Assortment Optimization Under General Choice

Srikanth Jagabathula
Stern School of Business, New York University, New York, NY 10012, sjagabat@stern.nyu.edu

We consider the key operational problem of optimizing the mix of offered products to maximize revenues when product prices are exogenously set and product demand follows a general discrete choice model. The key challenge in making this decision is the computational difficulty of finding the best subset, which often requires exhaustive search. Existing approaches address the challenge by either deriving efficient algorithms for specific parametric structures or studying the performance of general-purpose heuristics. The former approach results in algorithms that lack portability to other structures; whereas the latter approach has resulted in algorithms with poor performance. We study a portable and easy-to-implement local search heuristic. We show that it efficiently finds the global optimum for the multinomial logit (MNL) model and derive performance guarantees for general choice structures. Empirically, it is better than prevailing heuristics when no efficient algorithms exist, and it is within 0.02% of optimality otherwise.

Key words: assortment optimization, local search, choice models, nonparametric choice

1. Introduction

A key decision faced by operations managers is determining the optimal mix of products to offer to customers. For instance, a retailer must decide which subset of products to offer in the limited shelf space in order to maximize sales or revenues. A publisher must decide the collection of ads to show in the limited screen real-estate to maximize clicks. An airline must decide which collection of fare classes to offer as a function of the number of unsold seats and the number of days left for take-off. In all these settings, the optimal assortment must trade off losing revenues from not including low-revenue products with gaining revenues from inducing the switch to high-revenue products.

To determine the optimal product mix, the firm must forecast the impact of the offered subset of products on demand. Traditionally, the literature in operations and revenue management has assumed the independent demand model: customers decide the product they want to purchase and, upon arriving, make the purchase (if the product is available) or leave without purchasing otherwise. The independent demand model results in tractable decision problems but fails to capture product substitution: if the preferred product is unavailable, the customer may substitute an available one. Incorporating substitution significantly improves the accuracy of demand predictions. The seminal work of Talluri and van Ryzin (2004a) has demonstrated the importance of incorporating choice into revenue management models; see Ratliff et al. (2008), Newman et al. (2014), van Ryzin and Vulcano (2014) for recent references. As a result, demand models in the last decade have moved from independent to discrete choice-based models to capture the effects of substitution. In a choice-based demand model, customers are assumed to make choices from a menu of options according to a preference list; if the most preferred option is unavailable, they go down the list to pick an available option, as long as it is preferred to the outside or the no-purchase option; see Farias et al. (2013). Specific assumptions on how customers construct preference lists result in specific choice models. One then estimates the choice model using revealed preference data from customers in the form of choices from different offer sets; such revealed preference data are typically available in practice in the form of sales transactions. Using the predictions from the model, one determines the optimal assortment by searching through the space of all the feasible offer sets.

In this paper, we consider the problem of determining the optimal offer set, given a demand model that is already estimated from data. The subset must be chosen from a universe N, consisting of n products. Prices are exogenously set. The objective is to choose the offer set to maximize the expected revenue per customer, subject to potential constraints on the number of products that can be offered. This problem is popularly referred to in literature as the *static assortment optimization* problem. The key challenge in finding the optimal subset is that, in general, one must search through a combinatorial space of subsets, making the problem computationally intractable. For instance, the number of products in many practical settings easily exceeds 50, requiring a search of over 2^{50} subsets for an exhaustive search.

To address the computational challenge in searching for the optimal offer set, existing literature has taken two broad approaches: (a) impose parametric structures on the underlying choice models so that the resulting structure in the revenue function can be exploited to efficiently search for the optimal assortment and (b) propose general-purpose 'set optimization' heuristics and assess their theoretical and empirical performance. The first approach has been successfully applied for the following choice structures: multinomial logit (MNL) model (Talluri and van Ryzin 2004a, Rusmevichientong et al. 2010); multinomial logit model with uncertain parameters (Rusmevichientong and Topaloglu 2012); a specific variant of the two-level (Davis et al. 2014, Gallego and Topaloglu 2014) and d-level (Li et al. 2015) nested logit (NL) models; and specific variants of the mixed logit (ML) model (Rusmevichientong et al. 2014). The second approach has resulted in two general-purpose heuristics: (a) the *revenue-ordered* (RO) heuristic and (b) the *naive-greedy* (NG) heuristic.

The RO heuristic searches for the revenue maximizing subset from the n revenue-ordered subsets, where each revenue-ordered subset comprises the top-k priced products for $k = 1, 2, \ldots, n$. It is motivated by revenue management applications in which it is practically desirable to offer nested-by-fare offer sets. It has been shown to be optimal under the MNL model without a cardinality constraint (Talluri and van Ryzin 2004a) and under specific variants of the mixed-logit model (Rusmevichientong et al. 2014). It also has approximation guarantees for the optimality gap under general mixed-logit structures with a finite number of mixture components and excellent empirical performance, generally yielding profits within a fraction of a percent of the optimal for the instances they considered. On the other hand, the NG heuristic starts with an empty subset as the initial solution and iteratively adds the product that results in the maximum increase in revenues in each iteration, until there is no improvement or the cardinality constraint is reached; the reason for the *naive* moniker will become apparent below. It is a natural set optimization heuristic to consider. It has no theoretical guarantees but has been observed to have excellent empirical performance (Méndez-Díaz et al. 2014, Feldman and Topaloglu 2015).

The greatest advantage of the first approach, when it can be successfully executed, is that it results in efficient algorithms, often with theoretical approximation and convergence guarantees. The algorithms also shed light on the mathematical structures underlying the considered choice models. However, they suffer from two key disadvantages: (a) *lack of portability*: the structural insights and resulting algorithms do not apply to other choice models and (b) *challenging to implement*: different choice structures require different implementations, with some of them requiring intricate implementations (e.g., Rusmevichientong et al. (2010) and Li et al. (2015)) and others requiring solving linear and integer programs (e.g., Davis et al. (2014) and Bront et al. (2009)). The first disadvantage is critical because several other choice structures for which no efficient algorithms exist have been found to fit better in a large number of applications (see Greene (2003), Train (2009), Ben-Akiva and Lerman (1985)); examples include: general variants of the two-level nested logit model, the random parameters logit model, the latent-class logit model, the mallows and generalized mallows models, and other non-parametric 'black-box' approaches (see Farias et al. (2013)).

The general-purpose heuristics, on the other hand, score well on portability and ease-of-implementation. However, they suffer from two important performance issues: (a) failure to reach optimality, even for simple choice structures – for instance, both the RO and NG heuristics do not necessarily converge to the optimal solution, even for the MNL model when there is a cardinality constraint; and (b) converging to significantly sub-optimal solutions for common model structures – both RO and NG yield significantly sub-optimal solutions for common two-level nested logit and mixed logit instances (see Section 5). These performance issues have significant practical implications: a 1% increase in revenues can translate into a 20% increase in profits (see Méndez-Díaz et al. (2014)) for an average retailer because of razor-thin margins. Indeed, due to sub-optimality, these algorithms are leaving a significant amount of money on the table.

Our goal in this paper is to propose and study a general-purpose heuristic for assortment optimization that (a) is portable and easy-to-implement and (b) has excellent empirical performance and can be proved to converge to the optimal solution for common choice structures. We are primarily motivated by the portability consideration to obtain an algorithm that can be used with a large number of choice structures. Ideally, we want to gain the portability with little to no loss in performance.

In view of the above objective, we make the following contributions:

1. We study a local-search heuristic called ADXOpt that extends the NG algorithm by expanding the operation set to include additions, deletions, and exchanges (the A, D, and X respectively in ADXOpt) in each iteration. Because the deletions and exchanges can result in an exponential (in n) number of iterations before convergence, we cap the number of times each product may be removed from the solution (either through a deletion or an exchange) to ensure convergence in polynomial number of iterations.
2. Theoretically, we derive performance guarantees for general choice structures. The guarantees are problem dependent i.e., they depend on the parameter values of the underlying choice models. We specialize the bounds for the MNL, random parameters logit (RPL), and NL models. Our results show that ADXOpt converges to the optimal solution of the MNL model, both with and without a cardinality constraint on the number of offered products. Our guarantees for the RPL model characterize the performance of ADXOpt as a function of the heterogeneity in the underlying choice behavior. Finally, for specific sub-classes of the NL model, which are known to be NP-hard, we derive performance guarantees that are better than existing guarantees for particular parameter regimes. To the best of our knowledge, this work is the first to provide guarantees for the assortment problem under the RPL model, which is the work-horse model – and arguably more popular than the MNL model – in the areas of econometrics and marketing (Train 2009, Bajari et al. 2007).
3. We compare the empirical performance of ADXOpt against the RO and NG heuristics and the best model-based heuristic (which exploits specific model structure) for a large number of model instances under the RPL, NL, and latent-class multinomial logit model classes. Our results demonstrate that (a) ADXOpt performs significantly better than both the RO and NG heuristics across the board (up to 60% and 8% higher revenues than RO and NG, respectively); (b) ADXOpt is within 0.02% of the optimal solution, on average, in the instances in which the model-based heuristic reached the optimal solution; and (c) ADXOpt extracts significantly more (up to 5%) revenue than the best model-based heuristic in a large number of problem instances for which the model-based heuristic fails to reach the optimal solution.

A few remarks are in order. First, like the RO and NG heuristics, ADXOpt only requires access to a revenue subroutine that returns the expected revenues when queried with an offer set and, hence, is portable. Second, to put the theoretical results in context, we note that there is no a priori reason to expect that a structure-agnostic, general-purpose heuristic such as ADXOpt should converge to the optimal solution for any interesting choice structures. Indeed, it is surprising that we are able to prove convergence to optimality for interesting choice structures. Third, it is interesting to note the magnitude of improvements that ADXOpt obtains over NG. ADXOpt allows for more operations than NG in each iteration, but theoretically, it may converge to a "bad" local optimum and provide a worse solution than NG. However, our empirical results demonstrate that the additional operations not only result in an improved solution but essentially wipe-off any optimality gaps from NG. Finally, our results establish the following clear-cut value for ADXOpt: (a) better than existing heuristics for a large number of cases for which no efficient optimization algorithms exist; (b) within 0.02% of revenue optimality on average otherwise, translating to within 0.3% of profit optimality (see Méndez-Díaz et al. (2014)).

2. Relevant literature

Our work is part of the growing literature in operations management (OM) and revenue management (RM) on solving the assortment problem when the underlying demand is described by a choice model. It also has methodological connections to the literature on optimizing (submodular) set functions that study the performance of the local-search algorithm. Because the focus of our work is in applying the algorithm to the assortment problem, we focus the literature review on the relevant work in OM and RM and discuss the connections to the literature on optimizing submodular set functions in Section 4.

The problem of incorporating choice into operational decision-making has been extensively studied in the areas of operations management (OM) and revenue management (RM). Traditionally, operations models have ignored substitution and assumed an "independent demand" (Talluri and van Ryzin 2004b) model in which the demand for each product is independent of the availability of the other products. Over the years, several studies in airline RM have shown the value from using choice-based demand models: Belobaba and Hopperstad (1999) conducted simulation studies on the well-known Passenger Origin and Destination Simulator (PODS) to demonstrate the value of corrections to the independent demand model; Ratliff et al. (2008) and Vulcano et al. (2010) used real airline market data to demonstrate average revenue improvements from using MNL choice-based RM approaches. Following such studies, there has been a significant amount of research in incorporating choice behavior into operational decision problems: Talluri and van Ryzin (2004a), Gallego et al. (2006), Liu and van Ryzin (2008), Zhang and Adelman (2009), Meissner and Strauss

(2012a,b), Zhang (2011), Kunnumkal and Topaloglu (2008, 2010), van Ryzin and Vulcano (2008), Chaneton and Vulcano (2011). Static assortment optimization is a key problem that arises in most of the above work. We summarize below the key existing results on static assortment optimization.

Talluri and van Ryzin (2004a) consider a single-leg airline seat allocation problem in which an airline sells tickets to an aircraft over a finite number of discrete time periods so that, at most, one customer arrives in each time period. The customer chooses from the multiple fare classes that are offered. The airline must decide which set of fare classes to offer or 'open' as a function of the available seat capacity and remaining booking time. They formulate this problem as dynamic program (DP) with two state variables and show that solving the DP requires solving a variant of the static assortment optimization problem. When customers make choices according to an MNL model, and when there is no capacity constraint, the optimal revenue maximizing offer set must be one of the revenue-ordered assortments: $\{1\}, \{1,2\}, \ldots, \{1,2,\ldots,n\}$ where the products are indexed such that $p_1 \geq p_2 \geq \cdots \geq p_n$ with p_i denoting the price of product i. As a result, the optimal policy is to start with a revenue-ordered assortment of fare classes and close (open) lower-priced fare classes as the available seat capacity (remaining booking time) decreases, all else being equal. The resulting nested policy has practical appeal.

Rusmevichientong et al. (2010) consider the problem of finding the revenue maximizing offer set under the MNL choice model when the size of offer set is constrained. They show that, with the capacity constraint, the revenue maximizing offer set may no longer be revenue-ordered. However, the revenue maximizing offer set can be determined in a computationally efficient manner. In particular, the authors specialize the approach of Megiddo (1979) to solve combinatorial optimization problems with rational objective functions to obtain an efficient algorithm for determining the revenue maximizing offer set. The algorithm has $O(nC)$ complexity.

Rusmevichientong and Topaloglu (2012) consider the setting in which customers make choices according to an MNL model, but decision-maker knows the model parameters only up to a compact uncertainty set. The decision-maker wants to be protected from the uncertainty. As a result, her goal is to determine the offer set that maximizes the worst-case expected revenues over all the parameter values in the uncertainty set. The authors show that when there is no capacity constraint, surprisingly, the revenue-ordered subsets remain optimal even when the goal is to protect against the worst-case expected revenue. This result generalizes the result of Talluri and van Ryzin (2004a) and also provides a novel (and simpler) proof that leads to a new understanding of the structure of the MNL model.

Davis et al. (2014) study the problem of assortment optimization under the two-level nested logit model. The authors show that when there is no cardinality constraint, the revenue maximizing offer set can be determined in a computationally efficient manner when the nest dissimilarity parameters are all less than one and the no-purchase alternative is in a nest of its own. Gallego and Topaloglu (2014) extend this result to the case of when there is a capacity constraint per nest and Li et al. (2015) extend the result to the d-level nested logit model. Davis et al. (2014) also show that relaxing the assumption that all nest dissimilarity parameters are less than one or that the no-purchase alternative is in a nest of its own makes even the uncapacitated problem NP-hard. The reduction is from the well-known *partition problem* (Garey and Johnson 1979).

Bront et al. (2009) show that the uncapacitated assortment optimization problem under the Mixed-Logit model is NP-hard in the strong sense when there are at least n customer classes. The reduction is from the well-known *minimum vertex cover problem* (Garey and Johnson 1979). Méndez-Díaz et al. (2014) propose a branch-and-cut algorithm to find the optimal assortment under the general Mixed-Logit model. Rusmevichientong et al. (2014) extend the NP-hardness result to the case when there are at least two customer classes. The reduction, however, is from the *partition problem* (Garey and Johnson 1979) and, hence, proves NP-hardness in the weak sense. The authors in Rusmevichientong et al. (2014) consider two special instances of the Mixed-Logit model class for which they show that the optimal assortment must be one of the revenue-ordered subsets: (a) *product-independent price sensitivity:* the intrinsic mean utility of each product is given by a product-specific term that is not random, which is added to a term that is dependent on price with the price sensitivity being sampled from the uniform distribution; and (b) *value-conscious:* any realization of the mean utility vector satisfies $V_1 \leq V_2 \leq \ldots \leq V_n$ and $p_1 V_1 \geq p_2 V_2 \geq \cdots \geq p_n V_n$ where products are indexed such that $p_1 \geq p_2 \geq \cdots \geq p_n$ with p_i denoting the price of product i. The variant 'product-independent price sensitivity' can equivalently be formulated as the model in which the intrinsic utilities of the customers (belonging to different classes) vary only for the no-purchase alternative.

Other variants of assortment optimization have also been considered in literature. Mahajan and van Ryzin (1999) consider the assortment optimization problem where the products both generate revenues and result in operational costs. Cachon et al. (2005) extend this setup to include product search costs, where a customer may decide not to purchase an acceptable product from the store hoping that further search may surface a higher utility product.

Our algorithm is general and can accommodate choice models different from the MNL, NL, and ML model classes. Train (2009) describes several extensions to the Nested Logit and Mixed Logit models. Train (2009) also describes the Multinomial Probit Model (MNP), which is extensively used in several applications. Farias et al. (2013) describe a *non-parametric method* to choice modeling in which they allow the choice model to be *any* distribution over preference lists and design a revenue sub-routine that predicts the expected revenues for any offer set as the worst-case revenue over all choice models that are consistent with given data. They show that, while the model is very complex, the method has high predictive power. The paper, however, does not consider the optimization problem. Jagabathula and Rusmevichientong (2015) consider a two-stage model in which customers first form a consideration set and then choose according to a choice model. In the context of this model, the authors propose semi-parametric techniques to predict the revenues as a function of the price vector and offer set.

3. Local search algorithm: ADXOpt

The objective is to find the revenue maximizing subset containing at most C products from a universe N consisting of n products. We are given a revenue function $R\colon 2^N \to \mathbb{R}_+$ that returns the expected revenues when queried with an offer set; here, 2^N denotes the collection of all subsets of N, and $\mathbb{R}_+$ denotes the set of all non-negative real numbers. We assume that $R(\emptyset) = 0$. For simplicity of notation, let $S + j$ and $S - j$ denote $S \cup \{j\}$ and $S \setminus \{j\}$ respectively for any subset S. Our goal is to solve the following set optimization problem:

$$\underset{S \subseteq N\colon |S| \leq C}{\arg\max} \; R(S), \tag{1}$$

where $|\cdot|$ denotes the cardinality of set S and C is the available capacity. This problem is referred to as the *capacitated static assortment optimization problem* or simply the 'capacitated' problem. Setting $C = n$ yields the uncapacitated problem. We assume throughout that the revenue function $R(\cdot)$ returns the expected revenue for each offer set in a computationally efficient manner.

Finding the optimal assortment, in general, requires an exhaustive search over all the $O(n^C)$ feasible assortments, which is prohibitive for large values of n and C. Therefore, we adopt the following local-search heuristic. Start with the empty subset as the initial solution. In each iteration, determine the product that, when added, results in the maximum increase in revenues. If such a product exists *and* the size of the current solution is strictly less than C, then add it to the current solution and repeat. Otherwise, consider the following two operations: delete a product from the solution or exchange an existing product with a new product. Of these, pick the operation that results in the maximum increase in revenues. The algorithm may be allowed to terminate when it reaches a local optimum i.e., when none of the operations result in a strict increase in the revenues. This termination criterion, however, may require an exponential (in n and C) number of iterations to converge because of the deletions and exchanges. Therefore, to force convergence in a polynomial number of iterations, we cap the number of times each product can be removed from the solution, during deletion or exchange operations, to be an integer b. If at any point in the algorithm, a product is removed b times, we disregard it and do not consider it for any subsequent additions, as part of the add or exchange operations. The algorithm then terminates when either a local optimum is reached or when every product has been removed at least b times, whichever is

Figure 1 ADXOpt

Input: Universe N, Capacity C, Revenue function $R(\cdot)$, Max number b of removals per product, Initial estimate S^0

Output: Subset of size at most C

Algorithm:

Initialization: $\text{removals}(i) = 0$ for each $i \in N$

do for each $t \geq 0$

available products $= \{j \in N : \text{removals}(j) < b\}$
//try additions
$S^{\mathrm{A}} \leftarrow \arg\max_{S \in \mathcal{M}(\mathrm{A})} R(S)$, $\mathcal{M}(\mathrm{A}) = \{S^t + j : j \in \text{available products}\}$ if $|S^t| < C$ and $\emptyset$ o.w.
if $R(S^{\mathrm{A}}) \leq R(S^t)$ **then**
//try deletions and exchanges
$S^{\mathrm{D}} \leftarrow \arg\max_{S \in \mathcal{M}(\mathrm{D})} R(S)$, $\mathcal{M}(\mathrm{D}) = \{S^t - i : i \in S^t\}$
$S^{\mathrm{X}} \leftarrow \arg\max_{S \in \mathcal{M}(\mathrm{X})} R(S)$, $\mathcal{M}(\mathrm{X}) = \{S^t - i + j : i \in S^t \text{ and } j \in \text{available products}\}$
// choose operation in a greedy manner
$S^{t+1} = \arg\max \{R(S^{\mathrm{D}}), R(S^{\mathrm{X}})\}$
else
$S^{t+1} = S^{\mathrm{A}}$
// update removal count for removed products
$\text{removals}(i) \leftarrow \text{removals}(i) + 1$ for all $i \in S^t \setminus S^{t+1}$

while $R(S^{t+1}) > R(S^t)$ **and** $\text{removals}(i) < b$ for some $i \in N$

Output: $\arg\max \{R(S^{t+1}), R(S^t)\}$

sooner. We term this algorithm the ADXOpt algorithm to represent the three feasible operations: Add, Delete, and eXchange. The formal description of the algorithm is given in Figure 3.

The following remarks are in order. First, the above algorithm may be modified to remove some of the operations, usually the exchange operation. Dropping operations make the algorithm run faster, sometimes with little to no loss in performance. We explore this aspect further in our theoretical and empirical analysis. Second, the algorithm requires two parameters: the initial assortment S^0 and the removal cap b. Throughout, we set $S^0 = \emptyset$ in order to do a fair comparison with existing heuristics. It is clear, however, that the we can boost the performance of an existing heuristic through ADXOpt by taking S^0 to be the estimate obtained from the heuristic. This observation has practical relevance. Further, the number of removals b plays an important role in determining the running time and optimality gap of the algorithm. We explore these issues in subsequent sections, but numerically we found $b = 1$ to yield good performance. Third, dropping the delete and exchange operations results in the NG heuristic. Despite this fact, ADXOpt may not necessarily result in higher revenues because it may get stuck in bad local optima. Our theoretical and empirical analysis reveals that this is, indeed, not the case. Finally, the algorithm prioritizes additions over deletions/exchanges in order to avoid cycling; such prioritization is standard in literature (Feige et al. 2011).

Computational complexity of ADXOpt**.** It is easy to see that each iteration requires searching over $O(nC)$ assortments. Since each product can be removed at most b times, and the product must be added before it is removed, a product will be added or removed at most $2b$ times. Furthermore, since each iteration involves adding or removing at least one product, the number of iterations can be no more than $2nb$. Hence, the algorithm searches over $O(n^2Cb)$ assortments. Note that if we eliminate exchanges – which may be done when there is no capacity constraint – each iteration involves searching over $O(n)$ assortments, and the overall computational complexity improves to searching over $O(n^2b)$ assortments.

4. Theoretical guarantees for the ADXOpt algorithm

We now study the theoretical properties of the ADXOpt algorithm with the objective of determining the quality of the solution obtained. Like for any local search algorithm, two aspects are of interest: (a) polynomial-time convergence to either an exact or approximate local optimum and (b) quality of a local optimum S as an approximation of the global optimum S^*, measured in terms of the *locality ratio*, defined as

$$\text{locality ratio} = \min_{S \in \mathcal{S}} R(S)/R(S^*), \quad \mathcal{S} = \{S \colon S \text{ is feasible and a local optimum}\}.$$

A solution S is a local optimum with respect to the add, delete, and exchange operations if and only if $R(S) > R(S+j)$, $R(S) > R(S-i)$, and $R(S) > R(S+j-i)$ for any $j \notin S$ and $i \in S$.

Finding the local optimum, in general, is a hard problem. For instance, finding the local optimum for the Max Cut problem is PLS-complete (Schäffer 1991). In addition, a local optimum can be an arbitrarily bad approximation of the global optimum. For this reason, most existing work analyzes local-search algorithms when the set functions possess particular structures. The most commonly studied structure is *submodularity*, which requires the set function $f(\cdot)$ to satisfy $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$, for any two subsets A and B. Feige et al. (2011) provide a deterministic local search $\frac{1}{3}$−approximation algorithm when $f(\cdot)$ is a non-negative submodular function. Beyond submodular set functions, few general structures exist for which local-search algorithms have been analyzed. Unfortunately, existing results and analysis techniques for the submodular functions do not apply to our setting because the revenue function is not submodular, even for the special case when the underlying choice model is MNL. Instead, we develop new analysis techniques that rely on the underlying choice structures to derive performance guarantees for the local-search algorithm.

In particular, we identify choice structures and the corresponding parameter spaces for which ADXOpt obtains good quality approximations.

We organize the analysis of ADXOpt as follows. First, we derive a lower bound for the locality ratio for *any* "weakly rational" choice model, defined below. The bound is problem-dependent, i.e., it depends on the parameters of the underlying choice model. Because the bound is general, it can be immediately applied to obtain a performance bound for particular choice structures. To illustrate, we specialize the bound to the MNL, random parameters logit (RPL), and the NL models. Our bounds characterize the parameter space for these models, over which a local-optimum is a good approximation to the global optimum. Second, we identify sufficient conditions under which ADXOpt reaches a local optimum in a polynomial number of steps. We show that the MNL and (particular sub-classes of) NL models satisfy the sufficient conditions. When these sufficient conditions are not satisfied, we settle for an approximate local optimum. In particular, we use existing techniques to obtain an ε-approximate local optimum in $O(\frac{1}{\varepsilon})$ steps. Most of our analysis is focused on uncapacitated optimization. The case with a cardinality constraint is much harder. While we are not able to provide bounds for this general problem, we are able to show that ADXOpt converges to the global optimum efficiently for the MNL model.

We focus our analysis on the following setting. The choice behavior of the customers is governed by an underlying choice model. When offered subset S, customers purchase product $j \in S$ with probability $P_j(S)$ and leave without purchasing with probability $P_0(S)$, where 0 denotes the no-purchase option. We suppose that the choice probabilities exhibit strict weak rationality: $P_j(S) > P_j(S')$ for any two subsets $S \subset S'$ and $j \in S' \cup \{0\}$. We also assume that they constitute a valid probability mass function i.e., $P_j(S) \geq 0$ for all $j \in S \cup \{0\}$ and $S \subseteq N$ and $P_0(S) + \sum_{j \in S} P_j(S) = 1$. The expected revenue is given by $R(S) = \sum_{j \in S} p_j P_j(S)$. To avoid degenerate solutions and simplify exposition, we assume for this section that $p_1 > p_2 > \cdots > p_n$ and $P_j(S) > 0$ for any $j \in S$ and $S \subseteq N$.

4.1. Bounds for the locality ratio

We first focus on deriving bounds for the locality ratio. In the next subsection, we deal with the issue of finding a local optimum solution in polynomial time.

We start with the following definitions. For any subset S, define:

$$X_j(S) \stackrel{\text{def}}{=} \max_{i \in S} \frac{P_i(S) - P_i(S+j)}{P_i(S) P_j(S+j)} \text{ for any } j \notin S \text{ and } Y_j(S) \stackrel{\text{def}}{=} \min_{i \in S-j} \frac{P_i(S-j) - P_i(S)}{P_i(S-j) P_j(S)} \text{ for any } j \in S.$$

We extend the definition of $X_\cdot(\cdot)$ and define for any subset Δ such that $S \cap \Delta = \varnothing$

$$X_\Delta(S) \stackrel{\text{def}}{=} \max_{i \in S} \frac{P_i(S) - P_i(S+\Delta)}{P_i(S) P_\Delta(S+\Delta)},$$

where we use the notation $P_\Delta(S) \overset{\text{def}}{=} \sum_{i\in\Delta} P_i(S)$ for any subset S and $\Delta \subseteq S$. Note that under the assumption of weak rationality, $X_j(S), Y_j(S), X_\Delta(S) > 0$ for any j, S, and Δ. Further, define the constants:

$$\lambda_1 \overset{\text{def}}{=} \max_{\Delta, S\subseteq[n]} X_\Delta(S) \text{ and } \lambda_2 \overset{\text{def}}{=} \min_{j\in S, S\subseteq[n]} Y_j(S).$$

Note that it follows from our definitions that if $j \in S$, then $Y_j(S) \le X_j(S-j)$. Therefore, we have

$$\lambda_2 \le Y_j(S) \le X_j(S-j) \le \lambda_1. \tag{2}$$

We first derive the necessary conditions for local optimality:

Lemma 1 **(Necessary conditions for local optimality)**. *We have $R(S+j) < R(S) \implies p_j < X_j(S)R(S)$ and $R(S-j) < R(S) \implies p_j > Y_j(S)R(S)$.*

Using the result of Lemma 1, we can derive the following lower bound for the locality ratio.

Theorem 1 **(Locality ratio for general choice)**. *For any revenue function $R(S) = \sum_{j\in S} p_j P_j(S)$, where $P_j(\cdot)$ satisfy the strict weak rationality condition, the locality ratio is bounded below by*

$$\min\left\{\frac{\lambda_2}{\lambda_1}, \max\left\{\lambda_2 P_{S^*}(N), \frac{1-\lambda_1 P_N(N)}{(1-\lambda_2 P_N(N))^+}\right\}\right\},$$

where S^* is the global optimum and x^+ denotes $\max\{x, 0\}$ *for any real number x.*

A few remarks are in order. First, note that the result of Theorem 1 can be applied to *any* choice model that satisfies the strict weak rationality assumption. Most common choice models – in particular, all models belonging to the random utility maximization (RUM) class – satisfy the weak rationality assumption and, hence, the bound is broadly applicable. In order to apply the result, all we need is to obtain reasonable bounds for λ_1 and λ_2.

Second, the bound is close to 1 when λ_1 and λ_2 are close to each other. The gap between λ_1 and λ_2 essentially depends on the gap between $X_j(S)$ and $Y_j(S)$. Note that $(P_i(S+j) - P_i(S))/P_i(S)$ is the percentage of product i's demand that is cannibalized by the introduction of product j into assortment S; we call this the cannibalization effect of j on i (in subset S). Therefore, $P_j(S+j)X_j(S)$ measures the maximum cannibalization effect of j over all the products in S. Similarly, $Y_j(S)P_j(S)$ is the minimum cannibalization effect of j over all the products in $S-j$. Theorem 1 essentially shows that local optimality is sufficient to ensure global optimality if the cannibalization effect of each product j is uniform across all the products in S. More generally, the result of Theorem 1 provides a characterization of the locality ratio as a function of the "gap" between the maximum and minimum cannibalization effects of a product. This discussion becomes concrete (below) when we derive the locality ratios for particular choice structures.

We start with the basic attraction model (BAM), of which the popular MNL model is a special case. The attraction model is described by n non-negative parameters $v_j \geq 0$, one for each $j \in N$. The choice probability is given by $P_i(S) = v_i / \left(1 + \sum_{j \in S} v_j\right)$ for any product $i \in S$ and subset $S \subseteq N$. We can show that $\lambda_1 = \lambda_2 = 1$ for the attraction model; therefore, the locality ratio is 1. In other words, the local optimal solution is also the global optimum for the attraction model. More precisely, we have the following result:

THEOREM 2 **(Locality ratio for BAM)**. *The parameters $\lambda_1 = \lambda_2 = 1$ for a BAM model. Therefore, the locality ratio is equal to* 1*.*

For the BAM model, due to the Independence of the Irrelevant Alternatives (IIA) property, the cannibalization effect of any product j, as defined, is uniform across all products in the assortment. Therefore, the locality ratio is equal to 1.

Next, we consider the Random Parameters Logit (RPL) model, which is the work-horse model – and arguably more popular than the MNL model – in the areas of marketing and econometrics. The RPL model is widely used to model the heterogeneity in customer "tastes" when modeling the choice preferences of a population of customers. While an MNL model assumes that customers are homogeneous in the (mean) utilities they assign to each of the products, the RPL model allows the (mean) utilities to vary by customer. Particularly, the RPL model describes the preferences of a population by the multivariate distribution $\phi \colon \mathbb{R}^n \to [0,1]$. Each customer samples a parameter vector $\boldsymbol{\beta}$ according ϕ. When faced with an offer set S, the customer makes choices according to a logit model with parameters e^{β_j} for product j. As a result, the choice probability is given by:

$$P_i(S) = \int_{\boldsymbol{\beta} \in \mathbb{R}^n} L_i(\boldsymbol{\beta}, S)\, \phi(\boldsymbol{\beta})\, \mathrm{d}\boldsymbol{\beta}, \text{ where } L_i(\boldsymbol{\beta}, S) = \frac{e^{\beta_i}}{1 + \sum_{j \in S} e^{\beta_j}}$$

The RPL model has been widely used in marketing (both in Bayesian and non-Bayesian) settings to capture the heterogeneity in customer "tastes" (Berry et al. 1995, Rossi and Allenby 2003). The model has gained wide popularity in the field of econometrics starting with the seminal work of Berry et al. (1995). The popular assumption in literature has been to suppose that ϕ is a multivariate normal distribution, but several nonparametric distributions have also been considered Bajari et al. (2007). Existing literature in marketing and econometrics has proposed both parametric and nonparametric techniques to estimate the parameters of ϕ from choice data (Train 2009, Bajari et al. 2007). The RPL model has been recently gaining interest in operations. Recent work (Jagabathula and Vulcano 2015) has shown that the RPL models have better predictive power than the latent-class multinomial logit (LC-MNL) models in predicting the choice behavior of individuals as a function of the offer set. Despite their popularity, the RPL models have received little attention in the assortment optimization literature. The only exception is the work by Rusmevichientong et al. (2014), which shows that the assortment optimization problem is NP complete even when ϕ is a two-point distribution. The paper also derives theoretical guarantees for the RO heuristic. The best guarantee when ϕ is a continuous distribution is $R(\hat{S})/R(S) \geq 2/n$, which can be bad when n is large.

We derive the locality ratio for the RPL model when the distribution ϕ is a product of sub-Gaussian distributions. More precisely, we suppose that utility parameters $\boldsymbol{\beta}$ under the RPL model are specified as $\beta_i = \mu_i + Z_i$ for $i \in N$, where $\mu_i \in \mathbb{R}$ for all i and $(Z_i)_{i=1}^n$ are zero-mean and independent random variables. Each Z_i is σ_i-sub-Gaussian for some $\sigma_i > 0$ i.e., $\mathbb{E}\left[e^{tZ_i}\right] \leq e^{t^2\sigma_i^2/2}$ for all $t \in \mathbb{R}$. The parameter σ_i is called the variance proxy because it can be shown that $\mathrm{Var}(Z_i) \leq \sigma_i^2$. Several commonly used distributions in practice are sub-Gaussian. For instance, light-tailed distributions (Normal, Poisson, Gamma) and bounded distributions (Binomial, Uniform, finite-support) are sub-Gaussian. More precisely, if X is zero-mean and bounded so that $|X| \leq b$ a.s., then X is b-sub-Gaussian. Under these assumptions, we have the following result:

Theorem 3 **(Locality ratio for the RPL model)**. *Consider an RPL model with $\beta_i = \mu_i + Z_i$ for $i \in N$, where $\mu_i \in \mathbb{R}$ is the mean of β_i and $(Z_i)_{i=1}^n$ are zero-mean and independent random variables with each Z_i, a sub-Gaussian random variable having variance proxy σ_i. Let $\sigma^2 \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^n \sigma_i^2$ and $\gamma \stackrel{\text{def}}{=} \min_{i\in N} e^{\mu_i}/\left(1 + \sum_{j\in N} e^{\mu_j}\right)$. Then, the locality ratio for the RPL model is lower bounded by:*

$$(1-\varepsilon)\exp(-k_1\sigma\sqrt{cn}), \text{ where } c = \max\left\{1, \frac{1}{\sqrt{n}}\log\left(\frac{6}{\varepsilon\gamma}\right)\right\} \text{ and } \varepsilon \in (0,1),$$

where $k_1 = 6\sqrt{5}$, provided $P_N(N) \leq 1/(\lambda_1 + \lambda_2)$ or $\min_{i\in N} P_i(N) \geq 1/\lambda_1$ with $\lambda_1 \leq e^{k_1\sigma\sqrt{cn}}/(1-\varepsilon/2)$ and $\lambda_2 \geq e^{-k_1\sigma\sqrt{cn}}(1-\varepsilon/2)$.

Several remarks are in order. First, the bounds needed for $P_N(N)$ and $\min_i P_i(N)$ are technical conditions that we need to simplify the expression for the locality ratio because we are making no assumptions on the remaining parameters. Similar bounds for the locality ratio can be derived even when the technical conditions are not satisfied, but the expressions will be involved and lacking in any insights. As argued below, the technical conditions usually hold in regimes that are of interest.

Second, our result holds whenever ϕ is a product of *any* sub-Gaussian distributions. In particular, our result holds even when the random terms Z_i belong to *different* classes of distributions, so long as all they are all sub-Gaussian. Furthermore, because the random variables Z_i can have a finite support, our result also applies to a sub-class of the latent-class multinomial logit (LC-MNL) model class, in which ϕ is assumed to have a finite support.

Third, for a fixed value of n, setting $\sigma = 1/m$ and $\varepsilon = 1/m$ and m large enough, the locality ratio becomes $(1 - \frac{1}{m})\exp\left(-k_1 m^{-1} n^{1/4}\sqrt{\log(6m/\gamma)}\right)$, which converges to 1 as $m \to \infty$. Therefore, as

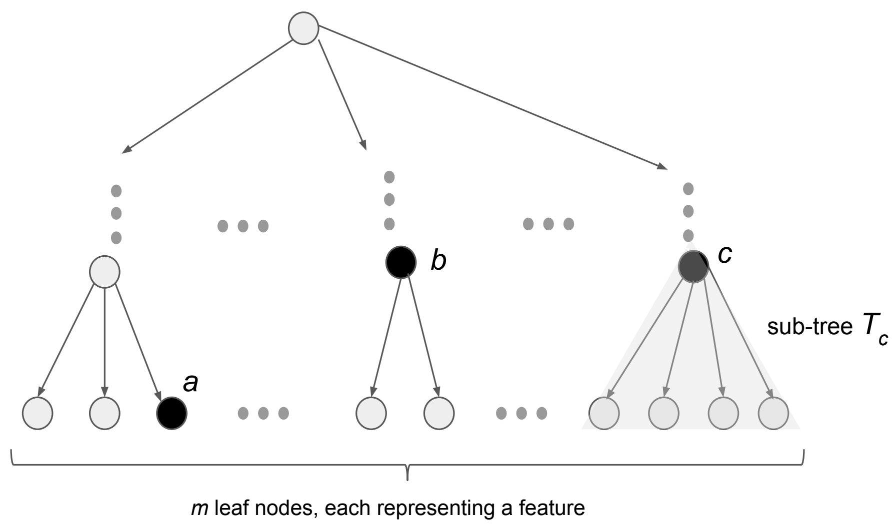

Figure 2 Tree-based representation of the horizontally differentiated products. Each leaf node represents an individual feature, such as a different shade of a color or a different design pattern. The shaded nodes, a, b, and c, denote the offered products. Each product is the bundle of features corresponding to leaf nodes in its sub-tree. The offered products have distinct features (because they are horizontally differentiated). Therefore, the sub-trees of the offered products are non-overlapping.

expected, the locality ratio goes to 1 as the heterogeneity in the underlying choice behavior goes to 0.

Similarly, when $n \to \infty$, suppose $\sigma\sqrt{n}$ remains a constant and $\gamma = \Omega(e^{-\sqrt{n}})$. In this regime, setting $\varepsilon \leq 1/2$ and choosing n to be large enough such that $c = 1$, we obtain a constant factor guarantee.

It is instructive to understand the settings in which Theorem 3 yields a constant factor guarantee for large values of n. As noted, the key required condition is $\sum_{i=1}^{n} \sigma_i^2 < \infty$ as $n \to \infty$ i.e., the "amount" of heterogeneity in the population, as captured by $\sum_{i=1}^{n} \sigma_i^2$, should be bounded above, as we offer a larger variety of products. In practice, we expect the cognitive capacity of individuals to be bounded above, and as a result, the heterogeneity in the population to be driven by only a few underlying factors, resulting in a bounded amount of heterogeneity. To illustrate this point, we discuss a generative model that imposes structure on how customers' utilities are generated as the number of products offered grows large. For this model, we show that the heterogeneity is bounded i.e., $\limsup_{n\to\infty} \sum_{i=1}^{n} \sigma_i^2 < \infty$.

Example 1 (Horizontally differentiated products). The standard assumption in literature is that the utility of a product is derived from the values that a customer assigns to its constituent features. Based on this, we consider the following generative model.

Suppose that the products are horizontally differentiated i.e., the products have the same "quality" but differ on other attributes such as color, flavor, packaging, etc. This assumption is common in literature (see Hotelling (1929), Lancaster (1990), Gaur and Honhon (2006)) and fits several examples: different flavors of yogurt, colors of bags, spice levels of ketchup, types of cereals, etc.

Products are bundles of features. There is a large number m of constituent features, such as different shades of color, ingredients for flavors, design patterns of bags, etc. that can be combined and included in products. The firm designs $n < m$ products by combining subsets of the m features, with each product corresponding to a particular subset of features.

The feasible combinations of features are described by a rooted tree consisting of m leaf nodes, each corresponding to a feature (see Figure 4.1 for an example). Every node a in the tree represents the collection of features corresponding to the set T_a of leaf nodes contained in the sub-tree rooted at a. The firm determines its offering of n products by choosing a collection of n nodes $\{a_1, a_2, \ldots, a_n\}$ in the tree with the requirement that $T_{a_i} \cap T_{a_j} = \varnothing$ for any $1 \le i < j \le n$. Note that, for any two nodes a and b in a tree, one of the following must be true: (i) $T_a \subseteq T_b$, (ii) $T_b \subseteq T_a$, or (iii) $T_a \cap T_b = \varnothing$. Therefore, our requirement that $T_{a_i} \cap T_{a_j} = \varnothing$ for distinct i and j eliminates the possibility of offering product bundles that are contained within one another to ensure that the products are horizontally differentiated.

The value X_ℓ, assigned to the feature corresponding to leaf node ℓ, follows a normal distribution with mean μ_ℓ and standard deviation $\sigma_\ell > 0$. The mean utility β_a, assigned to a product associated with node a, is equal to $\sum_{\ell \in T_a} X_\ell$. It is clear that β_a follows a normal distribution with mean $\mu_a = \sum_{\ell \in T_a} \mu_\ell$ and standard deviation $\sigma_a = \sqrt{\sum_{\ell \in T_a} \sigma_\ell^2}$. Because any node in the tree can be a product, we require that both $\sum_{\ell=1}^m \mu_\ell$ and $\sum_{\ell=1}^m \sigma_\ell^2$ are bounded sequences as $m \to \infty$. We, thus, have that

$$\limsup_{n\to\infty} \sum_{i=1}^n \sigma_{a_i}^2 = \limsup_{n\to\infty} \sum_{i=1}^n \sum_{\ell \in T_{a_i}} \sigma_\ell^2 \le \limsup_{m\to\infty} \sum_{\ell=1}^m \sigma_\ell^2 < \infty.$$

It follows from the result of Theorem 3 that ADXOpt obtains a constant factor approximation for the assortment problem. □

More precisely, we have the following corollary:

Corollary 1 **(Corollary to Theorem 3)**. *Suppose that, for the setting of horizontally differentiated products described in Example 1, we have* $\limsup_{n\to\infty} \sigma\sqrt{n} = \limsup_{n\to\infty} \sqrt{\sum_{i=1}^n \sigma_{a_i}^2} < 1/(6\sqrt{5})$. *Then, then the lower-bound of the locality ratio converges to* $e^{-1} \approx 0.37$.

Next, we consider a special case of the NL model. See Davis et al. (2014) for treatment of the general NL model class. We focus on the following special case with three nests. The first nest consists of only the no-purchase option, the second nest consists of the no-purchase option and one product, and the third nest consists of the no-purchase option and all the other products. Suppose that the price of the product in the second nest is large enough such that it is added in the first iteration and never removed in a subsequent iteration. The assortment problem now reduces to determining the subset of products in the second nest with the maximum expected revenues. Davis et al. (2014) show that this particular assortment problem is NP-hard. To simplify notation, we suppose that there is a set of n products, denoted by N, in the third nest, and any subset S is implicitly assumed to also contain the only product in the second nest, denoted by m. The choice probability is now given by:

$$P_i(S) = \frac{V(S)^\rho}{1 + V_m^\eta + V(S)^\rho} \cdot \frac{v_i}{V(S)} \text{ for } i \in S, i \neq m, \text{ and } P_m(S) = \frac{V_m^\eta}{1 + V_m^\eta + V(S)^\rho} \cdot \frac{v_m}{V_m},$$

where $V(S) = w_2 + \sum_{j \in S} v_j$ and $V_m = w_1 + v_m$ so that $v_j > 0$ is the parameter associated with product j and w_1 and w_2 are parameters associated with the no-purchase options in nests two and three, respectively. The parameters $\eta, \rho \in (0, 1]$ are the nest dissimilarity parameters. In line with Davis et al. (2014), we call this the partially-captured nested-logit (PC-NL) model.

We can prove the following result:

THEOREM 4 **(Locality ratio for the PC-NL model.)**. *The locality ratio for the PC-NL model is lower-bounded by* $1/(1+c)$ *whenever* $P_N(N) \leq 1/(2 \cdot (1+c))$, *where* $c = ((1-\rho)/\rho) \cdot (1 + v_0/w_2)$ *and* $v_0 = 1 + V_m^\eta$.

Davis et al. (2014) propose an algorithm that guarantees a $\frac{1}{2}-$approximation for the PC-NL model. It is clear from our result that the guarantee from Theorem 4 is better than $\frac{1}{2}$ when $c \leq 1$, which happens when $v_0/w_2 \leq \frac{2\rho - 1}{1-\rho}$. Therefore, when $\rho \approx 1$ or when $w_2 \gg v_0$, the local optimal solution provides a good approximation to the global optimum.

For the NL model, the IIA is satisfied only within the same nest. Therefore, a locality ratio of value that is strictly less than 1 is a consequence of the gap in the cannibalization effect between the products of the same nest and different nests, as well as the overall maximization and minimization involved in the definitions of λ_1 and λ_2.

4.2. Polynomial time convergence to a local optimum

We now identify conditions under which ADXOPT converges to a local optimum in a polynomial (in n) number of steps. As mentioned above, finding the local optimum is also a hard problem, in general. However, we show that, when the following sufficient condition is satisfied, the ADXOPT converges to a local optimum in a polynomial number of steps.

LEMMA 2 **(Monotonicity condition for convergence to local optimum.)**. *Suppose* $x_j(S') \geq X_j(S)$ *for any* $S' \subset S$ *and* $j \notin S$ *and* $y_j(S) \leq Y_j(S')$ *for any* $S' \subset S$ *and* $j \in S'$, *where*

$$x_j(S) \stackrel{\text{def}}{=} \min_{i \in S} \frac{P_i(S) - P_i(S+j)}{P_i(S) P_j(S+j)} \text{ and } y_j(S) = \max_{i \in S-j} \frac{P_i(S-j) - P_i(S)}{P_i(S-j) P_j(S)}.$$

Then, ADXOPT*, with $b = 1$ and using only the add and delete operations, converges to the solution that is a local optimum with respect to all the three (add, delete, and exchange) operations.*

We can show that both the MNL model and PC-NL model satisfy the monotonicity condition. Therefore, ADXOPT converges to the local optimum in a polynomial number of steps. More precisely, we have the following result:

THEOREM 5 **(Efficient convergence to local optimum for BAM and PC-NL)**. *Both the BAM and PC-NL models satisfy the monotonicity condition of Lemma 2. Therefore,* ADXOPT *converges to a local optimum in $O(n^2)$ steps.*

It is now a consequence of Theorems 2 and 5 that ADXOPT reaches the optimal solution in $O(n^2)$ steps for the BAM model. Similarly, it follows from Theorems 4 and 5 that ADXOPT obtains a factor $\frac{1}{1+c}-$approximation to the optimal solution in $O(n^2)$ steps, where $c = \frac{1-\rho}{\rho} \cdot (1 + v_0/w)$.

More generally, we have to settle for an approximate local optimum. Specifically, a $(1 + \alpha)-$approximate local optimum S_α is defined as (see, for instance, Feige et al. (2011)): (a) $(1 + \alpha)R(S_\alpha) \geq R(S_\alpha + j)$, (b) $(1 + \alpha)R(S_\alpha) \geq R(S_\alpha - i)$, and (c) $(1 + \alpha)R(S_\alpha) \geq R(S_\alpha + j - i)$ for any $i \in S_\alpha$ and $j \notin S_\alpha$. To assess the quality of approximation of a $(1 + \alpha)-$approximate local optimum, define the α-locality ratio: $\min_{S_\alpha} R(S_\alpha)/R(S^*)$, where S^* is the optimal solution. We show that we can obtain a $(1 + \alpha)-$approximate local optimal solution in $O(\frac{1}{\alpha})$ number of steps, and the $\alpha-$locality ratio scales as $O((1 - c\alpha) \times$ locality ratio for some constant $c > 0$. More precisely, we have the following result:

THEOREM 6 **(Guarantees for $(1 + \alpha)-$approximate local optimum)**. *For any $\alpha > 0$, let* ADXOPT(α) *be the* ADXOPT *algorithm in which the assortment is updated in iteration t only if $R(S^{t+1}) > (1 + \alpha)R(S^t)$. Then, given $\varepsilon > 0$,* ADXOPT(α) *converges to a solution S_α in $O(\frac{1}{\alpha} \cdot \frac{p_{\max}}{r_{\max}})$ steps such that $R(S_\alpha)/R(S) \geq (1 - \varepsilon) \times$ locality ratio bound, where locality ratio bound is the lower bound on the locality ratio given in Theorem 1, $\alpha = \varepsilon(\gamma - 1)/(2\gamma + \varepsilon)$ with $\gamma = \min_{i,\Delta,S:\, i \in S, \Delta \not\subseteq S} \frac{P_i(S)}{P_i(S+\Delta)}$, $p_{\max} = \max_{i \in N} p_i$, and $r_{\max} = \max_{i \in N} R(\{i\})$, whenever $P_N(N) \leq \frac{1}{(1+\varepsilon/2)\lambda_1 + \lambda_2}$ or $P_N(N) \geq \frac{1}{\lambda_1}$.*

The above result combined with Theorem 3 characterizes the trade-off between the computational complexity and the performance guarantee for the RPL model, as captured by ε.

4.3. Assortment optimization with a cardinality constraint

We now introduce a cardinality constraint so that the objective is to find the revenue maximizing subset containing at most C products. Unfortunately, the assortment problem now becomes very challenging. The only known result is for the BAM model, for which Rusmevichientong et al. (2010) develop an efficient algorithm that finds the optimal solution in $O(nC)$ steps. For the NL model, Rusmevichientong et al. (2009) provide a polynomial-time approximation scheme (PTAS) under the more general space constraint, in which each product is assumed to occupy some space, and the total space occupied by the assortment is constrained[1]. The PTAS, however, scales exponentially in the number of nests. Recent work by Gallego and Topaloglu (2014) provides an efficient algorithm for determining the revenue maximizing assortment under nest-specific cardinality constraints, in which the size of the assortment offered in each nest is constrained. The result, however, does not extend to the case where the size of the entire offered set is constrained. A straightforward application of the algorithm in Gallego and Topaloglu (2014) to deal with an assortment-level cardinality constraint requires searching over all possible m-partitions of the integer C, which scales exponentially in m, the number of nests.

Against this backdrop, as expected, analyzing ADXOpt for a general choice model becomes challenging. Nevertheless, we are able to exploit the MNL structure to show that ADXOpt with $b = C$ converges to the the optimal solution for the assortment problem with a cardinality constraint. Because $b = \min\{C, n-C+1\}$, it follows from the discussion below the description of ADXOpt that ADXOpt finds the optimal solution in $O(n^2C^2)$ steps. More precisely, we have the following result:

Theorem 7 **(Optimality of ADXOpt for MNL with cardinality constraints)**. *The ADXOpt algorithm, with $b = \min\{C, n-C+1\}$ and C, converges to the revenue maximizing subset of size at most C under the BAM model, provided the following technical assumption is satisfied:*

(a) $v_i > 0$ for all $i \in N$ and $v_i \neq v_j$ for all $i \neq j$ with $i, j \in N$

(b) $\mathcal{I}(i,j) \neq \mathcal{I}(s,t)$ for all $(i,j) \neq (s,t)$, $i \neq j$, $s \neq t$ with $i,j,s,t \in N$, where $\mathcal{I}(i,j) \stackrel{\text{def}}{=} (p_i v_i - p_j v_j)/(v_i - v_j)$ for $i \neq j$.

The technical conditions ease exposition by avoiding degenerate cases. The conditions are from Rusmevichientong et al. (2010). The proof is challenging and requires careful exploitation of the MNL structure that was established in Rusmevichientong et al. (2010); refer to the appendix for the details.

In recent, independent work, Davis et al. (2013) show that the assortment problem, with a cardinality constraint C, under the MNL model is equivalent to the following linear program (LP):

$$\max_{\boldsymbol{x}} \sum_{j\in N} p_j x_j, \qquad \text{s.t.} \quad x_0 + \sum_{j \in N} x_j = 1; \sum_{j\in N} x_j/v_j \leq C x_0; \; 0 \leq x_j \leq v_j x_0 \;\forall\, j \in N. \tag{3}$$

If $\boldsymbol{x}^*$ is the optimal solution to (3), then the revenue maximizing assortment of size at most C is given by $S^* = \{j \in N : x_j > 0\}$. Furthermore, the extreme points of the constraint polytope are of the

[1] Note that when all the products occupy the same space, a space constraint reduces to a cardinality constraint.

form $x_j = v_j y_j x_0$ for $j \in N$ and $x_0 = 1 - \sum_{j \in N} x_j$ such that $y_j \in \{0,1\}$ and $\sum_{j \in N} y_j \leq C$. Therefore, the objective value of the LP at an extreme point $\boldsymbol{x}$ is equal to $R(S)$, where $S = \{j \in N : x_j > 0\}$.

Based on the above result, we note that solving (3) using the simplex algorithm is equivalent to solving the assortment problem under the MNL model using ADXOpt. In each iteration, the simplex algorithm moves from an extreme point $\boldsymbol{x}$ to an adjacent extreme point $\tilde{\boldsymbol{x}}$ that results in the maximum increase in the objective value. Let $\tilde{S} \stackrel{\text{def}}{=} \{j \in N : \tilde{x}_j > 0\}$. Then, because $\tilde{\boldsymbol{x}}$ is an extreme point that is adjacent to $\boldsymbol{x}$, $\tilde{S}$ can be obtained from S by either adding, deleting, or exchanging a pair of products. Further, because the increase in the objective value when moving from $\boldsymbol{x}$ to $\tilde{\boldsymbol{x}}$ is equal to $R(\tilde{S}) - R(S)$, the simplex algorithm will move to $\tilde{S}$ that results in the maximum increase in the revenue from S.

The above work does not analyze the running time complexity of the simplex algorithm for (3). Although, in general, the simplex algorithm is known *not* to converge polynomial time, our result (Theorem 7) shows that, for the LP in (3), it converges in $O(n^2)$ steps in the unconstrained setting and $O(n^2C^2)$ steps in the constrained setting. In addition, the simplex algorithm will remove product j from the basis at most $\min\{C, n - C + 1\}$ times.

5. Numerical study

We complement our theoretical analysis with a numerical study to test the performance of the ADXOpt algorithm on a large number of instances of the random parameters logit (RPL), nested logit (NL), and the latent-class multinomial (LC-MNL) model classes. For each instance, we compared the expected revenues extracted by the ADXOpt algorithm against the revenues extracted by popular benchmark algorithms: the revenue-ordered (RO) heuristic, the naive-greedy (NG) heuristic, and the best model-based heuristic for the particular model class. The RO and NG heuristics share the desirable characteristics of the ADXOpt algorithm, portability and ease-of-implementation, but do not converge to the optimal solution, in general. The best model-based heuristics are model-specific by design and do not extend to other model classes, but can be shown to converge to the optimal solution in some cases.

Our results show that the ADXOpt algorithm extracts up to 60% higher revenues than RO and 8% higher than NG heuristics. In the retail context, with razor-thin margins, an increase of 1% revenues results in a 20% increase in profits (see Méndez-Díaz et al. (2014)). In the few instances in which the best model-based heuristic reaches the optimal solution, the ADXOpt algorithm is within an average of 0.02% of the optimal revenue. In the remaining instances, ADXOpt obtains up to an average of 5% higher revenues than the best model-based heuristic.

The broad experimental setup we used was as follows: (a) generate a random ground-truth model instance of the particular model class; (b) estimate the revenue maximizing subset using the different competing heuristics; and (c) compare the ground-truth revenues across the different estimates. We repeated the above sequence of steps for several model instances and various parameter combinations to cover the spectrum.

5.1. ADXOpt performance: Random Parameters Logit model class

For the purposes of the simulations, we focused on the setting in which the random parameters follow a normal distribution. We considered nine types of models, each distinguished by a correlation parameter ρ and a standard deviation parameter σ. We generated the types by varying ρ over the set $\{-.25, -.5, -1\}$ and σ over the set $\{2, 4, 10\}$. We set $n = 100$, and for each combination of the parameters (ρ, σ), we randomly generated 100 problem instances, with each instance generated as follows: for each product i, (a) sample σ_i uniformly at random from the interval $[\sigma, \sigma+1]$; (b) sample correlated uniform random variables U_1, U_2 by sampling X_1, X_2 according to a bivariate normal distribution with means 0, standard deviations 1, and correlation ρ, and setting $U_1 = F(X_1)$ and $U_2 = F(X_2)$, where $F(\cdot)$ is the cumulative distribution function (cdf) of the standard normal distribution; (c) set $\mu_i = -4 + 5 \cdot U_1$ and $p_i = 100 + 50 \cdot U_2$ so that μ_i and price p_i have marginal distributions that are uniform over $[-4, 1]$ and $[100, 150]$, respectively, and are negatively correlated to correspond with U_1 and U_2.

Because there is no closed form expression for the choice probabilities under the RPL model, we used the standard simulation technique. Specifically, for each model instance, we generated $T = 400$ samples $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_T$ with each component β_{ti} generated from a normal distribution with mean μ_i and standard deviation σ_i, for $i = 1, 2, \ldots, n$ and $t = 1, 2, \ldots, T$. The expected revenues from offer set S was then computed as $R(S) = \frac{1}{T}\sum_{t=1}^{T}\left(\sum_{i\in S} p_i e^{\beta_{ti}}/[1 + \sum_{i\in S} e^{\beta_{ti}}]\right)$.

For each model instance, we determined the revenue maximizing assortment using ADXOpt and the RO and NG heuristics. Table 1 summarizes the results. The second and fifth columns (entitled "Avg.") report the average, and the third and sixth columns (entitled "Max.") report the maximum percentage increase in revenues extracted by ADXOpt and the NG heuristic, relative to the RO heuristic, over the 100 problem instances. The fourth and seventh columns (entitled "Perc.") report the percentage of instances in which ADXOpt and the NG heuristic, respectively, beat the RO heuristic.

The key conclusion we draw from the results is that ADXOpt performs at the same level or, in many cases, better than both the NG and RO heuristics. The gains obtained by ADXOpt, relative to the RO heuristic, are particularly significant when there is a capacity constraint, with the gains being positive for 100% of the instances and the average gain reaching 28% in some cases. On the other hand, ADXOpt outperforms the NG heuristic by significant margins when the correlation (captured by ρ) between the prices and average "attractiveness" of the products is large and negative or when the heterogeneity (captured by σ) is large.

We observe that the RO heuristic performs well in uncapacitated cases. This may be explained by the fact that the RO heuristic is optimal when $\sigma = 0$, and as expected, its performance degrades as σ increases. The RO heuristic performs poorly compared to both ADXOpt and the NG heuristic when we introduce a cardinality constraint and more so when ρ is large and negative. To explain this, we note that, when there is a cardinality constraint, it may be optimal *not* to offer the highest revenue products that have low attractiveness. Indeed, because the parameter ρ captures the degree to which the attractiveness and price of products are negatively correlated, large and negative values of ρ indicate that higher priced products, on average, have low attractiveness. Therefore, we observe that the RO heuristic performs worse than ADXOpt for large (and negative) values of ρ.

Finally, we note that the NG heuristic performs worse than even the RO heuristic in almost all the instances when there is no capacity constraint. The reason is that, because price and attractiveness are negatively correlated, the NG heuristic starts adding products with high attractiveness but low prices, resulting in sub-optimal assortments with a large number of low-priced products. In contrast, ADXOpt has the ability to delete these products in later iterations, yielding higher revenues.

5.2. ADXOpt performance: Nested Logit model class

We considered twelve types of NL models, distinguished by the values of the nest dissimilarity parameters, price ranges, and the presence of the no-purchase alternative in a nest of its own. We denote each type by the triple $([\rho^L, \rho^U], [p^L, p^U], \text{FC})$, where $[\rho^L, \rho^U]$ is the range for the nest dissimilarity parameters, $[p^L, p^U]$ the range of prices of products, and $\text{FC} \in \{\text{Yes}, \text{No}\}$ is the indicator of whether the nests are fully-captured or not. Borrowing terminology from Davis et al. (2014), the nests are fully-captured if the no-purchase alternative is only present in a nest of its own and partially-captured if the no-purchase alternative is present in every nest. We generated nine types of partially-captured models fixing FC to be "No" and varying the ranges $[\rho^L, \rho^U]$ over $\{[0,1], [1,2], [2,3]\}$ and $[p^L, p^U]$ over $\{[100,150], [100,250], [100,350]\}$. We call these the *partially-captured* types. We then created three additional types of fully-captured models by fixing FC to be "Yes", $[\rho^L, \rho^U]$ to $[0,1]$ and varying the price range $[p^L, p^U]$ over $\{[100,150], [100,250], [100,350]\}$. We call these the *fully-captured* types.

As mentioned, the assortment problem is NP-hard for the partially-captured types and efficiently solvable for the fully-captured types of NL models. We used the LP-based heuristic proposed by Davis et al. (2014) as the best model-based heuristic for the NL class of models. The LP-based heuristic obtains the optimal solution for the three fully-captured types and an approximation for the remaining nine partially-captured types.

We chose $n = 200$ products and $m = 5$ nests. For each triple $([\rho^L, \rho^U], [p^L, p^U], \text{FC})$ under consideration, we randomly generated 1000 model instances as follows:

Parameters	ADXOpt(%)			NG(%)		
(ρ, σ)	Avg.	Max.	Perc.	Avg.	Max.	Perc.
Uncapacitated						
(−1.00, 2)	0.01	0.10	50	−8.09	−10.88	0
(−1.00, 4)	0.03	0.18	67	−7.08	−10.48	0
(−1.00, 10)	0.09	0.56	85	−1.75	−4.08	1
(−0.75, 2)	0.01	0.06	50	−1.89	−5.51	1
(−0.75, 4)	0.03	0.12	77	−2.12	−5.55	0
(−0.75, 10)	0.08	0.39	79	−1.13	−2.63	3
(−0.25, 2)	0.01	0.05	58	−0.38	−3.69	13
(−0.25, 4)	0.03	0.14	67	−0.57	−2.25	6
(−0.25, 10)	0.08	0.41	76	−0.52	−1.88	11
Capacitated 10%						
(−1.00, 2)	28.00	40.95	100	19.86	32.75	100
(−1.00, 4)	3.54	6.78	100	−6.14	−10.23	1
(−1.00, 10)	0.44	1.58	91	−2.07	−5.50	4
(−0.75, 2)	22.92	43.91	100	21.13	42.24	100
(−0.75, 4)	3.33	6.71	100	0.98	6.71	65
(−0.75, 10)	0.33	0.98	92	−1.30	−4.71	7
(−0.25, 2)	5.66	19.15	100	5.23	19.12	99
(−0.25, 4)	1.21	6.33	99	0.55	6.24	70
(−0.25, 10)	0.19	0.69	88	−0.54	−3.27	13

* "Max." values are min values when %s are < 0.

Table 1 Percentage increase in revenues of ADXOpt and naive greedy (NG) heuristics relative to the RO heuristic. The columns "Avg." and "Max." report the average and maximum percentage increases in revenues over the RO heuristic. The column "Perc." reports the percentage of instances in which ADXOpt and NG heuristics obtain higher revenues than the RO heuristic.

1. Randomly partition the n products into m nests, with each nest containing at least one product.
2. Sample the nest dissimilarity parameter ρ_i independently and uniformly at random from $[\rho^L, \rho^U]$ for each nest i.
3. Sample the price of each product independently and uniformly at random from $[p^L, p^U]$.
4. Sample the NL parameter for each product independently and uniformly at random from $[0, 10]$. Set the NL parameter v_0 of the no-purchase alternative in a nest of its own to 10.
5. If FC is "No", then sample the NL parameter v_{0i} of the no-purchase alternative in nest i independently and uniformly at random from $[0, 4]$, for each nest i. If, on the other hand, FC is Yes, then set $v_{0i} = 0$ for all nests i.

For each model instance generated as above, we determined the optimal assortment according to ADXOpt and the RO and NG heuristics. We compared the performance of the general-purpose heuristics against the LP-based heuristic, described next.

Borrowing notation from Davis et al. (2014), let N_i denote the set of products in nest i. The LP-based heuristic uses an LP-formulation to search over a collection of candidate assortments. The collection of candidate assortments are expressed as the cartesian product $\mathcal{A} = \mathcal{A}_1 \times \mathcal{A}_2 \cdots \mathcal{A}_m$, where $\mathcal{A}_i \subseteq 2^{N_i}$ with 2^{N_i} denoting the power set of N_i for each nest i. As a result, each candidate assortment is of the form $(S_1, \ldots, S_m)$, where each $S_i \subseteq N_i$. Different candidate assortments provide different approximations. Davis et al. (2014) consider the so-called nested-by-revenue collection of assortments, in which each $\mathcal{A}_i$ consists of $|N_i|$ nested assortments $\{\{(i,1),(i,2),\ldots,(i,|N_i|)\},\{(i,1),(i,2),\ldots,(i,|N_i|-1)\},\ldots,\{(i,1)\}\}$, where the products are indexed such that $p_{(i,1)} \geq p_{(i,2)} \geq \cdots \geq p_{(i,|N_i|)}$, with product j in nest i denoted by the duple (i,j) and $p_{(i,j)}$ denoting its price. With the nested-by-revenue collection, they propose to solve the following LP:

$$\begin{aligned} &\min_{x,\boldsymbol{y}} x \\ \text{subject to } &v_0 x \geq \sum_{i=1}^{m} y_i \\ &y_i \geq V_i(S_i)^{\rho_i}\left(R_i(S_i) - x\right), \text{ for all } S_i \in \mathcal{A}_i, 1 \leq i \leq m, \end{aligned}$$

where $V_i(S_i)$ denotes $v_{i0} + \sum_{j \in S_i} v_j$ and $R_i(S_i)$ denotes $\left(\sum_{j \in S_i} p_j v_j\right)/V_i(S_i)$. Note that the LP has $m+1$ variables and $n+1$ constraints. Once the LP is solved, the optimal assortment $(S_1^*, S_2^*, \ldots, S_m^*)$ can be reconstructed from the optimal solution $(x^*, \boldsymbol{y}^*)$ using the property that $y_i^* = V_i(S_i^*)^{\rho_i}\left(R_i(S_i^*) - x^*\right)$. See Davis et al. (2014) for further details. We note here that Davis et al. (2014) also numerically test the performance of the LP-based heuristic with other collections of candidate assortments – what they call the nested-by-preference-and-revenue and powers-of-δ assortments. Their findings (Davis et al. 2014, Tables 2, 3, 4) indicate that the nested-by-preference-and-revenue assortments obtain only a slight improvement and the powers-of-δ obtain no improvement over the nested-by-revenue assortments. For this reason, along with its intuitive appeal and ease-of-implementation, we chose to implement the LP-based heuristic with the nested-by-revenue collection.

Table 2 summarizes the percentage increase in the revenues from using the ADXOpt, revenue-ordered (RO), and naive greedy (NG) heuristics relative to the LP-based heuristics. The columns entitled ‘Avg.’ and ‘Max’ report the average and maximum percentage increase in revenues across the 1000 model instances, respectively. We draw the following conclusions:

Parameters	ADXOpt(%)		RO(%)		NG(%)	
$([\rho^L, \rho^U], [p^L, p^U], \text{FC})$	Avg.	Max.	Avg.	Max.	Avg.	Max.
$([0,1],[100,150],\text{Yes})^*$	−0.00	−0.03	−0.21	−2.36	−0.00	−0.87
$([0,1],[100,250],\text{Yes})^*$	−0.00	−0.27	−0.29	−9.01	−0.00	−2.09
$([0,1],[100,350],\text{Yes})^*$	−0.00	−0.10	−0.28	−7.43	−0.00	−1.73
$([0,1],[100,150],\text{No})$	0.45	3.83	0.27	2.84	0.45	3.83
$([0,1],[100,250],\text{No})$	0.72	10.47	0.43	8.51	0.71	10.47
$([0,1],[100,350],\text{No})$	0.78	7.08	0.47	4.70	0.77	7.08
$([1,2],[100,150],\text{No})$	1.78	8.46	2.10	8.22	1.33	8.46
$([1,2],[100,250],\text{No})$	2.36	10.55	2.36	10.08	1.79	10.49
$([1,2],[100,350],\text{No})$	2.43	13.41	2.47	13.23	1.88	13.41
$([2,3],[100,150],\text{No})$	6.76	28.11	6.58	27.71	6.41	28.11
$([2,3],[100,250],\text{No})$	7.43	29.22	7.07	27.23	6.96	29.22
$([2,3],[100,350],\text{No})$	7.18	27.90	6.89	29.73	6.73	27.90

* "Max." values are minimum values because % increases are negative.

Table 2 Percentage increase in revenues of ADXOpt**, revenue-ordered (RO), and naive greedy (NG) heuristics relative to the LP-based heuristic with nested-by-revenue assortments.**

1. For the fully-captured models, for which the LP-based heuristic is known to converge to the optimal solution, ADXOpt reaches remarkably close to optimality, with the average optimality gap less than a thousandth percent and maximum optimality gap less than 0.3%. Naive greedy also has excellent performance, with the average gap less than a hundredth percent.
2. For the partially-captured models, all the general-purpose heuristics outperform the LP-based heuristic, with ADXOpt having the best performance of all the heuristics. Specifically, ADXOpt obtains up to an average of 7.43% increase in revenues relative to the LP-based heuristic.
3. The performance improvement from the general-purpose heuristics vs. the LP-based heuristic increases as the nest dissimilarity parameters ρ_i increase beyond 1. As discussed in Davis et al. (2014), models with $\rho_i > 1$ capture applications in which the products within a nest are *synergistic* so that adding a product to a nest increases the sales of the existing products within the nest. Common real-world examples are product categories with 'loss leaders,' where different categories are modeled as different nests and the 'loss leaders' are included to attract traffic to the nests. This is in contrast to what Davis et al. (2014) call *competitive nests* in which adding a product to a nest necessarily decreases the probability of purchase for existing products in the nest; competitive nests are captured with $\rho_i < 1$. Our results suggest that, as the degree of synergy (as captured by ρ_i) increases, the revenue gains obtained from using ADXOpt over the LP-based heuristic become substantial.

5.3. ADXOpt performance: Latent-Class Multinomial Logit model class

Similar to the NL setting above, we tested the performance of the various heuristics on a large number of instances of the LC-MNL model class. We adopted two different simulation setups that have been considered in the existing literature to capture different types of market characteristics. The first simulation setup has been considered in Méndez-Díaz et al. (2014) to study the performance of a branch-and-cut (B&C) heuristic to solve a MILP formulation of the assortment problem under the LC-MNL model class. The setup focuses on the market in which different customer segments have different consideration sets but draw the preference weights of products according to the same underlying distribution. The second simulation setup has been considered in Feldman and Topaloglu (2015) to study the performance of a heuristic to determine an upper bound on the optimal revenues. It focuses on a market in which different segments have the same consideration set, the entire product universe, but differ on the preference weights assigned to the subset of 'specialty' products, which are non-mainstream or esoteric.

There are two prevailing heuristics for computing the optimal assortment under the LC-MNL model class: (a) the general-purpose RO heuristic studied in Rusmevichientong et al. (2014) and (b) the model-based MILP heuristic proposed by Bront et al. (2009). Rusmevichientong et al. (2014) derive theoretical guarantees and demonstrate good empirical performance in their simulation studies for the RO heuristic. Bront et al. (2009) formulated the assortment problem under the LC-MNL model class as an MILP, which was further refined by Méndez-Díaz et al. (2014) to obtain:

$$\max_{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}} \sum_{j=1}^{n}\sum_{k=1}^{K} \alpha_k p_j v_{kj} z_{kj} \tag{4}$$

$$\begin{aligned}
\text{subject to} \qquad & y_k + \sum_{j=1}^{n} v_{kj} z_{kj} = 1 \text{ for all } 1 \le k \le K \\
& y_k - z_{kj} \le 1 - x_j \text{ for all } 1 \le k \le K, 1 \le j \le n \\
& z_{kj} \le y_k \text{ for all } 1 \le k \le K, 1 \le j \le n \\
& (1 + v_{kj}) z_{kj} \le y_k \text{ for all } 1 \le k \le K, 1 \le j \le n \\
& x_j \in \{0,1\}, y_k, z_{kj} \ge 0 \text{ for all } 1 \le k \le K, 1 \le j \le n,
\end{aligned}$$

The optimal assortment can be obtained from the optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{z}^*)$ as $S^* = \left\{1 \le j \le n \colon x_j^* = 1\right\}$. See Méndez-Díaz et al. (2014) for further details on the MILP formulation (4). We compared the performance of ADXOpt against the RO heuristic (as part of the general-purpose heuristics) and the MILP-based heuristic. We note that solving the MILP can be time consuming for large problem instances. Hence, we solved the MILP with a time limit of $10\times$ the run time of the ADXOpt algorithm and report the best lower bound obtained. The MILPs were solved using Gurobi Optimizer version 6.0.2 on a computer with processor 3.5GHz Intel Core i5, RAM of 16GB, and operating system Mac OSX Yosemite.

Next, we describe the performance of the different optimization heuristics under the two simulation setups.

5.3.1. Market with differing consideration sets. For the first set of simulations, we considered $n = 200$ products and sampled their preference weights v_{kj} for each market segment k independently and uniformly at random from the interval $[0, 10]$. The preference weight v_{k0} for the no-purchase option was sampled independently and uniformly at random from $[0, 4]$ for each segment k. The segment sizes α_k, $1 \le k \le K$, were obtained by sampling β_k independently and uniformly at random from $[0, 1]$ and setting $\alpha_k = \beta_k / \sum_{k'=1}^{K} \beta_{k'}$. We sampled the prices of products independently and uniformly at random from an interval, specified below. As in Méndez-Díaz et al. (2014), we considered three types of markets.

Type 1 markets. The considerations sets for the different segments have the following structure: $C_2 \subset C_1$, $C_4 \subset C_3$, $\ldots$, $C_K \subset C_{K-1}$, where $C_k \subseteq N$ is the consideration set of segment k and K is the total of number of segments in the market. We set $K = 50$. Given product prices, we generated each consideration set C_k, $k = 1, 3, 5, \ldots, K-1$, by sampling 6 products uniformly at random from the universe. For each C_k, we set C_{k+1} to contain the three lowest priced products in C_k. These consideration sets are designed to capture a market structure in which customers belonging to the odd-numbered segments are price insensitive; whereas, the customers belonging to the even-numbered segments are price sensitive and consider only the three cheapest products among the ones considered by the corresponding odd-numbered segments.

As in Méndez-Díaz et al. (2014), we considered four different price ranges: $100-150$, $100-250$, $100-300$, and $100-350$, as well as both the uncapacitated and the capacitated settings, with a capacity constraint of 100. For each price range, we generated 100 problem instances and computed the percentage gap in the revenues obtained from using the ADXOpt, RO, and NG heuristics relative to the revenues obtained from solving the MILP. Table 3 presents the average and maximum gaps for the three heuristics – ADXOpt, RO, and NG – for the uncapacitated and capacitated settings, with the capacity constraint set to 50% i.e., 100. All the percentages are rounded to two decimal points. We note that the MILP reached the optimal solution in all of the reported instances and, hence, the percentage gaps are also the optimality gaps. The results demonstrate that the local-search heuristic has the best average and worst-case performance among the three general-purpose heuristics. Furthermore, its average and worst-case optimality gaps are no more than 0.02% and 0.70%, respectively.

Type 2 markets. These markets are characterized by a more general overlapping structure of consideration sets. We considered four different parameter settings denoted by (K, c), with the

Price range	ADXOpt(%)		RO(%)		NG(%)	
	Avg.	Min.	Avg.	Min.	Avg.	Min.
Uncapacitated						
$100-150$	-0.00	-0.11	-0.70	-1.97	-0.20	-0.87
$100-250$	-0.01	-0.26	-2.37	-4.50	-0.68	-3.05
$100-300$	-0.01	-0.38	-3.17	-6.94	-0.70	-3.98
$100-350$	-0.01	-0.52	-3.68	-8.68	-0.69	-3.56
Capacitated 50%						
$100-150$	0.00	-0.00	-31.94	-48.67	-0.13	-0.56
$100-250$	-0.02	-0.70	-25.70	-40.60	-0.79	-2.81
$100-300$	-0.00	-0.45	-24.72	-38.65	-0.67	-4.18
$100-350$	-0.02	-0.53	-22.50	-32.46	-0.75	-3.45

Table 3 Type 1 markets: Percentage increase in revenues of ADXOpt, revenue-ordered (RO), and naive greedy (NG) heuristics relative to the MILP-based heuristic for the LC-MNL models. The MILP reached optimality in all of the instances and, hence, the reported numbers are optimality gaps. All percentages are rounded to two decimal places; the average and maximum are over 100 simulation runs.

number of latent classes K varying over the set $\{30, 70\}$ and the overlap parameter c varying over the set $\{3, 9\}$. The overlap parameter c determined the generation of the consideration sets as follows. We started by sampling 15 products at random to construct C_1. Then, for any $2 \leq k \leq K$, we took the last c products from C_{k-1} and sampled another $15-c$ products to construct C_k. We set the price range to $100-200$. Table 4 presents the average and worst-case optimality gaps over 100 simulations for the three general-purpose heuristics. We considered both the uncapacitated and capacitated settings, with the capacity constraint set to 20% i.e., 40. The results, again, demonstrate the excellent performance of the ADXOpt algorithm. We note that the MILP converges to the optimal solution in all of the instances, and, hence, all the reported gaps are optimality gaps.

Type 3 markets. These markets have a less restrictive overlapping pattern of consideration sets. We set the number of segments $K = 70$. Given the cardinality c of the consideration sets, we sampled them as follows. Pick c products at random to generate C_1. Then, sample 3 products at random from C_1 and $c-3$ products at random from the remaining products to construct C_2. For C_3, sample 3 products at random from $C_1 \cup C_2$ and $c-3$ products from the remaining products. We repeated the above procedure for each $4 \leq k \leq K$ by sampling 3 products from $C_{k-2} \cup C_{k-1}$ and $c-3$ products from the remaining products to construct C_k. We set the price range to be $100-150$ and considered both the uncapacitated and the capacitated settings, with the capacity constraint set to 20% i.e., 40 products.

Table 5 presents the results from over 100 simulation runs. We note that, for some of the model instances, the MILP failed to reach optimality within the time limit of $10\times$ the running time of ADXOpt. The ADXOpt heuristic again outperforms all the other general-purpose heuristics. The performance of ADXOpt, relative to other general-purpose heuristics, is particularly significant

Parameters	ADXOpt(%)		RO(%)		NG(%)	
(K, c)	Avg.	Min.	Avg.	Min.	Avg.	Min.
Uncapacitated						
(30, 3)	−0.00	−0.14	−0.81	−2.93	−0.80	−3.87
(30, 9)	−0.00	−0.25	−1.58	−3.94	−0.68	−2.82
(70, 3)	−0.00	−0.18	−0.54	−1.33	−0.56	−1.86
(70, 9)	−0.00	−0.09	−1.36	−2.53	−0.70	−1.98
Capacitated 20%						
(30, 3)	−0.00	−0.07	−0.82	−2.48	−0.91	−3.73
(30, 9)	−0.00	−0.18	−3.46	−11.40	−0.72	−3.81
(70, 3)	−0.00	−0.31	−0.65	−2.54	−0.68	−3.24
(70, 9)	−0.00	−0.15	−2.73	−6.92	−0.69	−2.10

Table 4 Type 2 markets: Percentage increase in revenues of ADXOpt, revenue-ordered (RO), and naive greedy (NG) heuristics relative to the MILP-based heuristic for the LC-MNL models. The MILP reached optimality in all of the instances and, hence, the reported numbers are optimality gaps. All percentages are rounded to two decimal places; the average and maximum are over 100 simulation runs.

Parameters	ADXOpt(%)		RO(%)		NG(%)	
c	Avg.	Min.	Avg.	Min.	Avg.	Min.
Uncapacitated						
5	−0.00	−0.09	−0.61	−2.22	−0.38	−1.42
10	0.02	−0.10	−0.45	−1.72	−0.58	−1.49
15	−0.00	−0.06	−0.28	−0.73	−0.57	−1.97
Capacitated 20%						
5	−0.01	−0.30	−38.67	−59.63	−0.57	−2.03
10	0.01	−0.14	−13.43	−22.29	−1.15	−4.15
15	0.00	−0.08	−6.73	−12.42	−0.98	−3.06

Table 5 Type 3 markets: Percentage increase in revenues of ADXOpt, revenue-ordered (RO), and naive greedy (NG) heuristics relative to the MILP-based heuristic for the LC-MNL models. The MILP with a time limit of $10\times$ the running time of ADXOpt failed to reach optimality for some of the instances. All percentages are rounded to two decimal places; the average and maximum are over 100 simulation runs.

when there is a cardinality constraint. Finally, ADXOpt obtains slight performance gains, relative to the MILP, for some of the instances.

5.3.2. Market with specialty products. In order to demonstrate the value of the ADXOpt heuristic for the LC-MNL model class, this simulation setup focuses on problem instances for which the MILP in (4) cannot be solved to optimality within a reasonable amount of time. This setup is in contrast to the simulation setup described above in which we were able to solve the MILP to optimality or close to optimality for all the instances. As mentioned, the simulation setup below is similar to the one considered in Feldman and Topaloglu (2015).

For the purposes of the simulation, we considered $n = 100$ products. We sampled the product prices independently and uniformly at random from the interval $[0, 2000]$ for each model instance.

For K latent classes, we obtained the segment sizes α_k, $1 \le k \le K$, by sampling β_k independently and uniformly at random from $[0,1]$ and setting $\alpha_k = \beta_k / \sum_{k'=1}^{K} \beta_{k'}$. We selected a subset S of 40 products from the 100 products and designated them as staple products.

We considered a total of 27 types of LC-MNL models, distinguished by the number of latent classes, the magnitude of preference weights, and the no-purchase probability. We denote each type by the triple $(K, V, \bar{P}_0)$, where K denotes the number of latent classes, V captures the magnitude of the preference weights, as described below, and $\bar{P}_0$ denotes the no-purchase probability. We generated the 27 types by varying K over $\{25, 50, 75\}$, V over $\{5, 10, 20\}$, and $\bar{P}_0$ over $\{0.6, 0.8, 1.0\}$. Given $(K, V, \bar{P}_0)$, we generated preference weights for 1000 model instances as follows:

1. For each product j, sample ν_j independently and uniformly at random from $[1, V]$.
2. For each class k and product j, sample X_{kj} from $[0.3, 0.7]$ if j is a staple product and $[0.1, 0.3] \cup [0.7, 0.9]$ if j is a specialty product.
3. For each k, sample P_{0k}, the probability of no-purchase for customers of class k, independently and uniformly at random from $[0, \bar{P}_0]$.
4. Set the preference weights $v_{kj} = \nu_j X_{kj}(1 - P_{k0}) / (P_{k0} \sum_{i=1}^{n} \nu_i X_{ki})$, which ensures that the no-purchase probability is indeed P_{k0} for each k.

Note that K controls the degree of heterogeneity in the population, V controls the magnitude and variance across the attractiveness of the products, and $\bar{P}_0$ controls the magnitude and the variance across the classes of the no-purchase probability. Further, by design, specialty products have a higher variance in the preference weights across the classes when compared to the staple products.

For each model instance generated above, we determined the assortment from ADXOpt and compared its revenues to those obtained from the prevailing MILP-based heuristic.

Table 6 summarizes the results from our simulation study. The results show that the ADXOpt can obtain substantial improvements, viz. up to 9% on average and 67% in the best case, over the MILP-based heuristic for problem instances for which the MILP is computationally difficult to solve. This result, along with the above results, allow us to conclude that ADXOpt obtains solutions that are either close to optimal or substantially better than the other prevailing heuristics for the LC-MNL model class.

6. Summary and discussion

Summary. This paper studied both the theoretical and empirical performance of a general-purpose, local-search heuristic, called ADXOpt, for the problem of determining the revenue maximizing subset. ADXOpt is portable (directly applicable to a wide-range of choice models) and easy-to-implement. It iteratively builds an estimate of the best subset by starting with an empty subset and performing one of the three operations in each iteration: add, delete, or exchange. The

Parameters	Uncapacitated		Capacitated (50%)	
$(K, V, \bar{P}_0)$	Avg.	Max	Avg.	Max
(20, 5, 0.4)	5.86	26.61	8.99	45.85
(20, 5, 0.6)	4.32	11.46	1.42	11.95
(20, 5, 0.8)	3.63	10.43	1.00	4.66
(20, 10, 0.4)	5.53	27.04	8.67	67.34
(20, 10, 0.6)	4.27	14.64	1.27	9.51
(20, 10, 0.8)	3.74	11.52	1.05	6.85
(20, 20, 0.4)	5.67	30.13	8.64	49.82
(20, 20, 0.6)	4.18	11.80	1.12	6.68
(20, 20, 0.8)	3.73	13.17	1.11	7.36
(40, 5, 0.4)	5.72	29.16	8.58	44.78
(40, 5, 0.6)	4.21	10.46	1.36	10.40
(40, 5, 0.8)	3.70	14.34	1.03	5.87
(40, 10, 0.4)	5.68	29.36	8.74	66.14
(40, 10, 0.6)	4.24	14.40	1.17	7.46
(40, 10, 0.8)	3.67	10.98	1.07	6.24
(40, 20, 0.4)	5.57	32.65	8.51	58.80
(40, 20, 0.6)	4.20	13.38	0.97	7.62
(40, 20, 0.8)	3.66	10.88	1.06	6.11
(60, 5, 0.4)	5.37	20.69	8.15	64.09
(60, 5, 0.6)	4.15	12.41	1.31	10.64
(60, 5, 0.8)	3.58	11.94	1.07	5.19
(60, 10, 0.4)	5.73	32.59	8.24	40.07
(60, 10, 0.6)	4.21	11.94	1.16	7.71
(60, 10, 0.8)	3.70	11.95	1.06	5.60
(60, 20, 0.4)	5.68	27.35	7.98	49.00
(60, 20, 0.6)	4.12	11.21	1.04	6.38
(60, 20, 0.8)	3.70	11.98	1.08	7.39

Table 6 Percentage increase in revenues of ADXOpt **relative to the MILP-based heuristic. The MILP was solved with a time limit of** $10\times$ **the running time of** ADXOpt**.**

operation that results in the maximum increase in revenues is chosen in each iteration. In order to ensure convergence in polynomial time, we capped the number of times a product is removed (as part of either the delete or the exchange operations) to a fixed integer b.

Our approach is consistent with the stream of work focused on studying general-purpose heuristics, such as the revenue-ordered (RO) and the naive-greedy (NG) heuristics. It is in contrast to the approach of building model-based heuristics, which focus on exploiting model-specific structures to result in efficient algorithms. The model-based heuristics are not portable; whereas, the existing general-purpose heuristics suffer from poor performance. We showed that the local-search algorithm is both portable and has good performance.

In particular, we established the following theoretical properties. The algorithm converges to the optimal subset for both the capacitated and uncapacitated assortment optimization problems for the MNL model. We also derived guarantees for the RPL model class. Our guarantees characterize the quality of the solution obtained by ADXOpt as a function of the heterogeneity in the underlying choice behavior. To the best of our knowledge, our result is the first such characterization for the GMNL model. Finally, for a sub-class of NL models for which the assortment problem is known to be NP-hard, we derived performance guarantees. Our guarantees are better than existing guarantees for particular parameter regimes. We also performed extensive numerical studies that compare its performance against the general-purpose RO and NG heuristics and the best model-based heuristic. Our results indicate that the ADXOpt outperforms both RO and NG across the board, by significant margins in some cases. Further, relative to the best model-based heuristics, we established that ADXOpt is (a) significantly better than the best model-based heuristic for a large number of cases for which no efficient optimization algorithms exist; and (b) within 0.02% of revenue optimality otherwise.

Future work. Our work opens the doors for many exciting future research directions. First, the theoretical analysis of the algorithm can be extended to general nested logit models and other variants of the mixed logit model. We conjecture that the algorithm will converge to a local optimum for a general nested logit model with the no-purchase alternative in a separate nest of its own and the nest dissimilarity parameters taking values between 0 and 1. Furthermore, our numerical study suggests that even in the cases in which the algorithm does not converge to the optimal solution, it provides good approximations to the optimal decision. Understanding the settings in which the ADXOpt algorithm performs well and deriving theoretical guarantees are natural future steps.

The algorithm we proposed can be readily extended by allowing for additions, deletions, and exchanges of multiple products in each operation. Such flexibility certainly increases the computational complexity of the algorithm but will yield higher quality results. Understanding the benefits of such extensions is one potential future direction. Another direction is to restrict the class of subsets that are feasible. Our setting assumes that the subsets are restricted only by their cardinality. In practice, however, several considerations restrict the class of feasible subsets. Such restrictions, provided they possess sufficient structure, may make the optimization problem easier.

References

Patrick Bajari, Jeremy T Fox, and Stephen P Ryan. Linear regression estimation of discrete choice models with nonparametric distributions of random coefficients. *The American economic review*, 97(2):459–463, 2007.

P P Belobaba and C Hopperstad. Boeing/MIT simulation study: PODS results update. In *1999 AGIFORS Reservations and Yield Management Study Group Symposium, April*, pages 27–30, 1999.

M E Ben-Akiva and S R Lerman. *Discrete choice analysis: theory and application to travel demand*. CMIT press, Cambridge, MA, 1985.

Steven Berry, James Levinsohn, and Ariel Pakes. Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society*, pages 841–890, 1995.

Juan José Miranda Bront, Isabel Méndez-Díaz, and Gustavo Vulcano. A column generation algorithm for choice-based network revenue management. *Operations Research*, 57(3):769–784, 2009.

Gérard P Cachon, Christian Terwiesch, and Yi Xu. Retail Assortment Planning in the Presence of Consumer Search. *Manufacturing & Service Operations Management*, 7(4):330–346, October 2005.

J M Chaneton and G Vulcano. Computing bid prices for revenue management under customer choice behavior. *Manufacturing & Service Operations Management*, 2011.

James Davis, Guillermo Gallego, and Huseyin Topaloglu. Assortment planning under the multinomial logit model with totally unimodular constraint structures. *Department of IEOR, Columbia University. Available at http://www. columbia. edu/ gmg2/logit_const. pdf*, 2013.

James M Davis, Guillermo Gallego, and Huseyin Topaloglu. Assortment Optimization Under Variants of the Nested Logit Model. *Operations Research*, 62(2):250–273, April 2014.

Vivek F Farias, Srikanth Jagabathula, and Devavrat Shah. A Nonparametric Approach to Modeling Choice with Limited Data. *Management Science*, 59(2):305–322, February 2013.

Uriel Feige, Vahab S Mirrokni, and Jan Vondrak. Maximizing non-monotone submodular functions. *SIAM Journal on Computing*, 40(4):1133–1153, 2011.

Jacob Feldman and Huseyin Topaloglu. Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. *Production and Operations Management*, 2015.

Guillermo Gallego and Huseyin Topaloglu. Constrained assortment optimization for the nested logit model. *Management Science*, 60(10):2583–2601, 2014.

Guillermo Gallego, G Iyengar, R Phillips, and A Dubey. Managing flexible products on a network. 2006.

Michael R Garey and D S Johnson. *Computers and Intractability*. A Guide to the Theory of NP-Completeness. W H Freeman & Company, January 1979.

Vishal Gaur and Dorothée Honhon. Assortment planning and inventory decisions under a locational choice model. *Management Science*, 52(10):1528–1543, 2006.

W H Greene. Econometric analysis, 2003.

Harold Hotelling. Stability in competition. *The Economic Journal*, 39(153):41–57, 1929.

Daniel Hsu, Sham M Kakade, and Tong Zhang. A tail inequality for quadratic forms of subgaussian random vectors. *Electron. Commun. Probab*, 17(52):1–6, 2012.

Srikanth Jagabathula and Paat Rusmevichientong. A Nonparametric Joint Assortment and Price Choice Model. *Available at SSRN: http://ssrn.com/abstract=2286923*, 2015.

Srikanth Jagabathula and Gustavo Vulcano. A model to estimate individual preferences using panel data. *Available at SSRN: http://ssrn.com/abstract=2560994*, 2015.

Sumit Kunnumkal and Huseyin Topaloglu. A refined deterministic linear program for the network revenue management problem with customer choice behavior. *Naval Research Logistics (NRL)*, 55(6):563–580, September 2008.

Sumit Kunnumkal and Huseyin Topaloglu. A New Dynamic Programming Decomposition Method for the Network Revenue Management Problem with Customer Choice Behavior. *Production and Operations Management*, 19(5):575–590, September 2010.

Kelvin Lancaster. The economics of product variety: A survey. *Marketing Science*, 9(3):189–206, 1990.

Guang Li, Paat Rusmevichientong, and Huseyin Topaloglu. The d-level nested logit model: Assortment and price optimization problems. *Operations Research*, 62(2):325–342, 2015.

Qian Liu and G van Ryzin. On the choice-based linear programming model for network revenue management. *Manufacturing & Service Operations Management*, 2008.

S Mahajan and G J van Ryzin. On the Relationship between Inventory Costs and Variety Benefits in Retail Assortments. *Management Science*, 45(11):1496–1509, 1999.

Nimrod Megiddo. Combinatorial Optimization with Rational Objective Functions. *Mathematics of Operations Research*, 4(4):414–424, November 1979.

Joern Meissner and Arne Strauss. Improved bid prices for choice-based network revenue management. *European Journal of Operational Research*, 217(2):417–427, March 2012a.

Joern Meissner and Arne Strauss. Network revenue management with inventory-sensitive bid prices and customer choice. *European Journal of Operational Research*, 216(2):459–468, January 2012b.

Isabel Méndez-Díaz, Juan José Miranda-Bront, Gustavo Vulcano, and Paula Zabala. A branch-and-cut algorithm for the latent-class logit assortment problem. *Discrete Applied Mathematics*, 164:246–263, 2014.

J. Newman, M. Ferguson, L. Garrow, and T. Jacobs. Estimation of choice-based models using sales data from a single firm. *M&SOM*, 16(2):184–197, 2014.

R M Ratliff, V Rao, C P Narayan, and K Yellepeddi. A multi-flight recapture heuristic for estimating unconstrained demand from airline bookings. *Journal of Revenue and Pricing Management*, 7(2): 153–171, 2008.

P E Rossi and G M Allenby. Bayesian statistics and marketing. *Marketing Science*, 2003.

Paat Rusmevichientong and Huseyin Topaloglu. Robust Assortment Optimization in Revenue Management Under the Multinomial Logit Choice Model. *Operations Research*, 60(4):865–882, August 2012.

Paat Rusmevichientong, Z J Max Shen, and D B Shmoys. A PTAS for capacitated sum-of-ratios optimization. *Operations Research Letters*, 37(4):230–238, 2009.

Paat Rusmevichientong, Zuo-Jun Max Shen, and David B Shmoys. Dynamic Assortment Optimization with a Multinomial Logit Choice Model and Capacity Constraint. *Operations Research*, 58(6):1666–1680, December 2010.

Paat Rusmevichientong, David Shmoys, Chaoxu Tong, and Huseyin Topaloglu. Assortment optimization under the multinomial logit model with random choice parameters. *Production and Operations Management*, 23(11):2023–2039, 2014.

Alejandro A Schäffer. Simple local search problems that are hard to solve. *SIAM journal on Computing*, 20 (1):56–87, 1991.

K Talluri and G J van Ryzin. Revenue management under a general discrete choice model of consumer behavior. *Management Science*, 50(1):15–33, 2004a.

K T Talluri and G J van Ryzin. *The Theory and Practice of Revenue Management*. Springer Science+Business Media, 2004b.

K E Train. Discrete choice methods with simulation, 2009.

G J van Ryzin and G Vulcano. Computing Virtual Nesting Controls for Network Revenue Management Under Customer Choice Behavior. *Manufacturing & Service Operations Management*, 10(3):448–467, 2008.

Garrett van Ryzin and Gustavo Vulcano. A market discovery algorithm to estimate a general class of nonparametric choice models. *Management Science*, 61(2):281–300, 2014.

G Vulcano, G van Ryzin, and W Chaar. OM Practice—Choice-Based Revenue Management: An Empirical Study of Estimation and Optimization. *Manufacturing & Service Operations Management*, 12(3):371–392, 2010.

D Zhang. An improved dynamic programming decomposition approach for network revenue management. *Manufacturing & Service Operations Management*, 2011.

Dan Zhang and Daniel Adelman. An Approximate Dynamic Programming Approach to Network Revenue Management with Customer Choice. *Transportation Science*, 43(3):381–394, August 2009.

Appendix

A. Proofs for Section 4.1

Proof of Lemma 1 It follows from the definitions that

$$\begin{aligned}
0 > R(S+j) - R(S) &= p_j P_j(S+j) + \sum_{i\in S} p_i[P_i(S+j) - P_i(S)] \\
&= p_j P_j(S+j) - P_j(S+j) \sum_{i\in S} p_i P_i(S) \frac{P_i(S) - P_i(S+j)}{P_i(S)P_j(S+j)} \\
&\geq p_j P_j(S+j) - P_j(S+j)\left(\max_{i\in S} \frac{P_i(S) - P_i(S+j)}{P_i(S)P_j(S+j)}\right) \sum_{i\in S} p_i P_i(S) \\
&= p_j P_j(S+j) - P_j(S+j) X_j(S) R(S).
\end{aligned}$$

It thus follows that $0 > R(S+j) - R(S)$ implies that $p_j < X_j(S)R(S)$. In a similar fashion we can write

$$\begin{aligned}
R(S-j) - R(S) &= -p_j P_j(S) + \sum_{i\in S-j} p_i[P_i(S-j) - P_i(S)] \\
&= -p_j P_j(S) + P_j(S) \sum_{i\in S-j} p_i P_i(S-j) \frac{P_i(S-j) - P_i(S)}{P_i(S-j)P_j(S)} \\
&\geq -p_j P_j(S) + P_j(S)\left(\min_{i\in S-j} \frac{P_i(S-j) - P_i(S)}{P_i(S-j)P_j(S)}\right) \sum_{i\in S-j} p_i P_i(S-j) \\
&= -p_j P_j(S) + P_j(S) Y_j(S) R(S-j) \\
&= -p_j P_j(S) + P_j(S) Y_j(S) R(S) + P_j(S) Y_j(S)[R(S-j) - R(S)]
\end{aligned}$$

It thus follows that

$$(1 - P_j(S)Y_j(S))[R(S-j) - R(S)] \geq -p_j P_j(S) + P_j(S) Y_j(S) R(S).$$

Because LHS is < 0, it must be that $p_j > Y_j(S)R(S)$.

The result of the lemma now follows. □

Proof of Theorem 1 Let S^* denote the optimal solution and S a local optimum. There are two cases: (i) $S \setminus S^* \neq \varnothing$ and (ii) $S^* \subseteq S$.

In the first case, for any $j \in S^* \setminus S$, we must have $R(S+j) < R(S)$ (because S is a local optimum) and $R(S^* - j) < R(S^*)$ (because S^* is the global optimum). It thus follows from Lemma 1 that $Y_j(S^*)R(S^*) < p_j < X_j(S)R(S)$, which implies that $R(S)/R(S^*) \geq Y_j(S^*)/X_j(S) \geq \lambda_2/\lambda_1$.

In the second case, let $\Delta \stackrel{\text{def}}{=} S \setminus S^*$. Because S is a local optimum, for any $j \in S$, we have that $R(S-j) < R(S)$, which implies according to Lemma 1 that

$$p_i > Y_i(S)R(S) \;\forall\; i \in S \tag{5}$$

Similarly, because S^* is the global optimum, it follows from Lemma 1 that

$$p_j > Y_j(S)R(S) \;\forall\; j \in S^*. \tag{6}$$

Therefore, it follows from (5) that

$$R(S) = \sum_{j\in S} p_j P_j(S) > R(S) \sum_{j\in S} Y_j(S) P_j(S) \geq \lambda_2 R(S) P_S(S) \implies P_S(S)\lambda_2 < 1 \tag{7}$$

We can now write

$$R(S) = \sum_{j \in S^*} p_j P_j(S) + \sum_{i \in \Delta} p_i P_i(S) > R(S^*) \sum_{i \in S^*} Y_j(S) P_j(S) + R(S) \sum_{i \in \Delta} Y_i(S) P_i(S)$$
$$\geq \lambda_2 R(S^*) P_{S^*}(S) + \lambda_2 R(S) P_\Delta(S),$$

where the first inequality follows by invoking (6) for the first summation and (5) for the second summation. Through a re-arrangement and using the fact that $1 - \lambda_2 P_\Delta(S) > 0$ (because of (7)), we now obtain

$$R(S)/R(S^*) \geq \frac{\lambda_2 P_{S^*}(S)}{1 - \lambda_2 P_\Delta(S)} \geq \lambda_2 P_{S^*}(S) \geq \lambda_2 P_{S^*}(N),$$

For the second case, we can also bound the locality as follows: where the inequality

$$\begin{aligned}
R(S^*) - R(S) &= \sum_{j \in S^*} p_j [P_j(S^*) - P_j(S)] - \sum_{i \in \Delta} p_i P_i(S) \\
&\leq \max_{j \in S^*} \frac{P_j(S^*) - P_j(S)}{P_j(S^*)} \sum_{j \in S^*} p_j P_j(S^*) - R(S) \sum_{i \in \Delta} Y_i(S) P_i(S) \\
&= \max_{j \in S^*} \frac{P_j(S^*) - P_j(S)}{P_j(S^*) P_\Delta(S)} P_\Delta(S) R(S^*) - R(S) \sum_{i \in \Delta} Y_i(S) P_i(S) \\
&= X_\Delta(S^*) P_\Delta(S) R(S^*) - R(S) \sum_{i \in \Delta} Y_i(S) P_i(S) \\
&\leq \lambda_1 P_\Delta(S) R(S^*) - R(S) \lambda_2 \sum_{i \in \Delta} P_i(S) \\
&= \lambda_1 P_\Delta(S) R(S^*) - R(S) \lambda_2 P_\Delta(S).
\end{aligned}$$

where the first inequality follows from (5). It now follows through a re-arrangement by using the fact that $1 - \lambda_2 P_\Delta(S) > 0$ that

$$R(S)/R(S^*) \geq \frac{1 - \lambda_1 P_\Delta(S)}{1 - \lambda_2 P_\Delta(S)}.$$

We claim that

$$\frac{1 - \lambda_1 P_\Delta(S)}{1 - \lambda_2 P_\Delta(S)} \geq \frac{1 - \lambda_1 P_N(N)}{(1 - \lambda_2 P_N(N))^+}. \tag{8}$$

To see this, we consider two cases: (a) $\lambda_2 P_N(N) \geq 1$ and (b) $P_N(N) < 1$. In the first case, because $\lambda_1 \geq \lambda_2$, we must have that $1 - \lambda_2 P_N(N) < 0$. Therefore, the term on the RHS of (8) is $-\infty$ and the inequality is immediately satisfied. In the second case, invoking that $x \mapsto \frac{1-\lambda_1 x}{1-\lambda_2 x}$ over the interval $[0, 1/\lambda_2)$ is non-increasing in x because $\lambda_2 \leq \lambda_1$, we can obtain a lower bound for LHS by replacing $P_\Delta(S)$ by something larger. We now note that $P_\Delta(S) \leq P_S(S) = 1 - P_0(S) \leq 1 - P_0(N) = P_N(N)$, where the second equality follows by weak rationality of choice probabilities so that $P_0(S) \geq P_0(N)$. The claim in (8) thus follows.

We have thus shown that in the case when $S^* \subseteq S$, we must have

$$R(S)/R(S^*) \geq \max\left\{\lambda_2 P_{S^*}(N), \frac{1 - \lambda_1 P_N(N)}{(1 - \lambda_2 P_N(N))^+}\right\}.$$

Because at least one of the two cases must be satisfied, we must have that

$$R(S)/R(S^*) \geq \min\left\{\frac{\lambda_2}{\lambda_1}, \max\left\{\lambda_2 P_{S^*}(N), \frac{1 - \lambda_1 P_N(N)}{(1 - \lambda_2 P_N(N))^+}\right\}\right\}.$$

Because the above bound holds for any local optimum S, it is also a lower bound for the locality ratio. The result of the theorem now follows. □

Proof of Theorem 2 For any subset S and product $i \in S$, the choice probability under the MNL model is give by $P_i(S) = v_i / \left(v_0 + \sum_{j \in S} v_j\right)$. We can now write for any subset S, $i \in S$, and $\Delta \subseteq S$

$$\begin{aligned} P_i(S) - P_i(S+\Delta) = \frac{v_i}{v_0 + \sum_{j \in S} v_j} - \frac{v_i}{v_0 + \sum_{j \in S+\Delta} v_j} &= v_i \frac{\left(v_0 + \sum_{j \in S+\Delta} v_j\right) - \left(v_0 + \sum_{j \in S} v_j\right)}{\left(v_0 + \sum_{j \in S+\Delta} v_j\right)\left(v_0 + \sum_{j \in S} v_j\right)} \\ &= \frac{v_i \left(\sum_{j \in \Delta} v_j\right)}{\left(v_0 + \sum_{j \in S+\Delta} v_j\right)\left(v_0 + \sum_{j \in S} v_j\right)} \\ &= \frac{v_i}{v_0 + \sum_{j \in S} v_j} \left(\sum_{j \in \Delta} \frac{v_j}{v_0 + \sum_{j' \in S+\Delta} v_{j'}}\right) \\ &= P_i(S) P_\Delta(S+\Delta). \end{aligned}$$

It thus follows that

$$\lambda_1 = \max_{\Delta \subseteq S \subseteq N} X_\Delta(S) = \max_{\Delta \subseteq S \subseteq N} \max_{i \in S} \frac{P_i(S) - P_i(S+\Delta)}{P_i(S) P_\Delta(S+\Delta)} = 1.$$

Similarly, for any $j, i \in S$, we obtain that $P_i(S-j) - P_i(S) = P_i(S-j)P_j(S)$ by replacing S by $S-j$ and Δ by $\{j\}$ in the above set of equalities. We thus have

$$\lambda_2 = \min_{j \in S, S \subseteq N} Y_j(S) = \min_{j \in S, S \subseteq N} \min_{i \in S} \frac{P_i(S-j) - P_i(S)}{P_i(S-j) P_j(S)} = 1.$$

The locality ratio for the MNL model is thus 1. The result of the theorem now follows. □

Proof of Theorem 3 The choice probability under the RPL model is given by

$$P_i(S) = \mathbb{E}\left[L_i(S, \boldsymbol{Z})\right], \text{ where } L_i(S, \boldsymbol{Z}) = \frac{e^{\mu_i + Z_i}}{1 + \sum_{j \in S} e^{\mu_j + Z_j}}$$

and the expectation is with respect to $\boldsymbol{Z}$. Recall that Z_i is sub-Gaussian with variance proxy parameter $\sigma_i > 0$ so that $\mathbb{E}\left[e^{tZ_i}\right] \le e^{t^2 \sigma_i^2/2}$, for any $t \in \mathbb{R}$. Furthermore, $\sigma^2 = \frac{1}{n} \sum_{i=1}^n \sigma_i^2$.

Let $L_\Delta(S, \boldsymbol{Z})$ denote $\sum_{i \in \Delta} L_i(S, \boldsymbol{Z})$. It follows from our definitions that $L_i(S, \boldsymbol{Z}) - L_i(S+\Delta, \boldsymbol{Z}) = L_i(S, \boldsymbol{Z}) L_\Delta(S+\Delta, \boldsymbol{Z})$. We can now write

$$\frac{P_i(S) - P_i(S+\Delta)}{P_i(S) P_\Delta(S+\Delta)} = \frac{\mathbb{E}\left[L_i(S, \boldsymbol{Z}) - L_i(S+\Delta, \boldsymbol{Z})\right]}{\mathbb{E}\left[L_i(S, \boldsymbol{Z})\right] \cdot \mathbb{E}\left[L_\Delta(S+\Delta, \boldsymbol{Z})\right]} = \frac{\mathbb{E}\left[L_i(S, \boldsymbol{Z}) \cdot L_\Delta(S+\Delta, \boldsymbol{Z})\right]}{\mathbb{E}\left[L_i(S, \boldsymbol{Z})\right] \cdot \mathbb{E}\left[L_\Delta(S+\Delta, \boldsymbol{Z})\right]}.$$

We obtain upper and lower bounds for the ratio above by obtaining upper and lower bounds for the expectations. For that, we proceed as follows:

1. *Expectation bounds.* We obtain the lower and upper bounds for the expectation $\mathbb{E}\left[g(\boldsymbol{Z})\right]$:

$$e^{-\sigma\sqrt{5nc}} g(0)(1 - \varepsilon/6) \le \mathbb{E}\left[g(\boldsymbol{Z})\right] \le e^{\sigma\sqrt{5nc}} g(0)(1 + \varepsilon/6), \tag{9}$$

where ε and c are as defined in the hypothesis of the theorem and $g \colon \mathbb{R}^n \to [0,1]$ is any function with the property that

$$\gamma \le g(0) \text{ and } e^{-\rho} g(0) \le g(\boldsymbol{z}) \le e^{\rho} g(0) \text{ for all } \boldsymbol{z} \in I_\rho, \text{ where } I_\rho = \{\boldsymbol{z} \in \mathbb{R}^n \colon \|\boldsymbol{z}\| \le \rho\} \text{ for any } \rho > 0. \tag{10}$$

2. *Property satisfied.* We show that $L_i(S, \boldsymbol{Z})$, $L_\Delta(S+\Delta, \boldsymbol{Z})$, and $L_i(S, \boldsymbol{Z}) \cdot L_\Delta(S+\Delta, \boldsymbol{Z})$ satisfy (10).

Assuming that the two statements above are true, we complete the proof of the theorem; we prove the statements below. Because the bounds in (9) are true for the functions $L_i(S,\boldsymbol{Z})$, $L_\Delta(S+\Delta,\boldsymbol{Z})$, and $L_i(S,\boldsymbol{Z})\cdot L_\Delta(S+\Delta,\boldsymbol{Z})$, we must have that

$$\frac{P_i(S)-P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)}=\frac{\mathbb{E}\left[L_i(S,\boldsymbol{Z})\cdot L_\Delta(S+\Delta,\boldsymbol{Z})\right]}{\mathbb{E}\left[L_i(S,\boldsymbol{Z})\right]\cdot\mathbb{E}\left[L_\Delta(S+\Delta,\boldsymbol{Z})\right]}\geq\frac{e^{-\sigma\sqrt{5nc}}(1-\varepsilon/6)L_i(S,0)\cdot L_\Delta(S+\Delta,0)}{e^{2\sigma\sqrt{5nc}}(1+\varepsilon/6)^2L_i(S,0)\cdot L_\Delta(S+\Delta,0)}\geq e^{-3\sigma\sqrt{5nc}}(1-\varepsilon/2),$$

where the last inequality follows from noting that $(1+\varepsilon/6)^{-2}\geq 1-\varepsilon/3$ and $(1-\varepsilon/6)(1-\varepsilon/3)\geq 1-\varepsilon/2$.

In a similar fashion, we obtain the upper bound

$$\frac{P_i(S)-P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)}=\frac{\mathbb{E}\left[L_i(S,\boldsymbol{Z})\cdot L_\Delta(S+\Delta,\boldsymbol{Z})\right]}{\mathbb{E}\left[L_i(S,\boldsymbol{Z})\right]\cdot\mathbb{E}\left[L_\Delta(S+\Delta,\boldsymbol{Z})\right]}\leq\frac{e^{3\sigma\sqrt{5nc}}}{1-\varepsilon/2}.$$

We have thus obtained the bounds

$$\lambda_2\geq e^{-3\sigma\sqrt{5nc}}(1-\varepsilon/2)\text{ and }\lambda_2\leq\frac{e^{3\sigma\sqrt{5nc}}}{1-\varepsilon/2}.$$

Putting everything together, we obtain that

$$\frac{\lambda_2}{\lambda_2}\geq e^{-6\sigma\sqrt{5nc}}(1-\varepsilon/2)^2\geq e^{-6\sigma\sqrt{5nc}}(1-\varepsilon).$$

Now suppose that $P_N(N)\leq\frac{1}{\lambda_1+\lambda_2}$. We must then have that

$$\frac{1-\lambda_1P_N(N)}{(1-\lambda_2P_N(N))^+}\geq\frac{1-\lambda_1P_N(N)}{1-\lambda_2P_N(N)}\geq\frac{1-\lambda_1/(\lambda_1+\lambda_2)}{1-\lambda_2/(\lambda_1+\lambda_2)}=\frac{\lambda_2}{\lambda_1},$$

where the first inequality follows from the fact that $x\mapsto\frac{1-\lambda_1x}{1-\lambda_2x}$ is non-increasing in x for $lambda_2\leq\lambda_1$. Similarly, suppose $\min_{i\in N}P_i(N)\geq 1/\lambda_1$. Then, we have

$$\lambda_2P_{S^*}(N)\geq\lambda_2\min_{i\in S^*}P_i(N)\geq\frac{\lambda_2}{\lambda_1},$$

where S^* is the optimal solution. We have thus shown that when the conditions of the theorem are satisfied, we must have

$$\max\left\{\lambda_2P_{S^*}(N),\frac{1-\lambda_1P_N(N)}{1-\lambda_2P_N(N)}\right\}\geq\frac{\lambda_2}{\lambda_1}.$$

We can now conclude that the locality ratio is bounded below by

$$\min\left\{\frac{\lambda_2}{\lambda_1},\max\left\{\lambda_2P_{S^*}(N),\frac{1-\lambda_1P_N(N)}{(1-\lambda_2P_N(N))^+}\right\}\right\}\geq\frac{\lambda_2}{\lambda_1}\geq e^{-6\sigma\sqrt{5nc}}(1-\varepsilon).$$

We are now left with establishing the two statements above.

Expectation bounds. Suppose $g\colon\mathbb{R}^n\to[0,1]$ is a function that satisfies (10). Set $\rho=\sigma\sqrt{5nc}$. We must have

$$\int_{\boldsymbol{z}\in I_\rho}g(\boldsymbol{z})\phi(\boldsymbol{z})d\boldsymbol{z}\leq\mathbb{E}\left[g(\boldsymbol{z})\right]=\int_{\boldsymbol{z}\in I_\rho}g(\boldsymbol{z})\phi(\boldsymbol{z})d\boldsymbol{z}+\int_{\boldsymbol{z}\notin I_\rho}g(\boldsymbol{z})\phi(\boldsymbol{z})d\boldsymbol{z}\leq\int_{\boldsymbol{z}\in I_\rho}g(\boldsymbol{z})\phi(\boldsymbol{z})d\boldsymbol{z}+\int_{\boldsymbol{z}\notin I_\rho}\phi(\boldsymbol{z})d\boldsymbol{z},\quad(11)$$

where the left inequality follows because $g(\boldsymbol{z})\geq 0$ for all $\boldsymbol{z}$ and the right inequality follows from the fact that $g(\boldsymbol{z})\leq 1$ for all $\boldsymbol{z}$. It now follows from (10) that

$$e^{-\sigma\sqrt{5nc}}g(0)\mathbb{P}(\boldsymbol{z}\in I_\rho)\leq\mathbb{E}\left[g(\boldsymbol{Z})\right]\leq e^{\sigma\sqrt{5nc}}g(0)\mathbb{P}(\boldsymbol{z}\in I_\rho)+\mathbb{P}(\boldsymbol{z}\notin I_\rho).\quad(12)$$

To bound the probabilities, we invoke the tail inequality for quadratic forms of sub-gaussian random variables. In particular, we have the following inequality from Hsu et al. (2012):

$$\mathbb{P}\left(\|\boldsymbol{z}\|^2 > \sigma^2(n + 2\sqrt{nt} + 2t)\right) \leq e^{-t} \text{ for any } t > 0,$$

whenever $\boldsymbol{z}$ is the random sample from the independent and sub-Gaussian random random variables $(Z_i)_{i=1}^n$. Now choosing $t = c\sqrt{n}$, we note that

$$n + 2\sqrt{nt} + 2t = n + 2n\sqrt{c} + 2c\sqrt{n} \leq nc + 2nc + 2nc = 5nc,$$

where the inequality follows from $\sqrt{c} \leq c$ and $\sqrt{n} \leq n$ because both n and c are larger than 1. We can now write

$$\mathbb{P}(\boldsymbol{z} \notin I_\rho) = \mathbb{P}(\|\boldsymbol{z}\|^2 > 5cn\sigma^2) \leq \mathbb{P}\left(\|z\|^2 > \sigma^2(n + 2\sqrt{nt} + 2t)\right) \leq e^{-t} = e^{-c\sqrt{n}}.$$

Now because $c \geq \frac{1}{\sqrt{n}} \log(6/(\varepsilon\gamma))$, we have $e^{-c\sqrt{n}} \leq e^{-\log(6/(\varepsilon\gamma))} \leq \varepsilon\gamma/6$. We have thus obtained

$$\mathbb{P}(\boldsymbol{z} \notin I_\rho) \leq e^{-c\sqrt{n}} \leq \frac{\varepsilon\gamma}{6} \text{ and } \mathbb{P}(\boldsymbol{z} \in I_\rho) = 1 - \mathbb{P}(\boldsymbol{z} \notin I_\rho) \geq 1 - \frac{\varepsilon\gamma}{6}.$$

We can now write from (12) that

$$e^{-\sigma\sqrt{5nc}} g(0)(1 - \varepsilon/6) \leq e^{-\sigma\sqrt{5nc}} g(0)(1 - \varepsilon\gamma/6) \leq e^{-\sigma\sqrt{5nc}} g(0)\mathbb{P}(\boldsymbol{z} \in I_\rho) \leq \mathbb{E}\left[g(\boldsymbol{Z})\right],$$

where the first inequality follows because $\gamma < 1$, and

$$\mathbb{E}\left[g(\boldsymbol{Z})\right] \leq e^{\sigma\sqrt{5nc}} g(0)\mathbb{P}(\boldsymbol{z} \in I) + \mathbb{P}(\boldsymbol{z} \notin I) \leq e^{\sigma\sqrt{5nc}} g(0) + \varepsilon\gamma/6 = e^{\sigma\sqrt{5nc}} g(0)\left(1 + \frac{\varepsilon}{6} \cdot \frac{\gamma}{e^{\sigma\sqrt{5nc}} g(0)}\right) \leq e^{\sigma\sqrt{5nc}} g(0)(1 + \varepsilon/6)$$

where the last inequality follows because $e^{\sigma\sqrt{5nc}} > 1$ and $\gamma \leq g(0)$.

We have thus established the first statement.

Property satisfied. We now establish that $L_i(S, \boldsymbol{Z})$, $L_\Delta(S + \Delta, \boldsymbol{Z})$, and $L_i(S, \boldsymbol{Z}) \cdot L_\Delta(S + \Delta, \boldsymbol{Z})$ satisfy (10). We first focus on $L_i(S, \boldsymbol{Z})$. Let the function $h(\boldsymbol{\theta})$ denote the function $\log L_i(S, \boldsymbol{z}) = \mu_i + \theta_i - \log\left(1 + \sum_{j \in S} e^{\mu_j} e^{\theta_j}\right)$. Because the logsumexp function is convex, it is clear that $h(\cdot)$ is a concave function. Further, it may be verified that the curvature of $h(\cdot)$ is bounded.

We obtain upper and lower bounds for $h(\cdot)$ over the relaxed ℓ_1-ball: $\{\boldsymbol{\theta}\colon \sum_{k=1}^n |\theta_k| \leq \rho\}$. We claim that h is maximized at $\rho\boldsymbol{e}_i$ and minimized at $-\rho\boldsymbol{e}_i$, where $\boldsymbol{e}_i$ is the n-dimensional unit vector taking value 1 in the ith coordinate and 0 elsewhere. To establish the claim we first note that for any $j \in S, j \neq i$ and $\boldsymbol{\theta}$, we have $\partial h/\partial\theta_i > 0 > \partial h/\partial\theta_j$.

Now consider any $\boldsymbol{\theta}$ in the ℓ_1-ball such that $\theta_i \in (-\rho, \rho)$. We claim that h will not attain maximum or minimum over the ℓ_1-ball at such a $\boldsymbol{\theta}$. We consider two cases: (a) $\theta_j = 0$ for all $j \in S, j \neq i$ and (b) there exists a $j \in S, j \neq i$ such that $\theta_j \neq 0$. In the first case, because $\partial h/\partial\theta_i > 0$, we can decrease (increase) θ_i by some $\varepsilon > 0$ such that $L_i(S, \cdot)$ strictly decreases (increases) while remaining inside the ℓ_1-ball (which can be achieved, if required, by adjusting θ_k for some $\theta_k \neq 0$ and $k \notin S$). Therefore, h cannot attain the maximum or minimum value at such a $\boldsymbol{\theta}$. In the second case, again there exists an $\varepsilon > 0$ such that $\boldsymbol{\theta}^- \stackrel{\text{def}}{=} \boldsymbol{\theta} - \varepsilon\boldsymbol{e}_i + \text{sgn}(\theta_i\theta_j)\varepsilon\boldsymbol{e}_j$ is inside the ℓ_1-ball, where $\text{sgn}(x) = 1$ if $x > 0$ and -1 if $x \leq 0$. Because the curvature of h is bounded, for

ε small enough, we must have $h(\boldsymbol{\theta}^-) - h(\boldsymbol{\theta}) = \varepsilon\,(\text{sgn}(\theta_i\theta_j)\partial h/\partial\theta_j - \partial h/\partial\theta_i) + O(\varepsilon^2) < 0$. Similarly, taking $\boldsymbol{\theta}^+ = \boldsymbol{\theta} + \varepsilon\boldsymbol{e}_i - \text{sgn}(\theta_i\theta_j)\varepsilon\boldsymbol{e}_j$, we have $h(\boldsymbol{\theta}^+) - h(\boldsymbol{\theta}) = \varepsilon\,(-\text{sgn}(\theta_i\theta_j)\partial h/\partial\theta_j + \partial h/\partial\theta_i) + O(\varepsilon^2) > 0$. Therefore, h does not attain maximum or minimum at $\boldsymbol{\theta}$. It can now be verified that h attains the maximum value at $\theta_i = \rho$ and the minimum value at $\theta_i = -\rho$.

We can thus conclude that for any $\boldsymbol{z} \in \mathbb{R}^n$ such that $\|\boldsymbol{z}\|_1 \le \rho$, we have

$$L_i(S,\boldsymbol{Z}) \le \frac{e^{\mu_i+\rho}}{1 + e^{\mu_i+\rho} + \sum_{j\in S, j\neq i} e^{\mu_j}} \le e^{\rho}\frac{e^{\mu_i}}{1+\sum_{j\in S} e^{\mu_j}} = e^{\rho}L_i(S,0).$$

We can similarly show that for any $\boldsymbol{z}$ in the ℓ_1-ball $\|\boldsymbol{z}\|_1 \le \rho$, we have

$$L_i(S,\boldsymbol{Z}) \ge e^{-\rho}\frac{e^{\mu_i}}{1+\sum_{j\in S} e^{\mu_j}} = e^{-\rho}L_i(S,0).$$

Using a symmetric argument, we can also show that

$$e^{-\rho}L_k(S+\Delta,0) \le L_k(S+\Delta,\boldsymbol{z}) \le e^{\rho}L_k(S+\Delta,0),$$

by summing which over $k \in \Delta$, we obtain

$$e^{-\rho}L_\Delta(S+\Delta,0) \le L_\Delta(S+\Delta,\boldsymbol{z}) \le e^{\rho}L_\Delta(S+\Delta,0),$$

The fact that $\gamma \le L_\Delta(S+\Delta,0)$ for any $\Delta \subseteq S$ follows from the definition of γ.

We have thus established (10) for $L_i(S,\boldsymbol{z})$ and $L_\Delta(S,\boldsymbol{z})$. Now consider $L_i(S,\boldsymbol{z})\cdot L_k(S+\Delta,\boldsymbol{z})$ for some $k \in \Delta$. As above, let $f(\boldsymbol{\theta})$ denote $\log\left(L_i(S,\boldsymbol{\theta})\cdot L_k(S+\Delta,\boldsymbol{\theta})\right) = \mu_i + \mu_j + \theta_i + \theta_k - \log\left(1+\sum_{j\in S} e^{\mu_j}e^{\theta_j}\right) - \log\left(1+\sum_{j\in S+\Delta} e^{\mu_j}e^{\theta_j}\right)$. As above, $f(\cdot)$ is concave with a bounded curvature. It may be verifed that $\partial f/\partial\theta_k > 0 > \partial f/\partial\theta_j$ for any $j \in S+\Delta$, $j \neq i,k$.

We consider two cases: (a) $\partial f/\partial\theta_i < 0$ and $\partial f/\partial\theta_i \ge 0$.

In the first case, using an argument similar to the above, we can show that f does not attain maximum of minimum over the ℓ_1-ball at any $\boldsymbol{\theta}$ such that $\theta_j \neq 0$ for any $j \neq k$. Further, the optimum won't occur if $|\theta_k| < \rho$. Therefore, we have that maximum occurs at $\theta_k = \rho$ and the minimum occurs at $\theta_k = -\rho$.

We now consider the second case. First note again that f does not attain maximum of minimum over the ℓ_1-ball at any $\boldsymbol{\theta}$ such that $\theta_j \neq 0$ for any $j \neq k,i$. Therefore, at optimality, we must have that $\theta_j = 0$ for all $j \neq k,i$. The optimal solution depends on the relation between $\partial f/\partial\theta_i$ and $\partial f/\partial\theta_k$. If $\partial f/\partial\theta_k > \partial f/\partial\theta_i$, we can use the arguments above to show that the maximum occurs at $\theta_k = \rho$ and the minimum occurs at $\theta_k = -\rho$. On the other hand, if $\partial f/\partial\theta_k < \partial f/\partial\theta_i$, the maximum occurs at $\theta_i = \rho$ and the minimum occurs at $\theta_i = -\rho$.

In each of the cases, it may be verified that $e^{-\rho}L_i(S,0)\cdot L_k(S+\Delta,0) \le L_i(S,\boldsymbol{z})\cdot L_k(S+\Delta,\boldsymbol{z}) \le e^{\rho}L_i(S,0)\cdot L_k(S+\Delta,0)$ for any $\boldsymbol{z}$ in the ℓ_1-ball $\|\boldsymbol{z}\|_1 \le \rho$. Summing the inequalities over $k \in \Delta$ yields:

$$e^{-\rho}L_i(S,0)\cdot L_\Delta(S+\Delta,0) \le L_i(S,\boldsymbol{z})\cdot L_\Delta(S+\Delta,\boldsymbol{z}) \le e^{\rho}L_i(S,0)\cdot L_\Delta(S+\Delta,0)$$

The fact that $\gamma \le L_i(S,\boldsymbol{z})\cdot L_\Delta(S+\Delta,\boldsymbol{z})$ follows by definition of γ. We have thus established that (10) is satisfied by $L_i(S,\boldsymbol{z})\cdot L_\Delta(S+\Delta,\boldsymbol{z})$. This finishes the proof of the second statement.

The result of the theorem now follows. $\square$

Proof of Theorem 4 To simplify notation, let $v_0 \stackrel{\text{def}}{=} 1 + V_m^\eta$. It then follows that for any $i \in S$ and $\Delta \subseteq N$ such that $\Delta \cap S = \varnothing$,

$$\begin{aligned}
P_i(S) - P_i(S+\Delta) &= v_i \frac{V(S)^{\rho-1}}{v_0 + V(S)^\rho} - v_i \frac{V(S+\Delta)^{\rho-1}}{v_0 + V(S+\Delta)^\rho} \\
&= v_i \frac{v_0\left(V(S+\Delta)^{\rho-1} - V(S)^{\rho-1}\right) + V(S)^{\rho-1}V(S+\Delta)^{\rho-1}(V(S+\Delta)-V(S))}{(v_0 + V(S)^\rho)(v_0 + V(S+\Delta)^\rho)} \\
&= \frac{v_i V(S)^{\rho-1}}{v_0 + V(S)^\rho} \left(\sum_{j\in\Delta} \frac{v_j V(S+\Delta)^{\rho-1}}{v_0 + V(S+\Delta)^\rho} \right) \left[1 + v_0 \frac{V(S+\Delta)^{\rho-1} - V(S)^{\rho-1}}{V(S)^{\rho-1}V(S+\Delta)^{\rho-1}(V(S+\Delta)-V(S))} \right] \\
&= P_i(S) P_\Delta(S+\Delta) \left[1 + v_0 \frac{V(S+\Delta)^{1-\rho} - V(S)^{1-\rho}}{V(S+\Delta) - V(S)} \right].
\end{aligned}$$

We thus have

$$\frac{P_i(S) - P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)} = 1 + v_0 \frac{V(S+\Delta)^{1-\rho} - V(S)^{1-\rho}}{V(S+\Delta) - V(S)} \tag{13}$$

Similarly,

$$\begin{aligned}
P_m(S) - P_m(S+\Delta) &= \frac{v_m V_m^{\eta-1}}{v_0 + V(S)^\rho} - \frac{v_m V_m^{\eta-1}}{v_0 + V(S+\Delta)^\rho} \\
&= \frac{v_m V_m^{\eta-1}}{(v_0 + V(S)^\rho)(v_0 + V(S+\Delta)^\rho)} \cdot (V(S+\Delta)^\rho - V(S)^\rho) \\
&= P_m(S) \frac{V(S+\Delta) - V(S)}{v_0 + V(S+\Delta)^\rho} \frac{V(S+\Delta)^\rho - V(S)^\rho}{V(S+\Delta) - V(S)} \\
&= P_m(S) \left(\sum_{j\in\Delta} \frac{v_j V(S+\Delta)^{\rho-1}}{v_0 + V(S+\Delta)^\rho} \right) \cdot \frac{1 - V(S)^\rho / V(S+\Delta)^\rho}{1 - V(S)/V(S+\Delta)} \\
&= P_m(S) P_\Delta(S+\Delta) \cdot \frac{1 - V(S)^\rho / V(S+\Delta)^\rho}{1 - V(S)/V(S+\Delta)}.
\end{aligned}$$

It thus follows that

$$\frac{P_m(S) - P_m(S+\Delta)}{P_m(S)P_\Delta(S+\Delta)} = \frac{1 - V(S)^\rho / V(S+\Delta)^\rho}{1 - V(S)/V(S+\Delta)}. \tag{14}$$

We can bound the terms in (13) and (14) as follows. Because $\rho \in (0,1]$, using the fact that $x \mapsto x^{1-\rho}$ is concave, which by the sub-gradient inequality at 1 implies that $x^{1-\rho} \le 1 + (1-\rho)(x-1)$, we have

$$(V(S+\Delta)/V(S))^{1-\rho} \le 1 + (1-\rho)\frac{V(S+\Delta) - V(S)}{V(S)} \implies \frac{V(S+\Delta)^{1-\rho} - V(S)^{1-\rho}}{V(S+\Delta) - V(S)} \le (1-\rho)V(S)^{-\rho}.$$

Therefore, we have

$$1 \le 1 + v_0 \frac{V(S+\Delta)^{1-\rho} - V(S)^{1-\rho}}{V(S+\Delta) - V(S)} \le 1 + (1-\rho) v_0 V(S)^{-\rho},$$

where the left inequality follows from $V(S+\Delta) > V(S)$. Similarly, noting that $x^\rho \le 1 + \rho(x-1)$, we must have

$$(V(S)/V(S+\Delta))^\rho \le 1 + \rho\,(V(S)/V(S+\Delta) - 1) \implies \frac{1 - V(S)^\rho / V(S+\Delta)^\rho}{1 - V(S)/V(S+\Delta)} \ge \rho.$$

Further, because $V(S)/V(S+\Delta) < 1$, we have that

$$\frac{1 - V(S)^\rho / V(S+\Delta)^\rho}{1 - V(S)/V(S+\Delta)} \le 1.$$

Putting everything together, we get

$$\begin{aligned}\lambda_1 = \max_{\Delta\subseteq S\subseteq N} X_\Delta(S) &= \max_{\Delta\subseteq S\subseteq N} \max\left\{\frac{P_m(S)-P_m(S+\Delta)}{P_m(S)P_\Delta(S+\Delta)}, \max_{i\in S}\frac{P_i(S)-P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)}\right\}\\ &\le \max_{\Delta\subseteq S\subseteq N}(1+(1-\rho)v_0V(S)^{-\rho})\\ &\le 1+(1-\rho)v_0/w_2\end{aligned}$$

In a similar fashion, we have

$$\lambda_2 = \min_{j\in S, S\subseteq N}\min\left\{\frac{P_m(S-j)-P_m(S)}{P_m(S-j)P_j(S)}, \min_{i\in S}\frac{P_i(S-j)-P_i(S)}{P_i(S-j)P_j(S)}\right\}\ge\rho.$$

We now compute the locality ratio using the values of λ_1 and λ_2 derived above. For that, we first note that following. Because $P_N(N)\le 1/(1+c)$, we have

$$P_N(N)\le\frac{1}{2(1+c)} = \frac{1}{2\left(1+\frac{(1-\rho)\cdot(1+v_0/w_2)}{\rho}\right)} = \frac{\rho}{2\cdot(1+(1-\rho)v_0/w_2)}\le\frac{1}{2\lambda_1}\le\frac{1}{\lambda_1+\lambda_2},$$

where the first inequality follows because $\rho\le 1$ and the second inequality follows because $\lambda_2\le\lambda_1$. Now, because $x\mapsto\frac{1-\lambda_1 x}{1-\lambda_2 x}$ is non-increasing in x, we can write

$$\frac{1-\lambda_1 P_N(N)}{1-\lambda_2 P_N(N)}\ge\frac{1-\lambda_1/(\lambda_1+\lambda_2)}{1-\lambda_2/(\lambda_1+\lambda_2))}=\frac{\lambda_2}{\lambda_1}.$$

We thus have

$$\min\left\{\frac{\lambda_2}{\lambda_1},\frac{1-\lambda_1 P_N(N)}{1-\lambda_2 P_N(N)}\right\}=\frac{\lambda_2}{\lambda_1}\ge\frac{\rho}{1+(1-\rho)v_0/w}=1/(1+c).$$

We can thus conclude that the locality ratio is lower bounded by $1/(1+c)$. The result of the theorem now follows. ☐

B. Proofs for Section 4.2

Proof of Lemma 2 In order to prove this lemma, we first need the following technical result.

CLAIM 1. *$R(S+j)>R(S)\implies p_j>x_j(S)R(S)$ for any $j\in S$ and $R(S-j)>R(S)\implies p_j<y_j(S)R(S)$ for any $j\in S$.*

Proof. The proof is similar to that of Lemma 1. In particular, it follows from the definitions that

$$\begin{aligned}0<R(S+j)-R(S) &= p_jP_j(S+j)+\sum_{i\in S}p_i[P_i(S+j)-P_i(S)]\\ &= p_jP_j(S+j)-P_j(S+j)\sum_{i\in S}p_iP_i(S)\frac{P_i(S)-P_i(S+j)}{P_i(S)P_j(S+j)}\\ &\le p_jP_j(S+j)-P_j(S+j)\left(\min_{i\in S}\frac{P_i(S)-P_i(S+j)}{P_i(S)P_j(S+j)}\right)\sum_{i\in S}p_iP_i(S)\\ &= p_jP_j(S+j)-P_j(S+j)x_j(S)R(S).\end{aligned}$$

It thus follows that $0<R(S+j)-R(S)$ implies that $p_j>x_j(S)R(S)$. In a similar fashion we can write

$$\begin{aligned}R(S-j)-R(S) &= -p_jP_j(S)+\sum_{i\in S-j}p_i[P_i(S-j)-P_i(S)]\\ &= -p_jP_j(S)+P_j(S)\sum_{i\in S-j}p_iP_i(S-j)\frac{P_i(S-j)-P_i(S)}{P_i(S-j)P_j(S)}\\ &\le -p_jP_j(S)+P_j(S)\left(\max_{i\in S-j}\frac{P_i(S-j)-P_i(S)}{P_i(S-j)P_j(S)}\right)\sum_{i\in S-j}p_iP_i(S-j)\\ &= -p_jP_j(S)+P_j(S)y_j(S)R(S-j)\\ &= -p_jP_j(S)+P_j(S)y_j(S)R(S)+P_j(S)y_j(S)[R(S-j)-R(S)]\end{aligned}$$

It thus follows that

$$(1 - P_j(S)y_j(S))[R(S - j) - R(S)] \geq -p_j P_j(S) + P_j(S)Y_j(S)R(S).$$

It thus follows that $0 > R(S - j) - R(S)$ implies that $p_j < y_j(S)R(S)$. The claim now follows. □

We prove the result in two steps:

1. When $b = \infty$, the iterations of the ADXOpt algorithm may be partitioned into consecutive sequence of additions followed by consecutive sequence of deletions before the algorithm terminates.
2. When $b = 1$, the algorithm terminates with a revenue-ordered subset that is a local optimum with respect to additions and deletions.

We prove each of the statements in turn.

Proof of statement 1. If there are no deletions, then the statement is trivially true. Therefore, we suppose that there is at least one deletion. Let t_1 denote the iteration in which the first deletion occurred. Let S_1 denote the assortment just before the first deletion and let i denote the deleted product. We prove the statement by contradiction. Suppose $t > t_1$ is the first iteration in which an addition happens after t_1. Suppose product j is added in iteration t. Let S denote the assortment just before addition. Given our assumptions, we have the sequence of inequalities: $R(S_1) < R(S_1 - i) \leq R(S) < R(S + j)$ because the revenue strictly increases in each iteration. Further, it must be that $S \subseteq (S_1 - i)$. We consider two cases: (i) product j was never added and (ii) product j was added, deleted, and then being added again.

Case (i): product j was never added. In this case, we can invoke Claim 1 to write $R(S+j) > R(S) \implies p_j > R(S)x_j(S)$. We also know that $R(S) > R(S_1)$ and $x_j(S) \geq X_j(S_1)$ (from the hypothesis of the lemma because $S \subseteq S_1$), so that $p_j > R(S)x_j(S) > R(S_1)X_j(S_1)$, which in turn implies according to Lemma 1 that $R(S_1 + j) > R(S_1)$. In other words, we have concluded that starting with subset S_1, adding product j to S_1 increases revenues. Since $b = \infty$, adding j is a feasible operation. Further, since we are prioritizing additions over deletions, product j must have been added in iteration t_1, contradicting the fact that i was deleted in iteration t_1. This finishes the first case.

Case (ii): product j was added, deleted and then added again. Suppose product j was deleted in iteration t_2 such that $t_1 < t_2 < t$. Let S_2 denote the subset just before the deletion of product j in iteration t_2. We must then have the following sequence of inequalities: $R(S_2) < R(S_2 - j) \leq R(S) < R(S + j)$. Furthermore, we have $S \subseteq (S_2 - j)$. Because $R(S) < R(S + j)$, we can invoke Claim 1 to get $p_j > x_j(S)R(S)$. Further, we know that $R(S) \geq R(S_2 - j)$ and $x_j(S) \geq X_j(S_2 - j)$ (from the hypothesis of the lemma because $S \subseteq (S_2 - j)$). Hence, we can write $p_j > x_j(S)R(S) \geq X_j(S_2 - j)R(S_2 - j)$, which in turn must imply according to Lemma 1 that $R(S_2) > R(S_2 - j)$. This contradicts the assumption that $R(S_2) < R(S_2 - j)$. This finishes the second case.

We have thus shown that the ADXOpt algorithm with $b = \infty$, results in a sequence of consecutive additions, followed by a sequence of consecutive deletions. In other words, a product, once deleted, will never be added again.

Proof of statement 2. Let $\hat{S}$ denote the assortment at the termination of the algorithm. We establish that $\hat{S}$ is a local optimum: (a) $R(\hat{S} - i) < R(\hat{S})$ for any $i \in \hat{S}$, (b) $R(\hat{S} + j) < R(\hat{S})$ for any $j \notin \hat{S}$, and (c) $R(\hat{S} + j - i) < R(\hat{S})$ for any $j \notin \hat{S}$, $i \in \hat{S}$. Because deletion is always a feasible operation, it immediately follows that $R(\hat{S}-i) < R(\hat{S})$ for any $i \in \hat{S}$ at termination. In a similar fashion, for a product j that was never added, we must have $R(\hat{S}+j) \leq R(\hat{S})$ because such an addition is feasible. Now consider a product j that was added and then removed. We know from statement 1 above that even when there is no limit on the number of times a product was removed (i.e., $b=\infty$), once product j is removed, no product will be added. Hence, we must have that $R(\hat{S}+j) \leq R(\hat{S})$. We can thus conclude that $\hat{S}$ is a local optimum with respect to both additions and deletions.

To prove that $\hat{S}$ is also a local optimum with respect to exchanges, we argue that whenever an exchange improves the revenues, then there must be at least one addition or deletion that improves the revenues. For that we show that if $R(S+j) \leq R(S)$, then $R(S+j-i) > R(S)$ must imply that $R(S-i) > R(S)$. Since $R(S+j) \leq R(S)$ and $R(S) < R(S+j-i)$, it must be that $R(S+j) < R(S+j-i)$. Hence, we can invoke Claim 1 to get $p_i < R(S+j) y_i(S+j)$. Now since $y_i(S+j) < Y_i(S)$ (from the hypothesis) and $R(S+j) \leq R(S)$, we can write $p_i < R(S+j) y_i(S+j) \leq R(S) Y_i(S)$, which in turn implies by Lemma 1 that $R(S-i) > R(S)$.

The result of the lemma now follows. $\square$

Proof of Theorem 5 The proof is straightforward for the BAM model. In particular, it follows from the proof of Theorem 2 that

$$\frac{P_i(S)-P_i(S+j)}{P_i(S)P_j(S+j)} = 1 \text{ for any } j \notin S, S \subseteq N \text{ and } \frac{P_i(S-j)-P_i(S)}{P_i(S-j)P_j(S)} = 1 \text{ for any } j \in S, S \subseteq N.$$

Therefore, it follows by definition that $X_j(S) = x_j(S) = 1$ for any $j \in S, S \subseteq N$ and $Y_j(S) = y_j(S) = 1$ for any $j \notin S, S \subseteq N$. Thus, the monotonicity condition is trivially satisfied.

Now consider the PC-NL model. It follows from the proof of Theorem 4 that

$$\frac{P_i(S)-P_i(S+j)}{P_i(S)P_i(S+j)} = 1 + v_0 \frac{V(S+\Delta)^{1-\rho} - V(S)^{1-\rho}}{V(S+\Delta) - V(S)} \text{ for any } j \notin S, S \subseteq N$$

$$\text{and } \frac{P_m(S-j)-P_m(S)}{P_m(S-j)P_j(S)} = \frac{1 - V(S)^{\rho}/V(S+\Delta)^{\rho}}{1 - V(S)/V(S+\Delta)} \text{ for any } j \in S, S \subseteq N,$$

where recall that $v_0 \stackrel{\text{def}}{=} 1 + V_m^{\eta}$ and i is a product in the third nest. Because both the ratios don't depend on i or m, it follows that $X_j(S) = x_j(S)$ for any $j \notin S$. By replacing S by $S-j$ in the above ratios, we can also conclude that $Y_j(S) = y_j(S)$ for any $j \in S$. Therefore, to prove the monotonicity condition, it is sufficient to show that both ratios above decrease as S is enlarged. More precisely, letting x denote $V(S)$ and $x+a$ denote $V(S+j)$, define the functions

$$g_1(x) = 1 + (v_0/a) \frac{1}{(x+a)^{\rho}} \cdot h_{1-\rho}(x) \text{ and } g_2(x) = \frac{1}{a} \cdot h_{\rho}(x), \text{ where } h_{\delta}(x) = (x+a)\left[1 - \left(\frac{x}{x+a}\right)^{\delta}\right],$$

for any $\delta \in (0,1)$. It can be shown that to prove the monotonicity property, it is sufficient to show that both g_1 and g_2 are decreasing in x. Further, because $x \mapsto 1/(x+a)^{\rho}$ is decreasing in x, it is sufficient to show that $h_{\delta}(x)$ is decreasing in x for any $\delta \in (0,1)$. For that, we take the derivative of h with respect to x:

$$h'(x) = 1 - \left(\frac{x}{x+a}\right)^{\delta} \cdot [1 + a\delta/x] = \left(\frac{x}{x+a}\right)^{\delta} \left[\left(\frac{x+a}{x}\right)^{\delta} - (1 + a\delta/x)\right].$$

To show that $h'(x) < 0$, we note that $x \mapsto x^{\delta}$ is concave because $\delta \in (0,1)$. Therefore, it follows from the sub-gradient inequality at $x=1$ that $x^{\delta} \leq 1 + \delta(x-1)$ for any x. Using this inequality for the function $x \mapsto ((x+a)/x)^{\delta}$, we have that

$$h'(x) \leq \left(\frac{x}{x+a}\right)^{\delta} \left[1 + \delta\left(\frac{x+a}{x} - 1\right) - (1 + a\delta/x)\right] = 0.$$

We have thus shown that h is decreasing, which implies that both g_1 and g_2 are decreasing. Therefore, the monotonicity condition is satisfied for the PC-NL model.

The fact that ADXOpt converges in $O(n^2)$ steps for $b=1$ and with only additions and deletions follows from the discussion below the description of the algorithm.

The result of the theorem now follows. $\square$

Proof of Theorem 6 Note that our results on locality ratios critically depend on the result of Lemma 1. Therefore, we start with establishing a similar result for a $(1+\alpha)-$approximate local optimum. Define

$$X^\alpha_\Delta(S) = \max_{i\in S} \frac{(1+\alpha)P_i(S) - P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)} \text{ for } \Delta \not\subseteq S \text{ and } Y^\alpha_j(S) = \min_{i\in S-j} \frac{P_i(S-j)-(1+\alpha)P_i(S)}{(1+\alpha)P_i(S-j)P_j(S)} \text{ for } j\in S.$$

Using arguments similar to those in the proof of Lemma 1, we can show that

$$R(S+j) < (1+\alpha)R(S) \implies p_j < X^\alpha_j R(S) \text{ for } j\notin S \text{ and } R(S-j) < (1+\alpha)R(S) \implies p_j > Y^\alpha_j(S)R(S).$$

Using the above equation and arguments similar to those in the proof of Theorem 1, we can show that the α-locality ratio is bounded below by

$$\min\left\{\frac{\lambda^\alpha_2}{\lambda^\alpha_1}, \max\left\{\lambda^\alpha_2 P_{S^*}(N), \frac{1-P_N(N)\lambda^\alpha_1}{(1-P_N(N)\lambda^\alpha_2)^+}\right\}\right\}, \text{ where } \lambda^\alpha_1 \stackrel{\text{def}}{=} \max_{\Delta,S\subseteq N} X^\alpha_\Delta(S), \lambda^\alpha_2 \stackrel{\text{def}}{=} \min_{j\in S,S\subseteq N} Y_j(S)^\alpha.$$

Therefore, in order to express the α-locality ratio in terms of the locality ratio, it sufficient to express $X^\alpha_\Delta(S)$ and $Y^\alpha_j(S)$ in terms of $X_\Delta(S)$ and $Y_j(S)$ respectively.

We now claim that $X^\alpha_\Delta(S) \le (1+c\alpha)X_\Delta(S)$ and $Y_j(S) \ge Y_j(S)/(1+c\alpha)$, where $c = \gamma/(\gamma-1-\alpha)$. To see this, note that

$$\begin{aligned}(1+\alpha)P_i(S) - P_i(S+\Delta) \le (1+c\alpha)\left[P_i(S) - P_i(S+\Delta)\right] &\iff c\alpha P_i(S+\Delta) \le (c-1)\alpha P_i(S)\\ &\iff \frac{P_i(S+\Delta)}{P_i(S)} \le \frac{c-1}{c} = \frac{1+\alpha}{\gamma}.\end{aligned}$$

The last inequality is true by definition of γ and therefore we must have

$$X^\alpha_\Delta(S) = \max_{i\in S}\frac{(1+\alpha)P_i(S)-P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)} \le (1+c\alpha)\max_{i\in S}\frac{P_i(S)-P_i(S+\Delta)}{P_i(S)P_\Delta(S+\Delta)} = (1+c\alpha)X_\Delta(S).$$

In a similar fashion, note that

$$\begin{aligned}\frac{P_i(S-j)-(1+\alpha)P_i(S)}{1+\alpha} \ge \frac{P_i(S-j)-P_i(S)}{1+c\alpha} &\iff \left[\frac{1}{1+\alpha} - \frac{1}{1+c\alpha}\right]P_i(S-j) \ge \left[1 - \frac{1}{1+c\alpha}\right]P_i(S)\\ &\iff \frac{P_i(S)}{P_i(S-j)} \le \frac{c-1}{c(1+\alpha)} = \frac{1}{\gamma}.\end{aligned}$$

Again, because the last inequality is true by definition, it follows that

$$Y^\alpha_j(S) = \min_{i\in S-j}\frac{P_i(S-j)-(1+\alpha)P_i(S)}{(1+\alpha)P_i(S-j)P_j(S)} \ge \frac{1}{1+c\alpha}\max_{i\in S}\frac{P_i(S-j)-P_i(S)}{P_i(S-j)P_j(S)} = \frac{1}{1+c\alpha}Y_j(S).$$

Further, it follows from our definitions that $c\alpha = \varepsilon/2$. We have thus shown that

$$\lambda^\alpha_1 \le (1+\varepsilon/2)\lambda_1 \text{ and } \lambda^\alpha_2 \ge \frac{\lambda_2}{1+\varepsilon/2} \ge (1-\varepsilon/2)\lambda_2.$$

We now argue that the $\alpha-$locality ratio is lower bounded by $(1-\varepsilon)$ times the lower bound for the locality ratio given in Theorem 1. For that, we consider two cases: (a) $P_N(N) \le \frac{1}{(1+\varepsilon/2)\lambda_1+\lambda_2)}$ and (b) $P_N(N) \ge \frac{1}{\lambda_1}$.

In the first case, noting that $P_N(N) \le 1/(\lambda_1+\lambda_2)$, we get

$$\frac{1-P_N(N)\lambda_1}{1-P_N(N)\lambda_2} \ge \frac{1-\frac{\lambda_1}{\lambda_1+\lambda_2}}{1-\frac{\lambda_2}{\lambda_1+\lambda_2}} = \frac{\lambda_2}{\lambda_1},$$

where the inequality follows from the fact that $x \mapsto \frac{1-\lambda_1 x}{1-\lambda_2 x}$ is decreasing in x. We also have $\lambda_2 P_{S^*}(N) \le \lambda_2 P_N(N) \le \frac{\lambda_2}{\lambda_1+\lambda_2} \le \frac{\lambda_2}{\lambda_1}$. Therefore, we must have that

$$\max\left\{\lambda_2 P_{S^*}, \frac{1-P_N(N)\lambda_1}{1-P_N(N)\lambda_2}\right\} = \frac{1-P_N(N)\lambda_1}{1-P_N(N)\lambda_2} \ge \frac{\lambda_2}{\lambda_1} \implies \text{locality ratio bound} = \frac{\lambda_2}{\lambda_1}.$$

Now note that

$$\frac{1}{(1+\varepsilon/2)\lambda_1+\lambda_2} \le \frac{1}{\lambda_1^\alpha+\lambda_2} \le \frac{1}{\lambda_1^\alpha+\lambda_2^\alpha},$$

where the second inequality follows because $\lambda_2^\alpha \le \lambda_2$. It now follows that the condition $P_N(N) \le \frac{1}{(1+\varepsilon/2)\lambda_1+\lambda_2}$ implies that $P_N(N) \le \frac{1}{\lambda_1^\alpha+\lambda_2^\alpha}$. Following the sequence of inequalities above, we can conclude that $\frac{1-P_N(N)\lambda_1^\alpha}{1-P_N(N)\lambda_2^\alpha} \ge \frac{\lambda_2^\alpha}{\lambda_1^\alpha}$. It thus follows that

$$\min\left\{\frac{\lambda_2^\alpha}{\lambda_1^\alpha}, \max\left\{\lambda_2^\alpha P_{S^*}(N), \frac{1-P_N(N)\lambda_1^\alpha}{(1-P_N(N)\lambda_2^\alpha)^+}\right\}\right\} \ge \frac{\lambda_2^\alpha}{\lambda_1^\alpha} \ge \frac{1-\varepsilon/2}{1+\varepsilon/2}\frac{\lambda_2}{\lambda_1} \ge (1-\varepsilon) \times \text{ locality ratio bound.}$$

Now consider the second case where $P_N(N) \ge \frac{1}{\lambda_1}$. In this case, it must be that $\frac{1-P_N(N)\lambda_1}{(1-P_N(N)\lambda_2)^+} \le 0$ and therefore

$$\text{locality ratio bound} = \min\left\{\frac{\lambda_2}{\lambda_1}, \lambda_2 P_{S^*}(N)\right\}.$$

We can now conclude that

$$\min\left\{\frac{\lambda_2^\alpha}{\lambda_1^\alpha}, \max\left\{\lambda_2^\alpha P_{S^*}(N), \frac{1-P_N(N)\lambda_1^\alpha}{(1-P_N(N)\lambda_2^\alpha)^+}\right\}\right\} \ge \min\left\{\frac{\lambda_2^\alpha}{\lambda_1^\alpha}, \lambda_2^\alpha P_{S^*}(N)\right\} \ge (1-\varepsilon)\min\left\{\frac{\lambda_2}{\lambda_1}, \lambda_2 P_{S^*}(N)\right\}.$$

We have thus shown that in both the cases,

$$\alpha - \text{locality ratio} \ge (1-\varepsilon) \times \text{locality ratio bound.}$$

We now argue that ADXOpt(α) converges to a $(1+\alpha)-$approximate solution in $O(\frac{1}{\alpha}\cdot\frac{p_{\max}}{r_{\max}})$ steps. To see this, note that after the first iteration, we must have $R(S^1) \ge \max_{i\in N} R(\{i\}) = r_{\max}$. In each subsequent iteration, the revenue increases at least by factor $(1+\alpha)$. Further, the revenue of every subset $R(S) \le p_{\max}$. It thus follows that if there are k steps until convergence, we must have $(1+\alpha)^k \le p_{\max}/r_{\max}$. It now follows that $k = O(\frac{1}{\alpha}\cdot\frac{p_{\max}}{r_{\max}})$.

The result of the theorem now follows. □

C. Proofs for Section 4.3.

Our objective is to show that ADXOpt can find the optimal solution to the following optimization problem:

$$\arg\max_{S\subseteq N:\,|S|\le C} \frac{\sum_{i\in S} p_i v_i}{v_0+\sum_{i\in S} v_i}. \tag{15}$$

For that, we need some additional notation. For any $z\in\mathbb{R}$ and any $j\in N$, let $h_j(z)$ denote $v_j(p_j-z)$. Further, given capacity $0\le C\le n$, let $S_C^*(z)$ denote that subset of the top at most C products according to $h_j(z)$ among the products with $h_j(z)>0$ i.e.,

$$S_C^*(z) = \arg\max_{S\subset N:\,|S|\le C} \sum_{j\in S} h_j(z).$$

Finally, let z_C^* denote the optimal revenue of the capacitated assortment optimization problem. Given this, it is a known fact that that optimal solution S_C^* to the capacitated assortment optimization problem (15) is $S_C^* = S_C^*(z_C^*)$ (Rusmevichientong et al. 2010).

Define the function $f\colon \mathbb{R}_+ \to \mathbb{R}$ as

$$f(z) = \begin{cases} \min_{j \in S_C^*(z)} h_j(z), & \text{if } S_C^*(z) \neq \emptyset \\ 0, & \text{otherwise.} \end{cases}$$

Since $h_j(\cdot)$ is linear for any $j \in N$, it is easy to see that $f(\cdot)$ as defined above is a piece-wise linear function in which every break-point corresponds to a point of intersection between two lines $h_j(\cdot)$ and $h_{j'}(\cdot)$. Thus, there are at most n^2 break-points. We say that a line h_j intersects f from top at at point z if and only if $h_j(z) = f(z)$, $h_j(z^-) \geq f(z^-)$, and $h_j(z^+) < f(z^+)$. Similarly, we say that h_j intersects f from the bottom if and only if $h_j(z) = f(z)$, $h_j(z^-) < f(z^-)$, and $h_j(z^+) \geq f(z^+)$. Finally, we say that h_j intersects the line $h_{j'}$ at z from the top if and only if $h_j(z) = h_{j'}(z)$ and $h_j(z^+) < h_{j'}(z^+)$, and h_j intersects $h_{j'}$ from the bottom if and only if $h_j(z) = h_{j'}(z)$ and $h_j(z^+) > h_{j'}(z^+)$.

In order to prove the result for the capacitated case, we need the following propositions.

Proposition 1. *Any line h_j intersects f from the top in at most $\min\{C, n - C + 1\}$ points.*

Proof. We count the number of intersection points (from top or bottom) by obtaining an injective map between the intersection points and the products. Specifically, let T and B denote the sets of points at which h_j intersects f from the top and bottom respectively. Then, we define injective mappings $t\colon T \to N$ and $b\colon B \to N$ such that h_j intersects $h_{t(z)}$ from the top for every $z \in T$ and h_j intersects $h_{b(z)}$ from the bottom for every $z \in B$.

In order to establish this mapping, consider a point $z \in T$. It follows by definition that $h_j(z) = f(z)$, $h_j(z^-) \geq f(z^-)$, and $h_j(z^+) < f(z^+)$. We let $t(z)$ be the product such that $f(z^+) = h_{t(z)}(z^+)$. Since both f and $h_{t(z)}$ are continuous, we now have that $h_j(z) = f(z) = h_{t(z)}(z)$ and $h_j(z^+) < h_{t(z)}(z^+)$. It thus follows that h_j intersects $h_{t(z)}$ from the top at z. Since two lines intersect at most at one point, it is easy to see that $t(z) \neq t(z')$ for any $z, z' \in T$ and $z \neq z'$. Thus, the mapping t is injective and h_j intersects the line $h_{t(z)}$ from the top at every $z \in T$.

In a similar fashion, let $b(z)$ be the product such that $f(z^-) = h_{b(z)}(z^-)$ for any $z \in B$. We then have by definition that $h_j(z) = f(z) = h_{b(z)}(z)$ and $h_j(z^-) < h_{b(z)}(z^-)$. This implies that h_j intersects the line $h_{b(z)}$ from the bottom at z. Furthermore, $b(z) \neq b(z')$ for any $z, z' \in B$ and $z \neq z'$ because two lines intersect at at most one point. We thus have an injective mapping b and h_j intersects the line $h_{b(z)}$ from the bottom at every $z \in B$.

It follows from the injective mappings t and b that both T and B are finite sets. Letting $|T| = L$ and $|B| = K$, let $t_1 < t_2 < \cdots < t_L$ be the elements of T and $b_1 < b_2 < \cdots < b_K$ be the elements of B. Furthermore, since h_j and f are continuous, between any two points where h_j intersects f from above, there exists a point at which h_j intersects f from below. It thus follows that we must have either $b_1 < t_1 < b_2 < t_2 < \cdots < b_K < t_L$ or $t_1 < b_1 < t_2 < b_2 < \cdots < t_{L-1} < b_K < t_L$ (note that since f is always non-negative and h_j is negative for large enough z, at the largest intersection point, h_j always intersects f from the top). It thus follows that $L = K$ or $L = K + 1$. In other words, $L \leq K + 1$.

We now obtain upper bounds on L and K. First, suppose $|S_C(z_1)| = C' < C$. Since $h_j(z_1) = f(z_1)$, all the lines below h_j at z_1 must also be below the y-axis. This implies that h_j cannot intersect any other lines from top above the y-axis. Therefore, h_j intersect f from the above at at most one point yielding $L \leq 1$. Since $C \geq 1$, $L \leq 1 \leq \min\{C, n - C + 1\}$.

Now suppose $|S_C(z_1)| = C$. Then, since $h_j(t_1) = f(t_1)$, it follows by definition that there are at most $n - C$ lines that lie below $h_j(t_1)$. This implies that h_j can intersect at most $n - C$ more lines from the top beyond t_1. Thus, including t_1, the total number of points in T is at most $n - C + 1$ or equivalently $L \leq n - C + 1$. Similarly, since $h_j(t_1) = f(t_1)$, there are at most $C - 1$ lines that are above h_j at t_1. This implies that h_j can intersect at most $C - 1$ lines from the bottom beyond t_1, which in turn implies that $K \leq C - 1$. Since $L \leq K + 1$, we have that $L \leq C$. Therefore, we have show that $L \leq \min\{C, n - C + 1\}$ for some $C' \leq C$.

The result of the proposition now follows. □

PROPOSITION 2. *For any two assortments S_1 and S_2, the following must be true*

$$\sum_{j \in S_1} h_j(z) > \sum_{j \in S_2} h_j(z) \iff R(S_1) > R(S_2) \text{ for } z = R(S_1), R(S_2).$$

Proof. The result follows from the definitions. We prove the result for $z = R(S_2)$. The proof is symmetrical for $z = R(S_1)$. Let z_2 denote $R(S_2)$. Then, we have

$$z_2 = R(S_2) = \frac{\sum_{j \in S_2} p_j v_j}{v_0 + \sum_{j \in S_2} v_j} \implies \left(v_0 + \sum_{j \in S_2} v_j\right) z_2 = \sum_{j \in S_2} p_j v_j \implies v_0 z_2 = \sum_{j \in S_2} (p_j - z_2) v_j = \sum_{j \in S_2} h_j(z_2). \quad (16)$$

We must then have

$$\begin{aligned}
\sum_{j \in S_1} h_j(z_2) > \sum_{j \in S_2} h_j(z_2) &\iff \sum_{j \in S_1} h_j(z_2) > v_0 z_2 \\
&\iff \sum_{j \in S_1} v_j (p_j - z_2) > v_0 z_2 \\
&\iff \sum_{j \in S_1} p_j v_j > z_2 \left(v_0 + \sum_{j \in S_2} v_j\right) \\
&\iff \frac{\sum_{j \in S_1} p_j v_j}{v_0 + \sum_{j \in S_2} v_j} > z_2 \\
&\iff R(S_1) > R(S_2),
\end{aligned}$$

where the first equivalence follows from (16). The result of the proposition now follows. □

PROPOSITION 3. *Suppose product j is removed (as part of either the 'remove' or the 'exchange' operation) in iteration $t - 1$ of* ADXOPT *algorithm. Then, it must be that $h_j(z_t) < f(z_t)$, where $z_t = R(S^t)$ and S^t is the estimate of the optimal assortment at the end of iteration t.*

Proof. We prove this result by contradiction. In particular, suppose product j is such that $h_j(z_t) \geq f(z_t)$. Then if follows by the definition of f that $j \in S^*_C(z_t)$. We consider two cases: (i) $|S^t| < C$, and (ii) $|S^t| = C$.
Case (i). $|S^t| < C$. In this case, let $S = S^t \cup \{j\}$. Since $h_j(z_t) \geq f(z_t) \geq 0$, it is easy to see that $\sum_{\ell \in S} h_\ell(z_t) - \sum_{\ell \in S^t} h_\ell(z_t) = h_j(z_t) \geq 0$, which implies according to Proposition 2 that $R(S) > R(S^t)$ Since S can be

obtained from S^{t-1} through a feasible operation (*not* remove j), this contradicts the fact that S^t is obtained from S^{t-1} in a greedy manner. This finishes the first case.

Case (ii). $|S^t| = C$. Since $|S^*_C(z_t)| \leq C$ and $S^*_C(z_t) \setminus S^t \neq \emptyset$ (because it contains product j), it must be that $S^t \setminus S^*_C(z_t) \neq \emptyset$. Let $j' \in S^t \setminus S^*_C(z_T)$. We must then have that $h_{j'}(z_t) < h_j(z_t)$ because product j belongs to $S^*_C(z_t)$ whereas j' does not. This implies that for $S = S^t \setminus \{j'\} \cup \{j\}$, $\sum_{\ell \in S} h_\ell(z_t) - \sum_{\ell \in S^t} h_\ell(z_t) = h_j(z_t) - h_{j'}(z_t) > 0$. Thus, it follows from Proposition 2 that $R(S) > R(S^t)$. Since S can be obtained from S^{t-1} through a feasible operation (removing j' instead of j), we have arrived at a contradiction to the fact that S^t is obtained from S^{t-1} in a greedy manner. This finishes the second case.

The result of the proposition now follows. □

Proof of Theorem 7. We are now ready to prove the result of the theorem for the capacitated case. Let T be the iteration in which the ADXOpt algorithm terminates. Note that it is sufficient to prove that $S^T = S^*_C(z_T)$, where $z_T = R(S^T)$; then, it follows from the result in Rusmevichientong et al. (2010) that S^T is optimal. For that, we first claim that the number of removals at the beginning of iteration T of every product $j \in S^*_C(z_T)$ is strictly less than b. We defer the proof of the claim to the end. Assuming the claim is true, we prove that $S^T = S^*_C(z_T)$ by showing that $S^T \subseteq S^*_C(z_T)$ and $S^*_C(z_T) \subseteq S^T$.

We show that $S^*_C(z_T) \subseteq S^T$ by contradiction. Suppose there is a product $j \in S^*_C(z_T)$, but $j \notin S^T$. If $|S^T| < C$, then let $S = S^T \cup \{j\}$. Since the number of removals of j is strictly less than b, S can be obtained from S^T through a feasible operation (addition). Furthermore, it follows from the fact that $j \in S^*_C(z_T)$ that $h_j(z_T) > 0$. Therefore, $\sum_{\ell \in S} h_\ell(z_T) - \sum_{\ell \in S^T} h_\ell(z_T) = h_j(z_T) > 0$, which according to Proposition 2 implies that $R(S) > R(S^T)$ contradicting the fact that the algorithm terminated in assortment S^T.

In a similar fashion, suppose $|S^T| = C$. Then, we must have a product $j' \in S^T$ such that $j' \notin S^*_C(z_T)$; otherwise, $S^T \subseteq S^*_C(z_T)$, which in turn implies that $S^T = S^*_C(z_T)$ because $|S^*_C(z_T)| \leq C$. Since $j' \notin S^*_C(z_T)$ and $j \in S^*_C(z_T)$, we must have that $h_j(z_T) > h_{j'}(z_T)$. Letting $S = S^T \setminus \{j'\} \cup \{j\}$, it is easy to see that $\sum_{\ell \in S} h_\ell(z_T) - \sum_{\ell \in S^T} h_\ell(z_T) = h_j(z_T) - h_{j'}(z_T) > 0$, which according to Proposition 2 implies that $R(S) > R(S^T)$. Since the number of removals of j is strictly less than b, S can be obtained from S^T through a feasible operation (exchange). Thus, we have arrived at a contradiction.

We now show that $S^T \subseteq S^*_C(z_T)$. First suppose that $|S^*_C(z_T)| = C$. In this case, since $S^*_C(z_T) \subseteq S^T$ (from the step above) and $|S^T| \leq C$, we must have that $S^*_C(z_T) = S^T$. Now suppose $|S^*_C(z_T)| < C$. Then, to arrive at a contradiction, suppose there is a product $j \in S^T$, but $j \notin S^*_C(z_T)$. Since $|S^*_C(z_T)| < C$ and $j \notin S^*_C(z_T)$, it must be that $h_j(z_T) < 0$. Therefore, letting $S = S^T \setminus \{j\}$, we have $\sum_{\ell \in S} h_\ell(z_T) - \sum_{\ell \in S^T} h_\ell(z_T) = -h_j(z_T) > 0$, which according to Proposition 2 implies that $R(S) > R(S^T)$. Since S can be obtained from S^T through a feasible operation (removal), this is a contradiction.

We have thus shown that $S^T = S^*_C(z_T)$. In order to prove that S^T is optimal, it is sufficient to show that $R(S^T) \geq R(S)$ for any $S \subset N$ such that $|S| \leq C$. For that, consider any assortment $S \subset N$ such that $|S| \leq C$. Since $S^T = S^*_C(z_T)$, it follows by definition that $\sum_{\ell \in S^T} h_\ell(z_T) \geq \sum_{\ell \in S} h_\ell(z_T)$. Further, note that

$$z_T = R(S^T) = \frac{\sum_{\ell \in S^T} p_\ell v_\ell}{v_0 + \sum_{\ell \in S^T} v_\ell} \implies \sum_{\ell \in S^T} v_\ell(p_\ell - z_T) = v_0 z_T \implies \sum_{\ell \in S^T} h_\ell(z_T) = v_0 z_T.$$

It thus follows that

$$\sum_{\ell \in S^T} h_\ell(z_T) \geq \sum_{\ell \in S} h_\ell(z_T) \implies v_0 z_T \geq \sum_{\ell \in S} h_\ell(z_T) \implies z_T \geq \frac{\sum_{\ell \in S} p_\ell v_\ell}{v_0 + \sum_{\ell \in S} v_\ell} = R(S).$$

We have thus shown that $R(S^T) \geq R(S)$, establishing that S^T is the optimal solution.

We are now only left with proving the claim that the number of removals at the beginning of iteration T of every product $j \in S_C^*(z_T)$ is strictly less than b. We establish this result by actually proving a stronger result: in any iteration $t \leq T$, the number of removals of any product $j \in S_C^*(z_t)$ is strictly less than b.

We prove this result by induction. The result is trivially true for the first iteration because none of the products is yet removed. Suppose the result is true until iteration t. Now consider iteration $t+1$. We claim that at the end of iteration $t+1$, the number of removals of all products in $S_C^*(z_{t+1})$ is strictly less than b. Note that since $h_\ell(z_{t+1}) \geq f(z_{t+1})$ for all products $\ell \in S_C^*(z_{t+1})$, it follows from Proposition 3 that none of the products in $S_C^*(z_{t+1})$ will be removed in iteration $t+1$. Given this, we count the number of removals of a specific product $j \in S_C^*(z_{t+1})$ up to iteration t. We do this by obtaining a correspondence between the removals and the intersection points between h_j and f.

In order to obtain the correspondence, we first prove that in any iteration $t' \leq t$ in which j was added (as part of either the 'add' or the 'exchange' operation), we must have $h_j(z_{t'}) \geq f(z_{t'})$; in other words, h_j must be above f at $z_{t'}$. Suppose that is not the case and $h_j(z_{t'}) < f(z_{t'})$. Since $z_{t'} < z_{t+1}$ and $j \in S_C^*(z_{t+1})$, we must have that $h_j(z_{t'}) > h_j(z_{t+1}) \geq 0$. Since $h_j(z_{t'}) < f(z_{t'})$, it follows by definition that $j \notin S_C^*(z_{t'})$. Furthermore, since $h_j(z_{t'}) > 0$, it must be that $|S_C^*(z_{t'})| = C$; otherwise, j would be included in $S_C^*(z_{t'})$. Now because $S^{t'} \setminus S_C^*(z_{t'}) \neq \emptyset$ (since it contains product j) and $|S^{t'}| \leq C$, it follows that $S_C^*(z_{t'}) \setminus S^{t'} \neq \emptyset$. Let $j' \in S_C^*(z_{t'}) \setminus S^{t'}$. Then, we must have that $h_{j'}(z_{t'}) > h_j(z_{t'})$. Letting $S = S^{t'} \setminus \{j\} \cup \{j'\}$, it is easy to see that $\sum_{\ell \in S} h_\ell(z_{t'}) - \sum_{\ell \in S^{t'}} h_\ell(z_{t'}) = h_{j'}(z_{t'}) - h_j(z_{t'}) > 0$. Thus, it follows from Proposition 2 that $R(S) > R(S^{t'})$. Note that since $t' \leq t$, it follows from the induction hypothesis that the number of removals of all products in $S_C^*(z_{t'})$ is strictly less than b. Thus, S can be obtained from $S^{t'-1}$ through a feasible operation (adding j' instead of j). This contradicts the fact that $S^{t'}$ was obtained from $S^{t'-1}$ in a greedy manner.

Now, since a product cannot be removed before being added, there must be an addition between successive removals of product j. Further, since every removal of product j corresponds to a distinct point where h_j is be strictly below f (according to Proposition 3) and every addition of j corresponds to a point where h_j is above f, with each removal of j we can associate a distinct point at which h_j intersects f from the top. Thus, the number of removals is bounded above by the number of distinct points less than z_{t+1} at which h_j intersects f from the top. We know that the largest point of intersection between h_j and f is from the top (because $h_j(z) < 0$ for z large enough and $f(z) \geq 0$ for all $z \in \mathbb{R}_+$). However, at z_{t+1}, h_j is above f i.e., $h_j(z_{t+1}) \geq f(z_{t+1})$. Thus, letting T_j denote the total number of points where h_j intersects f from the top, the number of intersection points that are less than z_{t+1} is bounded above by $T_j - 1$ (where we subtract 1 to remove the largest point of intersection). We know from Proposition 1 that T_j is bounded above by $b = \min\{C, n - C + 1\}$. Thus, the number of removals of product j by iteration t is bounded above by $b - 1$. This establishes the induction step.

The result of the theorem now follows. $\square$