

# On Upper Bounds for Assortment Optimization under the Mixture of Multinomial Logit Models

Sumit Kunnumkal\*

September 30, 2014

## Abstract

The assortment optimization problem under the mixture of multinomial logit models is NP-complete and there are different approximation methods to obtain upper bounds on the optimal expected revenue. In this paper, we analytically compare the upper bounds obtained by the different approximation methods. We propose a new, tractable approach to construct an upper bound on the optimal expected revenue and show that it obtains the tightest bound among the existing tractable approaches in the literature to obtain upper bounds.

Assortment optimization has important applications in retailing and revenue management and has received much attention lately. In the assortment problem, we have a firm that is interested in maximizing revenues by selling products to customers, where each product has a revenue associated with it and customers choose among the offered products according to a given discrete choice model. The goal therefore is to figure out the set of products, or the assortment, that maximizes the expected revenue obtained from a customer. While there are a large number of discrete choice models that can be used to describe customer choice behavior, the multinomial logit model and its variants have been a popular choice in the assortment optimization literature.

In this paper, we consider the assortment problem under a mixture of multinomial logit models. In this model, we have multiple customer segments and an arriving customer belongs to a particular segment with a given probability and chooses among the offered products according to the multinomial logit model. The parameters of the multinomial logit model are allowed to depend on the segment to which the customer belongs.

The assortment optimization problem under the mixture of multinomial logit models is NP-complete; see Bront, Mendez-Diaz and Vulcano (2009) and Rusmevichientong, Shmoys, Tong and Topaloglu (2013). On the other hand, McFadden and Train (2000) show that the mixture of multinomial logits is a rich choice model that can approximate any random utility choice model arbitrarily closely. So there has been considerable interest in the assortment problem under the mixture of multinomial logit models and there is a growing literature that focuses on developing approximation methods that generate assortments with provable performance guarantees; see for example Rusmevichientong et al. (2013) and Mittal and Schulz (2013).

---

\*Indian School of Business, Hyderabad, 500032, India, email: sumit.kunnumkal@isb.edu

Another interesting research direction is to obtain upper bounds on the optimal assortment revenue, since upper bounds are useful in getting a better handle on the revenue performance gaps of the candidate assortments obtained by the different approximation methods. Bront et al. (2009) formulate the assortment problem under the mixture of multinomial logit models as a linear mixed integer program, the linear programming relaxation of which gives an upper bound on the optimal expected revenue. Feldman and Topaloglu (2013) propose a Lagrangian relaxation approach where they relax the constraints that the same assortment be offered to the different customer segments by associating Lagrange multipliers with them. While the Lagrangian relaxation approach obtains an upper bound, solving it turns out to be intractable. Therefore the authors propose a further approximation that is based on solving a continuous knapsack problem over a discrete set of grid points. They show that this approximation method is tractable and continues to provide an upper bound on the optimal expected revenue. One difficulty with the approximation method proposed by Feldman and Topaloglu (2013) is that the quality of the upper bound depends on the density of the grid, which makes it challenging to establish analytical results.

In this paper, we focus on solution methods to obtain upper bounds on the optimal expected revenue for the assortment problem under the mixture of multinomial logit models. We first provide a partial characterization of the optimal assortment. In particular, we show that an optimal assortment includes a certain revenue ordered subset of the products, and this subset can be obtained efficiently. This structure can be potentially exploited to speed up the solution methods by reducing the size of the search space. We then analytically compare the upper bounds obtained by the different solution methods in the literature. We show that the Lagrangian relaxation approach obtains the tightest bound among the available approaches to obtain upper bounds. However, since the Lagrangian relaxation is NP-complete, we propose a new, alternative approach to obtain an upper bound on the optimal expected revenue. Our approach builds on the ideas developed in Feldman and Topaloglu (2013) and Martinez-de-Albeniz and Kunnumkal (2014). The Lagrangian relaxation approach of Feldman and Topaloglu (2013) decomposes the assortment problem involving multiple customer segments into a number of single segment problems where there is a fixed cost of introducing a product into the assortment. Martinez-de-Albeniz and Kunnumkal (2014) study the assortment problem under the multinomial logit model when there is a fixed cost of introducing a product into the assortment and propose tractable solution methods. While our approach is a simple adaptation of the ideas developed in the above mentioned papers, it yields a solution method that is tractable and provably tighter than the other tractable approaches in the literature to obtain upper bounds on the optimal expected revenue.

In summary, we make the following contributions in this paper. 1) We provide a partial characterization of the solution to the assortment problem under the mixture of multinomial logit models. 2) We analytically compare the different methods proposed in the literature to obtain upper bounds on the optimal expected revenue. We show that the Lagrangian relaxation approach proposed by Feldman and Topaloglu (2013) obtains the tightest upper bound. However, the Lagrangian relaxation cannot be solved efficiently. 3) We propose a new approach to obtain an upper bound that remains tractable. The bound obtained by our approach is provably tighter than the other tractable solution methods in the literature to obtain upper bounds. Moreover, in contrast to Feldman and Topaloglu (2013), our solution method does not require a discrete set of grid points and thus reduces some of the subjectivity involved in choosing a discretization scheme. 4) Computational experiments indicate that our approach can be beneficial both in terms of tighter bounds and faster run times.

The rest of the paper is organized as follows. In Section 1, we describe the assortment problem under the mixture of multinomial logit models. In Section 2, we describe the solution methods proposed in the literature to obtain upper bounds on the optimal expected revenue. We analytically compare

the upper bounds obtained by the proposed methods in Section 3. In Section 4, we describe our solution approach and show that it obtains the tightest bound among the tractable solution methods in the literature to obtain upper bounds on the optimal expected revenue. Section 5 presents our computational experiments.

## 1 PROBLEM FORMULATION

We consider the assortment optimization problem under the mixture of multinomial logit models. We have  $N$  products and the revenue associated with product  $j \in \{1, \dots, N\}$  is  $p_j$ . We let  $x_j \in \{0, 1\}$  denote whether we offer product  $j$ . We have  $L$  customer segments interested in purchasing a product from the offered assortment. The preference weight for customer segment  $l \in \{1, \dots, L\}$  for product  $j$  is  $v_j^l$ , while the preference weight associated with a segment  $l$  customer not purchasing anything is normalized to be 1. Within each segment, choice is governed by the multinomial logit model and so the probability that a segment  $l$  customer purchases product  $j$  is  $v_j^l x_j / (1 + \sum_k v_k^l x_k)$ . We let  $\alpha^l$  denote the arrival rate of customer segment  $l$ . The assortment problem is to decide which products to make available to an arriving customer in order to maximize the expected total revenue. The optimal assortment can be obtained by solving the problem

$$Z^{OPT} = \max_{x|x_j \in \{0,1\}} \sum_l \alpha^l \frac{\sum_j p_j v_j^l x_j}{\sum_j v_j^l x_j + 1}. \quad (1)$$

Solving assortment problem (1) and obtaining the optimal expected revenue is intractable; see Bront et al. (2009) and Rusmevichientong et al. (2013). Therefore, it is unlikely that we can provide a complete characterization of the optimal solution to problem (1). However, Lemma 1 below provides a partial characterization of the structure of an optimal solution.

We begin with some preliminaries. Assume without loss of generality that the products are indexed in order of decreasing revenues so that  $p_1 \geq p_2 \geq \dots \geq p_N$ , and consider the assortment problem involving customer segment  $l$  alone

$$\max_{x^l|x_j^l \in \{0,1\}} \frac{\sum_j p_j v_j^l x_j^l}{\sum_j v_j^l x_j^l + 1}. \quad (2)$$

It is known that revenue ordered assortments are optimal for problem (2); see for example Liu and van Ryzin (2008). That is, there exists an optimal solution  $\hat{x}^l = \{\hat{x}_j^l | \forall j\}$  to problem (2) with  $\hat{x}_j^l = 1$  for  $j \in \{1, \dots, j^l\}$  and  $\hat{x}_j^l = 0$  for  $j \in \{j^l + 1, \dots, N\}$ , where  $j^l \in \{1, \dots, N\}$ . We have the following lemma.

**Lemma 1** *Let  $\hat{j} = \min_l \{j^l\}$ . There exists an optimal solution  $\hat{x} = \{\hat{x}_j | \forall j\}$  to problem (1) with  $\hat{x}_j = 1$  for  $j \in \{1, \dots, \hat{j}\}$ .*

**Proof.** Suppose that the statement of the lemma is false. Let  $\hat{x}$  be an optimal solution to problem (1) and let  $\kappa \in \{1, \dots, \hat{j}\}$  be the smallest index such that  $\hat{x}_\kappa = 0$ . We let  $\tilde{x} = \{\tilde{x}_j | \forall j\}$  be the same as  $\hat{x}$  except that  $\tilde{x}_\kappa = 1$ . We show below that the solution  $\tilde{x}$  generates as much revenue as  $\hat{x}$ , and is therefore also optimal.

Let  $S = \{1, \dots, N\} \setminus \{\kappa\}$  and  $T = \{1, \dots, \kappa - 1\}$ . Fix a segment  $l$  and note that

$$\frac{\sum_{j \in S} p_j v_j^l \hat{x}_j}{\sum_{j \in S} v_j^l \hat{x}_j + 1} = \left( \frac{\sum_{j \in T} v_j^l \hat{x}_j + 1}{\sum_{j \in S} v_j^l \hat{x}_j + 1} \right) \frac{\sum_{j \in T} p_j v_j^l \hat{x}_j}{\sum_{j \in T} v_j^l \hat{x}_j + 1} + \left( \frac{\sum_{j \in S \setminus T} v_j^l \hat{x}_j}{\sum_{j \in S} v_j^l \hat{x}_j + 1} \right) \frac{\sum_{j \in S \setminus T} p_j v_j^l \hat{x}_j}{\sum_{j \in S \setminus T} v_j^l \hat{x}_j}. \quad (3)$$

Since  $\kappa \in \{1, \dots, j\}$  and  $j \leq j^l$ , it follows that  $T \subsetneq \{1, \dots, j^l\}$ , the optimal revenue ordered assortment for segment  $l$ . Lemma 2 in Liu and van Ryzin (2008) then implies that  $\frac{\sum_{j \in T} p_j v_j^l \hat{x}_j}{\sum_{j \in T} v_j^l \hat{x}_j + 1} \leq \frac{\sum_{j \in T} p_j v_j^l \hat{x}_j + p_\kappa v_\kappa^l}{\sum_{j \in T} v_j^l \hat{x}_j + v_\kappa^l + 1}$ . This together with Lemma 3.1 in Rusmevichientong and Topaloglu (2012) implies that  $p_\kappa \geq \frac{\sum_{j \in T} p_j v_j^l \hat{x}_j}{\sum_{j \in T} v_j^l \hat{x}_j + 1}$ . On the other hand, since the products are revenue ordered,  $p_\kappa \geq p_j$  for all  $j \in S \setminus T$  and we have  $p_\kappa \geq \frac{\sum_{j \in S \setminus T} p_j v_j^l \hat{x}_j}{\sum_{j \in S \setminus T} v_j^l \hat{x}_j}$ . The above statements together with (3) imply that

$$\frac{\sum_{j \in S} p_j v_j^l \hat{x}_j}{\sum_{j \in S} v_j^l \hat{x}_j + 1} \leq p_\kappa. \quad (4)$$

Now consider the difference in the expected revenues from segment  $l$  under  $\tilde{x}$  and  $\hat{x}$ . We have

$$\begin{aligned} \frac{\sum_j p_j v_j^l \tilde{x}_j}{\sum_j v_j^l \tilde{x}_j + 1} - \frac{\sum_j p_j v_j^l \hat{x}_j}{\sum_j v_j^l \hat{x}_j + 1} &= \sum_{j \in S} p_j v_j^l \hat{x}_j \left[ \frac{1}{\sum_{j \in S} v_j^l \hat{x}_j + v_\kappa^l + 1} - \frac{1}{\sum_{j \in S} v_j^l \hat{x}_j + 1} \right] + \frac{p_\kappa v_\kappa^l}{\sum_{j \in S} v_j^l \hat{x}_j + v_\kappa^l + 1} \\ &= \frac{v_\kappa^l}{\sum_{j \in S} v_j^l \hat{x}_j + v_\kappa^l + 1} \left[ p_\kappa - \frac{\sum_{j \in S} p_j v_j^l \hat{x}_j}{\sum_{j \in S} v_j^l \hat{x}_j + 1} \right] \geq 0 \end{aligned}$$

where the first equality uses the fact that  $\tilde{x}$  is the same as  $\hat{x}$  except that  $\tilde{x}_\kappa = 1$ , while  $\hat{x}_\kappa = 0$ , and the inequality follows from (4). It follows that  $\sum_l \alpha^l \frac{\sum_j p_j v_j^l \tilde{x}_j}{\sum_j v_j^l \tilde{x}_j + 1} \geq \sum_l \alpha^l \frac{\sum_j p_j v_j^l \hat{x}_j}{\sum_j v_j^l \hat{x}_j + 1}$ .  $\square$

Lemma 1 implies that an optimal solution to problem (1) includes the revenue ordered subset  $\{1, \dots, j\}$ , and this subset can be determined efficiently. This may be useful in speeding up approximation methods by reducing the size of the solution space. We emphasize that revenue ordered assortments are not optimal for problem (1) in general; see Rusmevichientong et al. (2013).

## 2 UPPER BOUNDS

In this section, we describe approximation methods that obtain upper bounds on the optimal expected revenue. Upper bounds are useful in assessing the revenue performance gaps of the assortments obtained by the different heuristic methods.

### 2.1 LINEAR PROGRAMMING RELAXATION

Bront et al. (2009) formulate problem (1) as the mixed integer program

$$Z^{OPT} = \max \quad \sum_l \sum_j \alpha^l p_j v_j^l y_j^l \quad (5)$$

$$\text{s.t.} \quad \sum_j v_j^l y_j^l + w^l = 1 \quad \forall l \quad (6)$$

$$y_j^l \leq w^l \quad \forall l, j \quad (7)$$

$$y_j^l \leq x_j \quad \forall l, j \quad (8)$$

$$w^l - y_j^l \leq 1 - x_j \quad \forall l, j \quad (9)$$

$$y_j^l, w^l \geq 0, x_j \in \{0, 1\}. \quad (10)$$

In the above optimization problem, the decision variable  $x_j$  indicates if product  $j$  is offered, while the term  $v_j^l y_j^l$  can be interpreted as the probability that segment  $l$  purchases product  $j$  and  $w^l$  can be

interpreted as the probability that customer segment  $l$  does not purchase anything. It is intractable to solve the above mixed integer program, but its linear programming relaxation yields an upper bound on the optimal expected revenue. Letting  $Z^{LP}$  denote the optimal objective function value of the LP relaxation of problem (5)-(10), we have  $Z^{OPT} \leq Z^{LP}$ .

## 2.2 LAGRANGIAN RELAXATION

A second upper bound is due to Feldman and Topaloglu (2013) who consider relaxing the constraints that the same assortment be offered to the different customer segments by associating Lagrange multipliers with them. Letting  $x_j^l \in \{0, 1\}$  denote the decision to offer product  $j$  to segment  $l$ , we can view problem (1) as requiring  $x_j^l = x_j^\phi$  for all  $l$  and  $j$ , where  $x_j^\phi \in \{0, 1\}$ . These constraints ensure that the decision to offer a product is uniform across the different customer segments. Feldman and Topaloglu (2013) relax the constraints  $x_j^l = x_j^\phi$  by associating Lagrange multipliers  $\lambda_j^l$  with them. For a given set of Lagrange multipliers  $\lambda = \{\lambda_j^l | \forall l, j\}$ , they consider solving problem

$$Z^{LR}(\lambda) = \max_{x | x_j^\phi, x_j^l \in \{0, 1\}} \sum_l \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{\sum_j v_j^l x_j^l + 1} - \sum_l \sum_j \lambda_j^l x_j^l + \sum_j x_j^\phi [\sum_l \lambda_j^l]. \quad (11)$$

The authors show that for any set of Lagrange multipliers  $\lambda$ , we have  $Z^{OPT} \leq Z^{LR}(\lambda)$  and they find the set of Lagrange multipliers that yield the tightest upper bound on the optimal expected revenue by solving the problem

$$Z^{LR} = \min_{\lambda} Z^{LR}(\lambda).$$

However, the optimization problem for segment  $l$  in (11) involves finding the optimal assortment when there are fixed costs associated with introducing the products, and this is known to be NP-complete in general; see Kunnumkal, Rusmevichientong and Topaloglu (2009). An exception is when all the Lagrange multipliers are set to zero. If  $\lambda_j^l = 0$  for all  $j$  and  $l$ , then there are no costs associated with offering any of the products and the problem  $Z^{LR}(\bar{0})$  can be solved efficiently; see Talluri (2014). On the other hand, if the Lagrange multipliers are nonzero, then  $Z^{LR}(\lambda)$  cannot be computed in a tractable manner.

## 2.3 LAGRANGIAN RELAXATION OVER A DISCRETE GRID

Feldman and Topaloglu (2013) propose a tractable approximation to  $Z^{LR}(\lambda)$ , which is based on writing the optimization problem for segment  $l$  in (11) as

$$\max_{\theta^l} \max_{x | x_j^l \in [0, 1], \sum_j v_j^l x_j^l \leq \theta^l} \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{1 + \theta^l} - \sum_j \lambda_j^l x_j^l.$$

In the above representation, the inner maximization problem is a knapsack problem that depends on the parameter  $\theta^l$ , where  $\theta^l$  can be interpreted as the capacity of the knapsack and can take values in the interval  $[0, \sum_j v_j^l]$ . Instead of solving the optimization problem for all values of  $\theta^l$  in its range, the authors focus on a set of discrete grid points  $\Theta = \{\theta^{l(g)} | \forall l, g\}$ , where  $0 = \theta^{l(1)} < \theta^{l(2)} < \dots < \theta^{l(G)} = \sum_j v_j^l$  for all  $l$ , and  $G$  is the number of grid points. For an interval  $[\theta^{l(g)}, \theta^{l(g+1)}]$ , they solve the problem

$$Z^{LR-D, l, g}(\lambda, \Theta) = \max_{x | 0 \leq x_j^l \leq 1, \sum_j v_j^l x_j^l \leq \theta^{l(g+1)}} \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{1 + \theta^{l(g)}} - \sum_j \lambda_j^l x_j^l. \quad (12)$$

Feldman and Topaloglu (2013) show that for any set of discrete grid points

$$Z^{LR-D}(\lambda, \Theta) = \sum_l \left[ \max_{g \in \{1, \dots, G-1\}} Z^{LR-D, l, g}(\lambda, \Theta) \right] + \sum_j \max_{x_j^\phi \in \{0, 1\}} x_j^\phi \left[ \sum_l \lambda_j^l \right] \geq Z^{LR}(\lambda),$$

where the argument  $\Theta$  emphasizes the dependence of the choice of grid points on the upper bound obtained. Note that problem (12) is a continuous knapsack problem and can be solved efficiently. Therefore, the upper bound  $Z^{LR-D}(\lambda, \Theta)$  can be obtained in a tractable manner. Furthermore, the authors show that the problem

$$Z^{LR-D}(\Theta) = \min_{\lambda} Z^{LR-D}(\lambda, \Theta) \quad (13)$$

can be solved efficiently as a convex minimization problem. However, there is a trade-off between the quality of the upper bound and the computational efficiency, since a more refined set of grid points gives a tighter upper bound, but at the same time imposes a greater computational burden. Therefore, the choice of the grid points becomes crucial to the performance of the method. Feldman and Topaloglu (2013) propose a set of exponential grid points and bound the performance gap between the set of exponential grid points and an arbitrary set of grid points.

### 3 COMPARISON OF UPPER BOUNDS

In this section, we compare the upper bounds obtained by the different approximation methods. We first compare the upper bound  $Z^{LR}$  with  $Z^{LP}$ . We begin by making the change of variables  $z_j^l = v_j^l y_j^l$  so that problem (5)-(10) can be written as

$$\begin{aligned} Z^{OPT} = \max \quad & \sum_l \sum_j \alpha^l p_j z_j^l \\ \text{s.t} \quad & z_j^l / v_j^l + \sum_k z_k^l \leq 1 \quad \forall l, j \\ & z_j^l / v_j^l \leq x_j \quad \forall l, j \\ & z_j^l / v_j^l + \sum_k z_k^l \geq x_j \quad \forall l, j \\ & z_j^l \geq 0, x_j \in \{0, 1\}, \end{aligned}$$

where the first and third constraints use the fact that  $w^l = 1 - \sum z_j^l$ . Introducing variables  $x_j^l$  and  $x_j^\phi$  and constraining them to be equal, the above optimization problem can be written as

$$Z^{OPT} = \max \quad \sum_l \sum_j \alpha^l p_j z_j^l \quad (14)$$

$$\text{s.t} \quad z_j^l / v_j^l + \sum_k z_k^l \leq 1 \quad \forall l, j \quad (15)$$

$$z_j^l / v_j^l \leq x_j^l \quad \forall l, j \quad (16)$$

$$z_j^l / v_j^l + \sum_k z_k^l \geq x_j^l \quad \forall l, j \quad (17)$$

$$x_j^l = x_j^\phi \quad \forall l, j \quad (18)$$

$$z_j^l \geq 0, x_j^l, x_j^\phi \in \{0, 1\}. \quad (19)$$

Relaxing constraints (18) by associating Lagrange multipliers  $\lambda = \{\lambda_j^l | \forall l, j\}$  with them, we get a relaxed problem

$$\bar{Z}(\lambda) = \max \quad \sum_l \sum_j \alpha^l p_j z_j^l + \sum_l \sum_j \lambda_j^l [x_j^\phi - x_j^l] \quad (20)$$

$$\text{s.t} \quad z_j^l / v_j^l + \sum_k z_k^l \leq 1 \quad \forall l, j \quad (21)$$

$$z_j^l / v_j^l \leq x_j^l \quad \forall l, j \quad (22)$$

$$z_j^l / v_j^l + \sum_k z_k^l \geq x_j^l \quad \forall l, j \quad (23)$$

$$z_j^l \geq 0, x_j^l, x_j^\phi \in \{0, 1\}. \quad (24)$$

We have the following lemma.

**Lemma 2** *If  $\{x_j^l, x_j^\phi, z_j^l | \forall l, j\}$  satisfies constraints (21)-(24), then we have  $z_j^l = v_j^l x_j^l / (1 + \sum_k v_k^l x_k^l)$  for all  $l$  and  $j$ .*

**Proof.** Fix a customer segment  $l$  and a product  $j$ . The claim clearly holds if  $x_j^l = 0$  since constraint (22) forces  $z_j^l$  to be zero. On the other hand, if  $x_j^l = 1$ , let  $J_+^l = \{k | x_k^l = 1\}$ . Note that  $j \in J_+^l$ . Constraints (21) and (23) imply that

$$z_k^l / v_k^l + \sum_{k' \in J_+^l} z_{k'}^l = 1 \quad (25)$$

for all  $k$  in  $J_+^l$ . Considering equality (25) for products  $j$  and  $k \in J_+^l$ , we have  $z_k^l / v_k^l = z_j^l / v_j^l$  for all  $k \in J_+^l$ . Plugging this back into the constraint  $z_j^l / v_j^l + \sum_{k' \in J_+^l} z_{k'}^l = 1$ , we get  $z_j^l = v_j^l / (1 + \sum_{k \in J_+^l} v_k^l) = v_j^l x_j^l / (1 + \sum_k v_k^l x_k^l)$ .  $\square$

Lemma 2 implies that problem (20)-(24) can be equivalently written as

$$\max_{x | x_j^\phi, x_j^l \in \{0,1\}} \sum_l \sum_j \alpha^l p_j \frac{v_j^l x_j^l}{1 + \sum_k v_k^l x_k^l} - \sum_l \sum_k \lambda_j^l x_j^l + \sum_l \sum_j \lambda_j^l x_j^\phi,$$

which is precisely the Lagrangian relaxation (11) proposed by Feldman and Topaloglu (2013). So we have the following lemma.

**Lemma 3**  $\bar{Z}(\lambda) = Z^{LR}(\lambda)$ .

Next, consider the LP relaxation of problem (14)-(19). Letting  $\hat{\lambda} = \{\hat{\lambda}_j^l | \forall l, j\}$  be an optimal set of dual variables associated with constraints (18), strong duality implies that

$$Z^{LP} = \max \sum_l \sum_j \alpha^l p_j z_j^l + \sum_l \sum_j \hat{\lambda}_j^l [x_j^\phi - x_j^l] \quad (26)$$

$$\text{s.t.} \quad z_j^l / v_j^l + \sum_k z_k^l \leq 1 \quad \forall l, j \quad (27)$$

$$z_j^l / v_j^l \leq x_j^l \quad \forall l, j \quad (28)$$

$$z_j^l / v_j^l + \sum_k z_k^l \geq x_j^l \quad \forall l, j \quad (29)$$

$$z_j^l \geq 0, 0 \leq x_j^l, x_j^\phi \leq 1. \quad (30)$$

**Proposition 4**  $Z^{LR} = \min_\lambda \bar{Z}(\lambda) \leq \bar{Z}(\hat{\lambda}) \leq Z^{LP}$ .

**Proof.** The equality follows from Lemma 3, while the first inequality holds since  $\hat{\lambda}$  is a feasible but not necessarily optimal solution to the minimization problem and the last inequality holds since  $x_j^l$  and  $x_j^\phi$  are restricted to be binary in problem (20)-(24) while they are allowed to take on fractional values in problem (26)-(30).  $\square$

Proposition 4 shows that the  $Z^{LR}$  bound is tighter than  $Z^{LP}$ . It trivially follows that  $Z^{LR} \leq Z^{LR}(\bar{0})$ , while Feldman and Topaloglu (2013) show that  $Z^{LR} \leq Z^{LR-D}(\Theta)$ . Therefore the  $Z^{LR}$  bound is the tightest upper bound among the available bounds. However, as mentioned,  $Z^{LR}$  cannot be computed

in a tractable manner, and Feldman and Topaloglu (2013) propose the approximation  $Z^{LR-D}(\Theta)$  that involves solving a continuous knapsack problem over a set of discrete grid points. While the numerical study in Feldman and Topaloglu (2013) indicates that their method performs quite well in practice and generates tight bounds, one difficulty with the discretization approach is that the quality of the solution depends on how refined the grid is and it is not immediately clear how to analytically compare the upper bound with that obtained by the other approaches, such as the LP relaxation.

#### 4 AN ALTERNATIVE UPPER BOUND

We propose an alternative approach to obtain an upper bound on the optimal revenues. It is a simple adaptation of the ideas developed in Feldman and Topaloglu (2013) and Martinez-de-Albeniz and Kunnumkal (2014), but our approach is tractable and is provably tighter than the existing tractable approaches to obtain upper bounds. Moreover, our approach avoids the subjectivity involved in determining the density of the grid points.

We begin by relaxing the constraints that the same assortment be offered to the different segments by associating Lagrange multipliers  $\lambda = \{\lambda_j^l | \forall l, j\}$  with them and consider solving the problem

$$Z^{LR-C}(\lambda) = \max_{x|0 \leq x_j^l \leq 1, x_j^\phi \in \{0,1\}} \sum_l \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{\sum_j v_j^l x_j^l + 1} - \sum_l \sum_j \lambda_j^l x_j^l + \sum_j x_j^\phi \left[ \sum_l \lambda_j^l \right]. \quad (31)$$

Note that the problem we solve is almost the same as  $Z^{LR}(\lambda)$  except that we *relax* the integrality constraints on the  $x_j^l$  variables. Again note that problem (31) separates by segment and the optimization problem for segment  $l$  has the form

$$\max_{x|0 \leq x_j^l \leq 1} \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{\sum_j v_j^l x_j^l + 1} - \sum_j \lambda_j^l x_j^l.$$

It can be shown that the above optimization problem can be solved efficiently, in fact, essentially in closed form; see Martinez-de-Albeniz and Kunnumkal (2014). Therefore, given a set of Lagrange multipliers, problem (31) can be solved efficiently. Since  $Z^{LR-C}(\lambda) \geq Z^{LR}(\lambda) \geq Z^{OPT}$ , we obtain an upper bound on the optimal assortment revenue. We propose solving

$$Z^{LR-C} = \min_{\lambda} Z^{LR-C}(\lambda) \quad (32)$$

to find the optimal set of Lagrange multipliers that yield the tightest upper bound. By following the same line of reasoning as in Feldman and Topaloglu (2013), it can be shown that the optimal Lagrange multipliers can be obtained efficiently.

We now show that  $Z^{LR-C}$  obtains an upper bound that is uniformly tighter than the existing tractable approaches in the literature. Recall, that if all the Lagrange multipliers are set to zero, then problem (11) can be solved in a tractable manner. Proposition 5 below shows that the resulting upper bound  $Z^{LR}(\bar{0})$  is weaker than  $Z^{LR-C}$ , while Propositions 6 and 7 below show that the  $Z^{LR-C}$  bound is tighter than  $Z^{LP}$  and  $Z^{LR-D}(\Theta)$ .

**Proposition 5**  $Z^{LR-C} \leq Z^{LR}(\bar{0})$ .

**Proof.** Setting all the Lagrange multipliers to zero, we have

$$Z^{LR-C}(\bar{0}) = \max_{x|0 \leq x_j^l \leq 1} \sum_l \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{\sum_j v_j^l x_j^l + 1} = \max_{x|x_j^l \in \{0,1\}} \sum_l \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{\sum_j v_j^l x_j^l + 1} = Z^{LR}(\bar{0})$$



where the second equality holds since the objective function is quasiconvex; see for example, Lemma 1 in Liu and van Ryzin (2008). We have  $Z^{LR-C} = \min_{\lambda} Z^{LR-C}(\lambda) \leq Z^{LR-C}(\bar{0}) = Z^{LR}(\bar{0})$ .  $\square$

**Proposition 6**  $Z^{LR-C} \leq Z^{LP}$ .

**Proof.** Let  $\hat{\lambda} = \{\hat{\lambda}_j^l | \forall l, j\}$  be the optimal set of dual variables corresponding to constraints (18). Let  $\hat{x} = \{\hat{x}_j^l, \hat{x}_j^\phi | \forall l, j\}$  be an optimal solution to problem (31), when the Lagrange multipliers are set as  $\hat{\lambda}$ . Letting  $\hat{z}_j^l = v_j^l \hat{x}_j^l / (1 + \sum_k v_k^l \hat{x}_k^l)$ , and  $\hat{z} = \{\hat{z}_j^l | \forall l, j\}$ , we show that  $(\hat{x}, \hat{z})$  is a feasible solution to problem (26)-(30). Note that  $\sum_k \hat{z}_k^l = \sum_k v_k^l \hat{x}_k^l / (1 + \sum_k v_k^l \hat{x}_k^l)$ . So we have

$$\hat{z}_j^l / v_j^l = \hat{x}_j^l / (1 + \sum_k v_k^l \hat{x}_k^l) \leq 1 / (1 + \sum_k v_k^l \hat{x}_k^l) = 1 - \sum_k \hat{z}_k^l$$

and constraint (27) holds. Next,  $\hat{z}_j^l / v_j^l = \hat{x}_j^l / (1 + \sum_k v_k^l \hat{x}_k^l) \leq \hat{x}_j^l$  and so constraint (28) holds. Finally note that

$$\sum_k \hat{z}_k^l = \sum_k v_k^l \hat{x}_k^l / (1 + \sum_k v_k^l \hat{x}_k^l) \geq \hat{x}_j^l (\sum_k v_k^l \hat{x}_k^l) / (1 + \sum_k v_k^l \hat{x}_k^l) = \hat{x}_j^l (1 - 1 / (1 + \sum_k v_k^l \hat{x}_k^l)) = \hat{x}_j^l - \hat{z}_j^l / v_j^l$$

and constraint (29) also holds. Therefore  $Z^{LR-C}(\hat{\lambda}) \leq Z^{LP}$  and the result follows.  $\square$

**Proposition 7**  $Z^{LR-C} \leq Z^{LR-D}(\Theta)$ .

**Proof.** We have

$$Z^{LR-C}(\lambda) = \sum_l \left\{ \max_{\theta^l} \max_{x | 0 \leq x_j^l \leq 1, \sum_j v_j^l x_j^l \leq \theta^l} \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{1 + \theta^l} - \sum_j \lambda_j^l x_j^l \right\} + \sum_j \max_{x_j^\phi \in \{0,1\}} x_j^\phi [\sum_l \lambda_j^l].$$

Now, for  $\theta^l \in [\theta^{l(g)}, \theta^{l(g+1)}]$ ,

$$\max_{x | 0 \leq x_j^l \leq 1, \sum_j v_j^l x_j^l \leq \theta^l} \alpha^l \frac{\sum_j p_j v_j^l x_j^l}{1 + \theta^l} - \sum_j \lambda_j^l x_j^l \leq Z^{LR-D,l,g}(\lambda, \Theta).$$

Therefore,  $Z^{LR-C}(\lambda) \leq Z^{LR-D}(\lambda, \Theta)$  and the result follows.  $\square$

## 5 COMPUTATIONAL EXPERIMENTS

In this section we numerically test the quality of the upper bound obtained by our proposed method. Feldman and Topaloglu (2013) demonstrate that the upper bound obtained by the Lagrangian relaxation over a discrete grid tends to be significantly tighter than the linear programming relaxation bound and the bound obtained by setting all the Lagrange multipliers to zero. So, in our numerical study, we focus on comparing the upper bound and the solution time of our proposed method with that of the Lagrangian relaxation over a discrete grid.

We use LR-C to refer to the solution method described in Section 4 and LR-D( $\Theta$ ) to the solution method described in Section 2.3. Following Feldman and Topaloglu (2013), we use a set of exponential grid points in LR-D( $\Theta$ ). That is, we cover the interval  $[0, \sum_j v_j^l]$  using a set of grid points  $\{\theta^{l(g)} | \forall g \in$

$\{1, \dots, G\}$ , where  $\theta^{l(g)} = (1 + \rho)^{g-1} - 1$  and  $\rho > 0$ . Note that the parameter  $\rho$  determines the density of the grid. As  $\rho$  decreases, the spacing between the adjacent grid points decreases and we require a larger number of grid points to cover the interval  $[0, \sum_j v_j^l]$ . We consider three different grid densities in our computational experiments by varying  $\rho$  over the set  $\{0.1, 0.01, 0.001\}$ . We note that Feldman and Topaloglu (2013) use a value of  $\rho = 0.001$  in their computational study. We solve problems (32) and (13) using subgradient search. In both cases, we initialize all the Lagrange multipliers to zero and run 200 iterations of the subgradient search algorithm with a step size of  $20/(40 + t)$  at iteration  $t$ . Our computational experiments are carried out in MATLAB on a Core i7 desktop with 3.4-GHz CPU and 16-GB RAM.

In all of our test problems, we have 50 products and 25 customer segments. We generate our test problems by following an approach similar to that described in Feldman and Topaloglu (2013). We sample the product revenues from the uniform distribution over  $[0, 2000]$ . We sample  $\beta^l$  from the uniform distribution over  $[0, 1]$  and set the arrival rate of segment  $l$ ,  $\alpha^l$ , as  $\beta^l / \sum_{l'} \beta^{l'}$ . We set the preference weight for segment  $l$  for product  $j$ ,  $v_j^l$ , as  $\frac{\nu_j}{\sum_{j'} \nu_{j'}} \frac{1 - \phi^l}{\phi^l}$ . We sample  $\nu_j$  from the uniform distribution over  $[0, V]$  and  $\phi^l$  from the uniform distribution over  $[0, \Phi]$ , where  $V$  and  $\Phi$  are parameters that we vary in our computational experiments. Note that  $\sum_j v_j^l / (1 + \sum_j v_j^l) = 1 - \phi^l$  and so the probability that a segment  $l$  customer leaves without making a purchase when all the products are offered is  $\phi^l$ .

In our computational experiments we vary  $V \in \{5, 10, 20\}$  and  $\Phi \in \{0.4, 0.6, 0.8\}$ , which gives us a total of nine parameter combinations. For each parameter combination, we generate 100 test problems by following the approach described above. For each test problem, we obtain the LR-C bound as well as the solution time. We obtain the corresponding numbers for LR-D( $\Theta$ ) at different grid densities by varying  $\rho \in \{0.1, 0.01, 0.001\}$ . We use LR-C as a benchmark and for each of the three grid densities, we compute the percentage gap in the upper bound with respect to the upper bound obtained by LR-C, as well as the solution time relative to the solution time of LR-C. We report the averages across the 100 test problems.

Table 1 summarizes the results of our computational study. The first column in this table gives the parameter combinations. The second column gives the percentage gap in the upper bounds obtained by LR-D( $\Theta$ ) with  $\rho = 0.1$  and LR-C, averaged over the 100 problem instances, where as the third column gives the average of the ratio of the corresponding solution times. Columns four and five have a similar interpretation except that they compare LR-D( $\Theta$ ) with  $\rho = 0.01$  with LR-C. The last two columns, respectively, compare the upper bounds and solution times of LR-D( $\Theta$ ) with  $\rho = 0.001$  with LR-C. We observe that the upper bound obtained by LR-C is on average 9.553% tighter than that obtained by LR-D( $\Theta$ ) with  $\rho = 0.1$ , 0.986% tighter than that obtained by LR-D( $\Theta$ ) with  $\rho = 0.01$  and 0.099% tighter than that obtained by LR-D( $\Theta$ ) with  $\rho = 0.001$ . The quality of the upper bound obtained by LR-D( $\Theta$ ) improves with the density of the grid, but this comes at the expense of an increase in the solution time. LR-C is able to obtain bounds that are comparable to LR-D( $\Theta$ ) with  $\rho = 0.001$  but at a fraction of the computational cost. The running time of LR-C is in between those of LR-D( $\Theta$ ) with  $\rho = 0.1$  and  $\rho = 0.01$ . LR-C can be viewed as the limit of LR-D( $\Theta$ ) as  $\rho$  tends to zero and the grid becomes infinitely dense. Proposition 3 in Feldman and Topaloglu (2013) implies that no set of grid points, no matter how dense, can improve on the upper bounds obtained by LR-D( $\Theta$ ) with  $\rho = 0.1, 0.01$  and  $0.001$  by more than 10%, 1% and 0.1%, respectively. It is worthwhile noting that the gains from LR-C are quite close to the theoretical limits.

Problem ( $V, \Phi$ )	LR-D( $\Theta$ ) $\rho = 0.1$ vs LR-C		LR-D( $\Theta$ ) $\rho = 0.01$ vs LR-C		LR-D( $\Theta$ ) $\rho = 0.001$ vs LR-C	
	% Gap in upper bound	Ratio of soln. times	% Gap in upper bound	Ratio of soln. times	% Gap in upper bound	Ratio of soln. times
(5, 0.4)	9.782	0.9	0.992	1.4	0.099	5.9
(5, 0.6)	9.571	0.9	0.987	1.3	0.098	5.0
(5, 0.8)	9.282	0.8	0.977	1.1	0.098	3.6
(10, 0.4)	9.751	0.9	0.992	1.3	0.099	5.6
(10, 0.6)	9.552	0.8	0.987	1.1	0.099	4.1
(10, 0.8)	9.340	0.8	0.979	1.0	0.098	3.3
(20, 0.4)	9.797	0.8	0.993	1.2	0.099	4.6
(20, 0.6)	9.590	0.8	0.988	1.1	0.099	3.7
(20, 0.8)	9.309	0.8	0.980	1.0	0.098	3.1
Average	9.553	0.8	0.986	1.2	0.099	4.3

Table 1: Comparison of the upper bounds and solution times of LR-D( $\Theta$ ) with LR-C.

## REFERENCES

- Bront, J. J. M., Mendez-Diaz, I. and Vulcano, G. (2009), ‘A column generation algorithm for choice-based network revenue management’, *Operations Research* **57**, 769–784.
- Feldman, J. and Topaloglu, H. (2013), Bounding expected revenues for assortment optimization under mixtures of multinomial logits, Technical report, Cornell University, School of OR&IE, Ithaca.
- Kunnumkal, S., Rusmevichientong, P. and Topaloglu, H. (2009), Assortment optimization under the multinomial logit model with product costs, Technical report, Cornell University, School of OR&IE, Ithaca.
- Liu, Q. and van Ryzin, G. (2008), ‘On the choice-based linear programming model for network revenue management’, *M&SOM* **10**(2), 288–310.
- Martinez-de-Albeniz, V. and Kunnumkal, S. (2014), Tractable models and algorithms for assortment planning with product costs, Working paper, IESE Business School, Barcelona.
- McFadden, D. and Train, K. (2000), ‘Mixed mnl models for discrete response’, *Journal of Applied Economics* **15**.
- Mittal, S. and Schulz, A. (2013), ‘A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one’, *Operations Research* **61**.
- Rusmevichientong, P., Shmoys, D., Tong, C. and Topaloglu, H. (2013), ‘Assortment optimization under the multinomial logit model with random choice parameters’, *Production and Operations Management* . (to appear).
- Rusmevichientong, P. and Topaloglu, H. (2012), ‘Robust assortment optimization under the multinomial logit choice model’, *Operations Research* **60**(4).
- Talluri, K. (2014), ‘New formulations for choice network revenue management’, *INFORMS Journal on Computing* .