\zeta; x) = \frac{f_K^{(3)}(\zeta)}{6} x^3$$

for some ζ between 0 and x . For any $K \geq 3$, we can easily show that $f_K(x) \geq 1/2$ for all $0 \leq x \leq 1$. Therefore, $R_K(\zeta; x) \leq 0$ for all $0 \leq \zeta \leq x \leq 1$ and we can set $C_K = f''_K(0)/2$ to satisfy the upper bound inequality.

On the other hand, for $K = 1$ and $K = 2$, we can numerically show that

$$\max_{0 \leq \zeta \leq 1} f_K^{(3)}(\zeta) = f_K^{(3)}(0).$$

and $f_K^{(3)}(0) \leq f''_K(0)$. Therefore, we have

$$R_K(\zeta; x) \leq \frac{f''_K(0)}{6} \cdot x^2$$

for all $0 \leq \zeta \leq x \leq 1$ when $K = 1$ or $K = 2$. As a result, we can set $C_K = 2f''_K(0)/3$ to satisfy the upper bound inequality.

Since $f''_K(0) < f'_K(0)$ for any $K \geq 1$, the selected constant C_K also satisfies $C_K < \frac{2}{3}f'_K(0)$.

□

H Experimental Details

We numerically evaluate our algorithms over 20 independently generated problem instances and provide our results in Figure 3. We run experiments with $n = 100$ items for various assortment sizes K and various numbers of feature dimensions d . In each instance, the parameter ψ^* is uniformly chosen from $\{\psi : \|\psi\|_2 = 1/2\}$. On the other hand, price sensitivity parameter ϕ^* is generated by independently drawing its entries from a uniform distribution over $[\sqrt{L_0}/\sqrt{d}, 1/\sqrt{2d}]$ for some parameter $L_0 > 0$. Each context vector \mathbf{x}_{ti} is generated by independently drawing its entries over $[\sqrt{L_0}/\sqrt{d}, 1/\sqrt{2d}]$. This construction ensures that we satisfy both Assumptions 3.1 and 4.1.

Figure 4 demonstrates that the regret of CAP algorithm follows a T^α dependency with an empirically observed slope of $\alpha \approx 0.5$. This result aligns with the theoretical regret rate of $O(\sqrt{T})$ we obtained in this work.

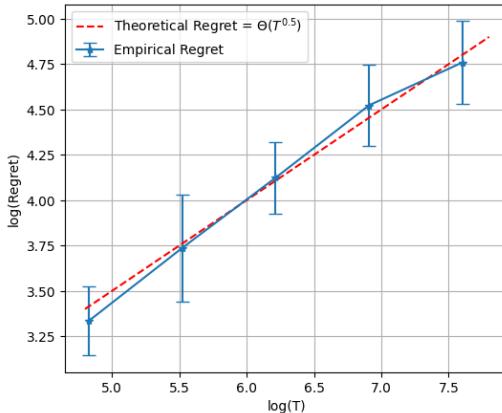


Figure 4: Log-log plot illustrating the dependency of regret for our proposed algorithm CAP. The slope of the curve reflects the empirical growth rate of regret with respect to time horizon T .