Assortment Planning under the Multinomial Logit Model with Totally Unimodular Constraint Structures

James Davis
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
jmd338@cornell.edu

Guillermo Gallego
Department of Industrial Engineering and Operations Research,
Columbia University, New York, New York 10027, USA
gmg2@columbia.edu

Huseyin Topaloglu
School of Operations Research and Information Engineering,
Cornell University, Ithaca, New York 14853, USA
topaloglu@orie.cornell.edu

April 10, 2013

Abstract

We consider constrained assortment problems assuming that customers select according to the multinomial logit model (MNL). The objective is to find an assortment that maximizes the expected revenue per customer and satisfies a set of totally unimodular constraints. We show that this fractional binary problem can be solved as an equivalent linear program. We use this result to solve five classes of practical assortment optimization and pricing models under MNL, including (1) assortment models with various bounds on the cardinality of the assortment, (2) assortment models where we need to decide the display location of the selected products, (3) pricing models with a finite menu of possible prices, (4) quality consistent pricing models where the prices of the products have to follow a specified quality ordering, (5) assortment models with precedence constraints. We show that all of these classes of problems can be solved as linear programs. In some instances, constraints can be combined as long as total unimodularity is preserved. In addition, we show how the results extend to a larger class of attraction choice models that avoid some of the shortcomings of MNL.

The problem of selecting assortments to maximize expected profits or expected welfare arises in a variety of industries ranging from transportation and retailing to travel and leisure. There is a growing concern in these industries to find the right set of products given that offering additional products cannibalizes demand for existing products, while the demand lost from excluding products can be partially recaptured among the remaining products. In revenue management, for example, airlines have developed fare admission control policies assuming that demand for different fares are independent. This assumption was tenable when fares were sufficiently differentiated in terms of price and restrictions, but clearly does not hold in the current environment where fares are far more similar and demand substitution and cannibalization are prevalent. For airlines, it is essential to update their pricing and admission control models to include the possibility of demand substitution among fare classes. This is also an important problem in retailing where substitution between products may occur based on prices, product attributes and display locations. Assortment problems often have associated constraints that make them more challenging.

In this paper, we consider constrained assortment optimization problems assuming customers choose according to the multinomial logit model (MNL) or according to other more general attraction choice models. The objective is to maximize the expected revenue obtained from each customer and the constraints on the offered assortment can be captured by a totally unimodular constraint matrix. We formulate this constrained assortment problem as a fractional program with binary decision variables to model the inclusion of products in the assortment. Our main result shows that we can transform this fractional program into a linear program where the integrality constraints can be relaxed because of the totally unimodular nature of the constraints. This result allows us to formulate and solve a variety of practical assortment and pricing problems, the majority of which were not known to be tractable in the literature.

Main Contributions. Let N be the product consideration set from which we want to select an assortment $S \subset N$ to offer to customers. An assortment can be identified by its incident vector $x = \{x_j : j \in N\} \in \{0,1\}^{|N|}$, where $x_j = 1$ if $j \in S$ and $x_j = 0$ if $j \notin S$. The set of feasible assortments is given by $\mathcal{F} = \{x \in \{0,1\}^{|N|} : \sum_{j \in N} a_{ij}\, x_j \leq b_i \ \forall\, i \in M\}$ for a totally unimodular matrix $[a_{ij}]_{i \in M, j \in N}$. Given feasible product offer decisions $x \in \mathcal{F}$, each customer chooses among the offered products according to MNL. The objective is to choose $x \in \mathcal{F}$, a feasible set of products to offer, to maximize the expected revenue obtained from each customer. Although this assortment problem is a fractional program with binary decision variables, our main result shows that it can directly be solved as a linear program.

Building on our main result, we show how to solve five classes of practical assortment and pricing problems. First, we work with assortment problems under MNL where there are cardinality constraints on the assortment. Rusmevichientong, Shen and Shmoys (2010) consider assortment problems with a limit on the total number of offered products. Their approach generates candidate assortments and checks the performance of the candidates, whereas we give a direct linear programming formulation. Our approach also extends to more general cardinality constraints. In particular, our results apply when the set of products are partitioned into a number of subsets and we limit the number of products offered in each subset. Such constraints occur when, for example, television sets are partitioned as small, medium and large and we limit the number of offered television sets of each size. We can also handle overlaps between successive subsets. That is, if some television sets can be considered as both small and medium and some can be considered as both medium and large, then we can limit the numbers of offered small, medium and large television sets. We can deal with nested cardinality constraints as well, which occur when products are categorized into subsets such that one subset includes another one and we limit the number of products offered in each subset. For example, some products may be specialty products and we may want to offer at most k specialty products, while limiting the number of all offered products to ℓ. Such overlapping or nested cardinality constraints are not studied in the earlier work.

Second, we consider assortment problems with display location effects, where the attractiveness of a product depends not only on its own attributes but also on the location at which it is displayed. Such problems arise when products are displayed in a window or a shelf and the products with prime locations may have a better chance of attracting attention, which can be modeled by a higher attractiveness parameter in MNL. Another relevant setting is online retail, where the propensity of a customer to choose a product depends not only on the attributes of the product, but also on where the product is shown within search results or web site. We show that the assortment optimization problem with display specific attractiveness parameters can be formulated as a linear program. To our knowledge, tractability of the assortment problem with display location effects was not previously known.

Third, we consider pricing problems under MNL, when there are a finite number of possible price levels for the products and the attractiveness of a product depends on its price. The objective is to choose the product prices to maximize the expected revenue from each customer. We show how to obtain an optimal solution to this problem by using a linear program. Earlier pricing formulations assume that the attractiveness parameter of product j is a parametric function $e^{\alpha_j - \beta_j p}$ of the price p of this product, where α_j and β_j are fixed coefficients, in which case, the pricing problem can be formulated as a continuous optimization problem involving the product prices. In our approach, the attractiveness of a product depends on its price arbitrarily, not limited to the parametric form $e^{\alpha_j - \beta_j p}$. Also, since we work with a finite number of possible price levels, we can put explicit restrictions on the product prices. For example, we can limit ourselves to prices in increments of a dollar or make sure that the prices of the products are chosen within specific intervals.

Fourth, we consider quality consistent pricing problems. In such pricing problems, there is an inherent ordering between the products, where some products are considered lower in quality than others. The prices should be chosen to be quality consistent, so that lower quality products are priced lower than higher quality products. We show how to obtain the optimal solution to such quality consistent pricing problems. The quality consistent pricing terminology is introduced by Gallego and Stefanescu (2009) in the context of MNL, but to our knowledge, there is no work on finding optimal prices under quality consistency constraints when customers choose according to MNL. Quality consistency constraints are especially important when there is excessive demand for lower quality products, in which case, one might price lower quality products higher than higher quality products when quality consistency constraints are ignored. Finally, we consider assortment problems with product precedence constraints where a product cannot be offered unless a certain set of related products are offered. We give a linear programming formulation of the assortment problem under such precedence constraints.

Related Literature. In this paper, we use MNL to capture the customer choice process, which is a special case of the axiomatic attraction model by Luce (1959). It is possible to show that MNL is based on utility maximizing customer behavior. Consider the case where a customer associates a random utility with each product and a random utility with the no purchase option. The customer, being a utility maximizer, chooses the option providing the largest utility. If the random utilities are independent and have a Gumbel distribution, then it is possible to show that the resulting choice model corresponds to MNL; see McFadden (1974). Due to the independence of the utilities, MNL possesses the independence of irrelevant alternatives property, which refers to the fact that if a product is added to the offered assortment, then the market share of all other offered products decreases by the same relative amount. Clearly, this property should not hold when products cannibalize each other to different extents; see Ben-Akiva and Lerman (1994).

Despite the independence of irrelevant alternatives property, MNL has some advantages. In many applications, the mean utility of a product is modeled as a linear combination of its observable features, such as price and quality. If the observable features are adequate to capture the means of the utilities that a customer associates with the products, then the residuals between the utilities and their means may be considered as independent noise and MNL becomes an appealing candidate for capturing customer choices. The parameters of MNL can be estimated efficiently from customer choice data as the corresponding estimation problem has a concave log likelihood function; see McFadden (1974). There exist likelihood ratio tests to check the goodness of fit between the data and the fitted MNL; see Hausman and McFadden (1984). The estimated parameters have economic interpretations to assess tradeoffs between different product features. For example, we can compute how much a customer is willing to pay for a unit of increase in the durability of a product. Finally, if the modeler believes that the independence of irrelevant alternatives property is indeed satisfied by the customer choice process, then she can exploit it during model estimation. Train (2003) describes a situation where data have been collected on the choices of customers among a set of products N. If we want to estimate the choice process of customers among a smaller set of products $S \subset N$, then we can simply discard the data on customers that did not choose a product in the set S. Due to independence of irrelevant alternatives, one can show that the choice model estimated in this fashion captures the choice process among the set of products S without bias.

Our linear programming formulation under MNL extends to a more general attraction model that mitigates some of the shortcomings of MNL. Unlike MNL, the purchase probabilities in the general attraction model depend on all of the products, both those offered and those not offered. This model was proposed by Gallego et al. (2011) and involves a shadow attraction value for each product that impinges on the choice probabilities when the products are not offered. It turns out that a simple transformation of our formulation for MNL allows for the assortment problem to be solved under this more general attraction model.

Assortment problems under variants of MNL is closely related to our work. Talluri and van Ryzin (2004) study assortment problems under MNL without any constraints and show that the optimal assortment can be obtained by greedily adding products into the offered assortment in the order of decreasing revenues. Rusmevichientong, Shen and Shmoys (2010) consider cardinality constraints on the offered assortment, whereas Rusmevichientong et al. (2009) assume that each product has a space requirement and they limit the total space requirement of the offered assortment. Gallego et al. (2011) give a linear programming formulation for the assortment problem under MNL without considering constraints on feasible assortments and extend their formulation to the general attraction model described in the previous paragraph. Bront et al. (2009), Mendez-Diaz et al. (2010) and Rusmevichientong, Shmoys and Topaloglu (2010) consider the case where there are multiple customer types and customers of each type choose according to a different MNL. They provide heuristics, integer programming formulations and approximation methods. Davis et al. (2011) study assortment problems under the nested logit model without any constraints, whereas Rusmevichientong et al. (2009) and Gallego and Topaloglu (2012) consider constraints under the nested logit model. Gallego et al. (2004), Liu and van Ryzin (2008), Kunnumkal and Topaloglu (2008), Zhang and Adelman (2009), Talluri (2011), Meissner and Strauss (2012) and Meissner et al. (2012) use assortment problems to make extensions to network revenue management, where itinerary products consume the capacities on flight legs in bundles.

For pricing problems, Hanson and Martin (1996) observe that the expected revenue is not a concave function of prices when customers choose according to MNL, but Song and Xue (2007) and Dong et al. (2009) are able to recover a concave objective function by using the market share of each product as the decision variable. Li and Huh (2011) solve pricing problems under the nested logit model. They also use market shares as decision variables, but work under the assumption that the products in the same nest share the same price sensitivity. Gallego and Wang (2011) study the same problem, but they relax the assumption that the price sensitivities in a nest are the same. Chen and Hausman (2000) and Wang (2012) study joint assortment and pricing problems, where one chooses the products to offer and their corresponding prices.

Organization. In Section 1, we formulate our assortment problem and give our main result, providing an equivalent linear programming formulation. In Section 2, we use our main result to show how to solve assortment problems with cardinality constraints, display location effects and product precedence constraints, along with pricing problems with a finite price menu and quality consistency constraints. In Section 3, we point out possible extensions of our work, including extensions to the general attraction model. In Section 4, we conclude.

1 Problem Formulation and Main Result

The set of products is N. The revenue and the preference weight associated with product j are respectively r_j and v_j. To capture the product offer decisions, we define the decision variable $x_j \in \{0,1\}$ such that $x_j = 1$ if we offer product j, otherwise $x_j = 0$. Under MNL, if the product offer decisions are given by the vector $x = \{x_j : j \in N\} = \{0,1\}^{|N|}$, then a customer purchases product j with probability $P_j(x) = v_j\, x_j / (1 + \sum_{k \in N} v_k\, x_k)$, where we normalize the preference weight of the no purchase option to one. Therefore, if the products that we offer correspond to the vector x, then the expected revenue obtained from a customer can be written as

$$R(x) = \sum_{j \in N} r_j\, P_j(x) = \frac{\sum_{j \in N} r_j\, v_j\, x_j}{1 + \sum_{j \in N} v_j\, x_j}. \tag{1}$$

Without loss of generality, if selling a unit of product j involves a constant cost for each unit sold, then we assume that r_j is the profit from the sale, given by the difference between revenue and cost. For a totally unimodular matrix $A = [a_{ij}]_{i \in M, j \in N}$ with dimensions $|M| \times |N|$, the feasible set of product offer decisions are given by $\mathcal{F} = \{x \in \{0,1\}^{|N|} : \sum_{j \in N} a_{ij}\, x_j \leq b_i \;\forall\, i \in M\}$. For the moment, we do not go into the specific structure of the matrix A, but we give specific examples for A in the next section. Our goal is to find a set of feasible products to offer so as to maximize the expected revenue obtained from each customer, yielding the problem

$$z^* = \max_{x \in \mathcal{F}} R(x). \tag{2}$$

Our use of "less than" constraints to capture the feasible set of product offer decisions is without loss of generality. A "greater than" constraint can be multiplied by -1 to get a "less than" constraint and an equality constraint can be replaced a pair of "less than" and "greater than" constraints. Recalling that multiplying a row of a matrix by -1 or duplicating a row of a matrix does not change its total unimodularity properties, we end up with "less than" constraints and a totally unimodular constraint matrix after transforming equality and "greater than" constraints to "less than" constraints; see Proposition 2.1 in Chapter III.1 of Nemhauser and Wolsey (1988).

Problem (2) has a nonlinear objective function and integrality requirements on its decision variables. Our main result shows that problem (2) is equivalent to the problem

$$\begin{aligned} \max \quad & \sum_{j \in N} r_j\, w_j && \qquad (3) \\ \text{st} \quad & \sum_{j \in N} w_j + w_0 = 1 \\ & \sum_{j \in N} a_{ij}\, \frac{w_j}{v_j} \leq b_i\, w_0 && \forall\, i \in M \\ & 0 \leq \frac{w_j}{v_j} \leq w_0 && \forall\, j \in N, \end{aligned}$$

where the decision variables are $\{w_j : j \in N \cup \{0\}\}$. The problem above is a linear program. In this problem, we interpret the decision variable w_j as the probability that a customer purchases product

j and w_0 as the probability that a customer leaves without making a purchase. The first constraint ensures that a customer either purchases a product or leaves without purchasing. Interestingly, the second set of constraints are enough to ensure that the product offer decisions are chosen within the feasible set $\mathcal{F}$. The third set of constraints ensure the connection between the probability that a customer purchases a product and leaves without purchasing anything. In the next theorem, we show that problems (2) and (3) are equivalent to each other.

Theorem 1. *Problems* (2) *and* (3) *have the same optimal objective value and we can construct an optimal solution to one of these problems by using an optimal solution to the other.*

Proof. Using the decision variables $y = \{y_j : j \in N\} \in \{0,1\}^{|N|}$, we claim that problem (2) is equivalent to the problem

$$
\begin{aligned}
\max \quad & \sum_{j \in N} (r_j - z^*)\, v_j\, y_j && && (4)\\
\text{st} \quad & \sum_{j \in N} a_{ij}\, y_j \le b_i && \forall\, i \in M\\
& 0 \le y_j \le 1 && \forall\, j \in N,
\end{aligned}
$$

which is a linear program. To see this equivalence, let x^* and y^* respectively be optimal solutions to problems (2) and (4). Since $z^* = R(x^*) = \sum_{j \in N} r_j\, v_j\, x_j^* / (1 + \sum_{j \in N} v_j\, x_j^*)$, we get $\sum_{j \in N} r_j\, v_j\, x_j^* = z^*\,(1 + \sum_{j \in N} v_j\, x_j^*)$. In this case, evaluating the objective value of problem (4) at the feasible solution x^*, we obtain $\sum_{j \in N} (r_j - z^*)\, v_j\, x_j^* = z^*\,(1 + \sum_{j \in N} v_j\, x_j^*) - z^* \sum_{j \in N} v_j\, x_j^* = z^*$, which implies that the optimal objective value of problem (4) is at least as large as the optimal objective value of problem (2). On the other hand, since A is totally unimodular and the objective function of problem (4) is linear, we can assume that $y^* \in \{0,1\}^{|N|}$ without loss of generality. Thus, $y^* \in \mathcal{F}$. In this case, evaluating the objective value of problem (2) at the feasible solution y^*, we have $z^* \ge R(y^*) = \sum_{j \in N} r_j\, v_j\, y_j^* / (1 + \sum_{j \in N} v_j\, y_j^*)$. Focusing on the first and last expressions in the last chain of inequalities and arranging the terms, we get $z^* \ge \sum_{j \in N} (r_j - z^*)\, v_j\, y_j^*$, which implies that the optimal objective value of problem (4) is at most as large as the optimal objective value of problem (2). Thus, problems (2) and (4) are equivalent to each other, sharing the same optimal objective value, establishing the claim. So, it is enough to show that problems (3) and (4) are equivalent to each other, which is what we do in the rest of the proof.

We let $w^* = \{w_j^* : j \in N \cup \{0\}\}$ be an optimal solution to problem (3) with objective value ζ^* and y^* be an optimal solution to problem (4). The discussion above shows that y^* provides the objective value z^* for problem (4). We construct the solution $\hat{w} = \{\hat{w}_j : j \in N \cup \{0\}\}$ to problem (3) as $\hat{w}_j = P_j(y^*) = v_j\, y_j^* / (1 + \sum_{i \in N} v_i\, y_i^*)$ for all $j \in N$ and $\hat{w}_0 = 1 - \sum_{j \in N} P_j(y^*) = 1/(1 + \sum_{i \in N} v_i\, y_i^*)$. Since y^* is feasible to problem (4), it is simple to check that $\hat{w}$ is feasible to problem (3). In particular, $\hat{w}$ clearly satisfies the first constraint in problem (3). Furthermore, noting that $\hat{w}_j / (v_j\, \hat{w}_0) = y_j^*$, we have $b_i \ge \sum_{j \in N} a_{ij}\, y_j^* = \sum_{j \in N} a_{ij}\, \hat{w}_j / (v_j\, \hat{w}_0)$ showing that $\hat{w}$ satisfies the second set of constraints in problem (3). Finally, noting that $1 \ge y_j^* = \hat{w}_j / (v_j\, \hat{w}_0)$, $\hat{w}$

satisfies the third set of constraints in problem (3). In this case, the objective value provided by the feasible solution $\hat{w}$ to problem (3) satisfies $\zeta^* \geq \sum_{j \in N} r_j \, \hat{w}_j = \sum_{j \in N} r_j \, P_j(y_j^*) = R(y^*) = z^*$. So, we have $\zeta^* \geq z^*$. On the other hand, we construct the solution $\hat{y} = \{\hat{y}_j : j \in N\}$ to problem (3) as $\hat{y}_j = w_j^* / (v_j \, w_0^*)$ for all $j \in N$. By using an argument similar to the one we just used, it is possible to show that $\hat{y}$ is a feasible solution to problem (4). In this case, the objective value provided by the feasible solution $\hat{y}$ to problem (4) satisfies $z^* \geq \sum_{j \in N} r_j \, v_j \, \hat{y}_j - z^* \sum_{j \in N} v_j \, \hat{y}_j = \sum_{j \in N} r_j \, w_j^* / w_0^* - z^* \sum_{j \in N} w_j^* / w_0^* \geq \sum_{j \in N} r_j \, w_j^* / w_0^* - \zeta^* \sum_{j \in N} w_j^* / w_0^* = (\zeta^* - \zeta^*(1 - w_0^*)) / w_0^* = \zeta^*$, where the second inequality uses the fact that $\zeta^* \geq z^*$ shown above and the second equality uses the fact that $\zeta^* = \sum_{j \in N} r_j \, w_j^*$ and $\sum_{j \in N} w_j^* = 1 - w_0^*$ by the first constraint in problem (3). So, we have $z^* \geq \zeta^*$, establishing that the solutions w^*, y^*, $\hat{w}$ and $\hat{y}$ all provide the objective value z^* for their respective problems. Given w^*, the solution $\hat{y}$ constructed through w^* is optimal to problem (4) and given y^*, the solution $\hat{w}$ constructed through y^* is optimal to problem (3). $\square$

Theorem 1 shows that problems (2) and (3) are equivalent and we can obtain an optimal solution to the former problem simply by solving the latter. This result is quite useful in practice since problem (3) is a linear program.

2 Applications

In this section, we give a number of specific cases where we can use Theorem 1 to obtain the optimal solutions to certain assortment optimization and pricing problems under MNL with a variety of constraints. For majority of these cases, efficient algorithms for obtaining the optimal solution do not exist in the earlier literature. Our results appear to provide the first tractable algorithms to obtain the optimal solution.

2.1 Cardinality Constraints

Consider the case where the total number of products that can be offered is limited to b. So, the feasible set of product offer decisions can be written as $\mathcal{F} = \{x \in \{0,1\}^{|N|} : \sum_{j \in N} x_j \leq b\}$, in which case, A is given by $(1, \ldots, 1)$ with dimensions $1 \times |N|$. This matrix is clearly totally unimodular, which implies that we can find the optimal set of products to offer under a cardinality constraint directly by solving problem (3) with $A = (1, \ldots, 1)$. Rusmevichientong, Shen and Shmoys (2010) give an efficient algorithm for finding the optimal set of products to offer under MNL with a cardinality constraint, but our use of Theorem 1 allows us to solve this problem directly by using a linear program. Furthermore, by building on this theorem, we can find the optimal solution under more general cardinality constraints. We proceed to giving two examples of such more general cardinality constraints below.

Consider the case where there are K nested subsets of products such that $S_1 \subset S_2 \subset \ldots \subset S_K \subset N$ and there are integers $b_1 \leq b_2 \leq \ldots \leq b_K$ associated with each one of these subsets. The total number of products that we can offer in subset S_k is limited to b_k. These constraints may arise when, for example, S_1 corresponds to the specialty products, whereas S_2 corresponds to the set of all products and we do not want to offer more than a total of b_2 products, while limiting the number of offered specialty products to b_1. Thus, the feasible set of product offer decisions can be written as $\mathcal{F} = \{x \in \{0,1\}^{|N|} : \sum_{j \in S_k} x_j \leq b_k \; \forall \, k = 1, \ldots, K\}$. Since $S_1 \subset S_2 \subset \ldots \subset S_K$, this constraint matrix includes consecutive ones in each row. Such matrices are called interval matrices and they are known to be totally unimodular; see Corollary 2.10 in Chapter III.1 of Nemhauser and Wolsey (1988). Thus, we can solve a linear program to find the optimal assortment to offer under nested cardinality constraints.

As another example, consider the case where the products are partitioned into K disjoint subsets $S_1, S_2, \ldots, S_K$. Without loss of generality, we assume that the products are indexed by the integers $\{1, \ldots, n\}$ and we have $S_k = \{i_k, \ldots, i_{k+1} - 1\}$ with $i_1 = 1$ and $i_{K+1} = n + 1$. The total number of products we can offer in subset S_k is limited to b_k. In this case, the feasible set of product offer decisions is $\mathcal{F} = \{x \in \{0,1\}^{|N|} : \sum_{j=i_k}^{i_{k+1}-1} x_j \leq b_k \; \forall \, k = 1, \ldots, K\}$ and this constraint matrix still corresponds to an interval matrix. Such constraints may arise when, for example, S_1 corresponds to the specialty products, whereas S_2 corresponds to the general interest products and we separately want to limit the numbers of offered specialty and general interest products. Furthermore, if S_k overlaps with only S_{k-1} and S_{k+1}, then the constraint matrix is still an interval matrix. So, even if some products may count both as specialty and general interest products, we can still limit the numbers of offered specialty and general interest products. To our knowledge, it is difficult to capture such overlapping cardinality constraints with earlier frameworks.

2.2 Display Location Effects

Consider the case where the preference weight of a product depends on where it is displayed. For example, if the products are displayed in a store window or a shelf, then customers may tend to overlook a product when it is displayed at the back of the window or the shelf, which can be captured by a smaller preference weight when the product is displayed at the back. In online retail, customers may be more likely to choose products that are displayed at the top of search results, which can be captured by using preference weights that depend on the display order of the product. It turns out we can build on Theorem 1 to find the optimal set of products to offer when the preference weights depend on the display location.

We use v_{jl} to denote the preference weight of item j when this item is displayed at location l. Without loss of generality, we assume that there are as many possible locations as the number of items so that we can offer all items at once. In this case, we can index both the items and the locations by N. If the number of possible locations is smaller than the number of items, then we can define additional locations with $v_{jl} = 0$ for all $j \in N$, for each additional location l, in which case, using one of these additional locations for an item is equivalent to not displaying the item at all. To capture the product offer decisions, we use $x = \{x_{jl} : j, l \in N\} \in \{0,1\}^{|N| \times |N|}$, where $x_{jl} = 1$ if we offer item j in location l, otherwise $x_{jl} = 0$. If the product offer decisions are given by x, then we obtain an expected revenue of $\sum_{j,l \in N} r_j \, v_{jl} \, x_{jl} / (1 + \sum_{j,l \in N} v_{jl} x_{jl})$. Therefore, we are interested in solving the problem

$$
\begin{aligned}
\max \quad & \frac{\sum_{j,l \in N} r_j \, v_{jl} \, x_{jl}}{1 + \sum_{j,l \in N} v_{jl} \, x_{jl}} \\
\text{st} \quad & \sum_{l \in N} x_{jl} \le 1 \qquad && \forall \, j \in N \\
& \sum_{j \in N} x_{jl} \le 1 \qquad && \forall \, l \in N \\
& x_{jl} \in \{0, 1\} \qquad && \forall j, l \in N,
\end{aligned}
$$

where the first set of constraints ensure that each item is offered at most in one location and the second set of constraints ensure that each location is used by at most one item. The problem above is a special case of problem (2) where each product is indexed by $(j, l) \in N \times N$. Its constraint matrix is the constraint matrix of an assignment problem, which is totally unimodular; see Corollary 2.9 in Chapter III.1 of Nemhauser and Wolsey (1988). So, by Theorem 1, we can use a linear program to find the optimal set of items to offer under display location effects.

2.3 Pricing with a Finite Price Menu

Consider the case where the price of a product is a decision variable, rather than being fixed. The preference weight of a product depends on its price. Increasing the price of a product is expected to make it less desirable to customers, effectively decreasing its preference weight, but we are not strictly tied to the assumption that higher prices result in lower preference weights. The goal is to choose the prices of the products so as to maximize the expected revenue from each customer. We show how to solve this pricing problem as a linear program as long as the prices are chosen within a finite set of possible price levels.

We let K be the set of possible price levels for an item. The price corresponding to price level k for an item is given by r_k. Therefore, $\{r_k : k \in K\}$ becomes the possible prices for an item. If we use the price level k for item j, then its preference weight is v_{jk}. Our notation indicates that the set of possible prices for each item is the same, but it is straightforward to extend our formulation to incorporate different sets of possible prices for different items. To capture our pricing decisions, we use $x = \{x_{jk} : j \in N, \; k \in K\} \in \{0, 1\}^{|N| \times |K|}$, where $x_{jk} = 1$ if we set the price of item j at price level k, otherwise $x_{jk} = 0$. Thus, we want to solve the problem

$$
\begin{aligned}
\max \quad & \frac{\sum_{j \in N} \sum_{k \in K} r_k \, v_{jk} \, x_{jk}}{1 + \sum_{j \in N} \sum_{k \in K} v_{jk} \, x_{jk}} \\
\text{st} \quad & \sum_{k \in K} x_{jk} = 1 \qquad && \forall \, j \in N \\
& x_{jk} \in \{0, 1\} \qquad && \forall \, j \in N, \; k \in K,
\end{aligned}
$$

where the constraints ensure that each item is offered at one price level. Similar to our observations for display location effects, the problem above is a special case of problem (2) where each product is indexed by $(j,k) \in N \times K$. Each row of the constraint matrix corresponds to an item j and it includes consecutive ones, corresponding to the different price levels for item j. Thus, the constraint matrix is an interval matrix, which is totally unimodular. In this case, using Theorem 1, we can find the optimal prices by solving a linear program. Our formulation above assumes that each product has to be offered. If we need to jointly decide which products to offer and the prices of the offered products, then we can simply replace the equality constraint in the problem above with a "less than" constraint. Furthermore, if we want to impose a limit of b on the number of products we offer, then we can add the constraint $\sum_{j \in N} \sum_{k \in K} x_{jk} \leq b$. The additional constraint amounts to adding a row of ones to the constraint matrix, which does not change the fact that the constraint matrix is an interval matrix.

Pricing models traditionally assume a parametric relationship between the price of an item and its preference weight. In particular, it is usually assumed that if the price of item j is p, then its preference weight is $e^{\alpha_j - \beta_j p}$, for some constants α_j and β_j. Using this parametric form, the pricing problem can be formulated as a smooth optimization problem, involving prices as the decision variables, but the objective function of this problem is not concave. Song and Xue (2007) and Dong et al. (2009) formulate the problem in terms of the market share of an item to get a concave objective function, but their work is based on the specific parametric relationship between price and preference weight. In our formulation, the preference weight of an item can depend on its price in an arbitrary fashion. Furthermore, since we work with discrete price levels, we can limit attention to operationally appealing prices, such as those in increments of a dollar.

2.4 Quality Consistent Pricing

Similar to our pricing model above, consider the case where the prices of the products are decision variables, but there is an inherent ordering between the products in terms of their quality. In particular, the products are indexed such that the first product is lower quality than the second one, the second product is lower quality than the third one and so on. When setting the prices of the products, we need to ensure that the lower quality products have lower prices. Such a pricing scheme is called quality consistent pricing or price laddering. It occurs when products have a clear ordering in terms of quality, richness of features or durability. For example, Rusmevchientong et al. (2006) describe an application where option rich automobiles of the same model have to be priced higher than the option poor ones. The objective is to choose the prices of the products to maximize the expected revenue from each customer, while adhering to the quality consistency constraint.

We index the items by $N = \{1, 2, \ldots, n\}$ and the possible price levels for an item by $K = \{1, \ldots, m\}$. The price corresponding to price level k of an item is r_k. Therefore, the possible price levels for an item are given by $\{r_k : k \in K\}$. Without loss of generality, we assume that the possible price levels are indexed such that $r_1 \leq r_2 \leq \ldots \leq r_m$ so that lower indices correspond to lower prices. Also, we assume that the items are indexed such that the first item is the lowest quality one and the last item is the highest quality one. So, the price of the first item must be smaller than the price of the second item, which, in turn, must be smaller than the price of the third item and so on. We let v_{jk} be the preference weight of item j when we price this item at price level k. To capture our pricing decisions, we use $x = \{x_{jk} : j \in N,\ k \in K\} \in \{0,1\}^{|N|\times|K|}$, where $x_{jk} = 1$ if we price item j at price level k, otherwise $x_{jk} = 0$. To impose the quality consistency constraint, we use the additional decision variables $z = \{z_{jk} : j \in N,\ k \in K \setminus \{m\}\} \in \{0,1\}^{|N|\times(|K|-1)}$, where $z_{jk} = 1$ if we price item j at price level $k+1$ or higher, otherwise $z_{jk} = 0$. Note that we do not need the decision variables $\{z_{jm} : j \in N\}$. We want to solve the problem

$$
\begin{aligned}
\max \quad & \frac{\sum_{j\in N}\sum_{k\in K} r_k\, v_{jk}\, x_{jk}}{1+\sum_{j\in N}\sum_{k\in K} v_{jk}\, x_{jk}} \\
\text{st} \quad & x_{11} + z_{11} = 1 \\
& x_{1k} + z_{1k} = z_{1,k-1} && \forall\, k = 2,\dots,m-1 \\
& x_{1m} = z_{1,m-1} \\
& x_{j1} + z_{j1} = x_{j-1,1} && \forall\, j = 2,\dots,n \\
& x_{jk} + z_{jk} = x_{j-1,k} + z_{j,k-1} && \forall\, j = 2,\dots,n,\ k = 2,\dots,m-1 \\
& x_{jm} = x_{j-1,m} + z_{j,m-1} && \forall\, j = 2,\dots,n \\
& x_{jk}, z_{jl} \in \{0,1\} && \forall\, j \in N,\ k \in K,\ l \in K \setminus \{m\}.
\end{aligned}
$$

We give interpretations for the last three sets of constraints above and the first three sets can be interpreted in a similar fashion. The fourth set of constraints ensure that if item $j-1$ is priced at level 1, then item j is either priced at level 1 or it is priced at level 2 or higher. The fifth set of constraints ensure that if item $j-1$ is priced at level k or if we decide to price item j at level k or higher, then item j is either priced at level k or it is priced at level $k+1$ or higher. The sixth set of constraints ensure that if item $j-1$ is priced at level m or if we decide to price item j at level m or higher, then item j is priced at level m.

The constraint matrix above corresponds to the constraints of a shortest path problem. To see this result, consider a network composed of the nodes $\{(j,k) : j \in N,\ k \in K\}$ and a sink node. Figure 1 shows a sample network with $n = 4$ and $m = 3$. In the problem above, the decision variable x_{jk} for $j \in N \setminus \{n\}$, $k \in K$ corresponds to an arc from node (j,k) to node $(j+1,k)$. The decision variable x_{nk} for $k \in K$ corresponds to an arc from node (n,k) to the sink node. Finally, the decision variable z_{jk} for $j \in N$, $k \in K \setminus \{m\}$ corresponds to an arc from node (j,k) to node $(j,k+1)$. The first three sets of constraints are the flow balance constraints for the nodes in $\{(1,k) : k \in K\}$, whereas the last three sets of constraints are the flow balance constraints for the nodes in $\{(j,k) : j \in N \setminus \{1\},\ k \in K\}$. The flow balance constraint of the sink node is redundant and it is omitted. The supply of node $(1,1)$ is one. So, the horizontal arcs in Figure 1 correspond to the decision variables $\{x_{jk} : j \in N,\ k \in K\}$, whereas the vertical arcs correspond to the decision variables $\{z_{jk} : k \in N,\ k \in K \setminus \{m\}\}$. The important observation is that if a unit of flow from node $(1,1)$ to the sink node follows the arc corresponding to the decision variable x_{jk}, then it can never follow the arc corresponding to the decision variable $x_{j+1,k-1}$, which means that if we choose price level k for item j, then we cannot choose price level $k-1$ for item $j+1$, which is enough to impose the price ladder. Lastly, we observe that if we add the first three sets of constraints above, then we obtain $\sum_{k \in K} x_{1k} = 1$, whereas if we add the last three sets of constraints over $k \in K$ for a particular item j, then we obtain $\sum_{k \in K} x_{jk} = \sum_{k \in K} x_{j-1,k}$. Thus, the constraints above ensure that $\sum_{k \in K} x_{jk} = 1$ for all $j \in N$, indicating that we offer each item at one price level.

Network matrices are totally unimodular by Proposition 3.1 in Chapter III.1 of Nemhauser and Wolsey (1988). So, by Theorem 1, we can solve quality consistent pricing problems by using a linear program. To our knowledge, tractability of incorporating quality consistency or laddering constraints into pricing problems under MNL was not known previously.

The development in this section assumes that the qualities of the items satisfy the full ordering $1 \preceq 2 \preceq \ldots \preceq n$, in which case, the prices of the items have to satisfy this ordering as well. By modifying the network in Figure 1 slightly, we can handle the case where we have a partial ordering between the qualities of the items. For example, assuming that there are five items, the qualities of the items may satisfy the partial ordering $1 \preceq \{2, 3, 4\} \preceq 5$, which is to say that the first item is lower quality than the second, third and fourth items, which are, in turn, lower quality than the fifth item, but there is no clear ordering between the qualities of the second, third and fourth items. In this case, the price of the first item should be lower than the prices of the second, third and fourth items. Similarly, the prices of the second, third and fourth items should be lower than the price of the fifth item, but there is no constraint on how the prices of the second, third and fourth items are ordered. Building on the approach described in this section, it is possible to show that we can still solve a linear program to enforce such a partial quality consistency constraint. Finally, our approach continues to apply when there are disjoint quality consistency constraints in the sense that the items are partitioned into disjoint subsets $S_1, \ldots, S_L$ and we impose a separate quality consistency constraint for the items in each one of the subsets S_l for $l = 1, \ldots, L$.

2.5 Product Precedence Constraints

We consider assortment optimization problems where a particular product cannot be offered to customers unless a certain set of related products are also offered. This kind of a constraint can arise when a company offers multiple versions of a product and company policy or law prohibits offering a more expensive or sophisticated version of the product unless a more inexpensive or basic version is offered. For example, it may not be possible to offer the brand name version of a drug unless the generic version is offered. To model such product precedence constraints, we use $S_j \subset N$ to denote the set of products that we need to offer to be able to offer product j. So, the feasible set of product offer decisions is given by $\mathcal{F} = \{x \in \{0,1\}^{|N|} : x_j - x_i \leq 0 \ \forall\, j \in N,\ i \in S_j\}$, indicating that we can have $x_j = 1$ only when $x_i = 1$ for all $i \in S_j$. In this constraint matrix, each row includes only a $+1$ and a -1. Such matrices are known to be totally unimodular; see Proposition 2.6 in Chapter III.1 of Nemhauser and Wolsey (1988), along with Heller (1957) and Heller and Hoffman (1962). Thus, by Theorem 1, it follows that we can find the optimal assortment under product precedence constraints by solving a linear program. We observe that the subsets $\{S_j : j \in N\}$ in product precedence constraints can be completely arbitrary. In particular, they can be overlapping and products can have circular dependencies on each other.

Closing this section, we note that if we want to enforce offering a certain product in any of the applications considered in this section, then we can impose a lower bound of one on the sum of the decision variables corresponding to the offer decisions for this product. We can check that the constraint matrix remains totally unimodular under this additional constraint. Furthermore, joining two totally unimodular matrices does not yield a totally unimodular matrix in general, but it may be possible to combine some of the constraints considered in this section without destroying the total unimodularity of the constraint matrix. For example, joining two interval matrices yields an interval matrix. Since a cardinality constraint and pricing with a finite price menu both yield interval constraint matrices, we can impose a cardinality constraint on the offered assortment when solving a pricing problem with a finite price menu.

3 Extensions of the Theory

In this section, we give some additional implications of Theorem 1 that may potentially become useful in broader settings. In particular, we show how to extend this theorem when customers choose according to the general attraction model proposed by Gallego et al. (2011), rather than MNL. Furthermore, we investigate working with alternative expected revenue functions.

3.1 General Attraction Model

Gallego et al. (2011) propose a general attraction model to describe the customer choice process that can alleviate some of the shortcomings of MNL. The main idea behind this choice model is that if a certain product is not in the offered assortment, then the absence of this product can potentially increase the propensity of a customer to leave without making a purchase. To pursue this line of thought, the general attraction model associates two preference weights $\tilde{v}_j$ and $\tilde{w}_j$ with product j. If product j is offered, then its preference weight is $\tilde{v}_j$, but if product j is not offered, then the preference weight of the no purchase option increases by $\tilde{w}_j$. Thus, under the general attraction model, if the product offer decisions are given by the vector $x = \{x_j : j \in N\} \in \{0,1\}^{|N|}$, then a customer purchases product j with probability

$$\tilde{P}_j(x) = \frac{\tilde{v}_j \, x_j}{1 + \sum_{k \in N} \tilde{w}_k \, (1 - x_k) + \sum_{k \in N} \tilde{v}_k \, x_k},$$

where the expression $\sum_{k \in N} \tilde{w}_k \, (1 - x_k)$ captures the increase in the preference weight of the no purchase option as a function of the products that are not offered. Gallego et al. (2011) justify the choice probability above by using an axiomatic approach that is similar to the one used by Luce (1959). Their axiomatic justification requires that $\tilde{v}_j \geq \tilde{w}_j$ for all $j \in N$. Furthermore, the authors show that the specific form of the choice probability above can be derived by using a random utility based choice model. In particular, the customers are utility maximizers. When making their choices among an assortment of products offered by a company, the customers associate random utilities with the products offered by the company, a random utility with the no purchase option and random utilities with the products offered by the competitors. Each customer follows the option providing the largest utility. If a customer decides to leave without making a purchase or chooses a product that is offered by a competitor, then the company does not obtain any sale from this customer. Assuming that the random utilities are independent and they have a Gumbel distribution, it is possible to show that we obtain the general attraction model. If the customers choose according to the general attraction model and the products that we offer correspond to the vector x, then the expected revenue obtained from a customer is

$$\tilde{R}(x) = \sum_{j \in N} r_j \, \tilde{P}_j(x) = \frac{\sum_{j \in N} r_j \, \tilde{v}_j \, x_j}{1 + \sum_{j \in N} \tilde{w}_j \, (1 - x_j) + \sum_{j \in N} \tilde{v}_j \, x_j}.$$

We want to solve the assortment problem $\max_{x \in \mathcal{F}} \tilde{R}(x)$, where $\mathcal{F} \subset \{0,1\}^{|N|}$ is the feasible set of product offer decisions as defined in Section 1. We can make a transformation on the expected revenue function above, which ultimately allows us to use Theorem 1 to solve this assortment problem. To see this transformation, we use the definition of $\tilde{R}(x)$ to write it as

$$\tilde{R}(x) = \frac{\sum_{j \in N} r_j \, \tilde{v}_j \, x_j}{1 + \sum_{j \in N} \tilde{w}_j + \sum_{j \in N} (\tilde{v}_j - \tilde{w}_j) \, x_j} = \frac{\displaystyle\sum_{j \in N} \frac{r_j \, \tilde{v}_j}{\tilde{v}_j - \tilde{w}_j} \, \frac{\tilde{v}_j - \tilde{w}_j}{1 + \sum_{k \in N} \tilde{w}_k} \, x_j}{\displaystyle 1 + \sum_{j \in N} \frac{\tilde{v}_j - \tilde{w}_j}{1 + \sum_{k \in N} \tilde{w}_k} \, x_j}, \tag{5}$$

where the second equality above is obtained by dividing both the denominator and the numerator of the first fraction by $1 + \sum_{j \in N} \tilde{w}_j$. Comparing the expected revenue expressions in (1) and (5), we observe that these two expressions have precisely the same form as long as we identify $r_j \, \tilde{v}_j / (\tilde{v}_j - \tilde{w}_j)$ and $(\tilde{v}_j - \tilde{w}_j)/(1 + \sum_{k \in N} \tilde{w}_k)$ in (5), respectively, with r_j and v_j in (1). Thus, to solve the assortment problem $\max_{x \in \mathcal{F}} \tilde{R}(x)$, we can simply replace r_j and v_j in (3), respectively, with $r_j \, \tilde{v}_j / (\tilde{v}_j - \tilde{w}_j)$ and $(\tilde{v}_j - \tilde{w}_j)/(1 + \sum_{k \in N} \tilde{w}_k)$ and solve the resulting linear program. So, our observations in this section establish that we can extend Theorem 1 to cover the case where customers choose according to the general attraction model and we can formulate the assortment problem $\max_{x \in \mathcal{F}} \tilde{R}(x)$ as an equivalent linear program.

All of the applications in Section 2 make use of Theorem 1. So, we can incorporate cardinality constraints, display location effects and product precedence constraints when solving assortment problems under the general attraction model and solve pricing problems with a finite price menu and incorporate quality consistency constraints when making pricing decisions.

3.2 Alternative Expected Revenue Functions

In this section, we consider a more general version of problem (2), where the objective function is a ratio of the form $R(x) = f(x)/g(x)$ for $x \in \{0,1\}^{|N|}$. In problem (2), both $f(\cdot)$ and $g(\cdot)$ are linear functions, but we entertain the possibility of more general forms throughout this section. Thus, letting A be a totally unimodular matrix and $\mathcal{F} \subset \{0,1\}^{|N|}$ be the feasible set of product offer decisions as defined in Section 1, we consider problems of the form

$$z^* = \max_{x \in \mathcal{F}} \frac{f(x)}{g(x)}, \tag{6}$$

where, for compatibility with problem (2), we assume that $f(x)$ and $g(x)$ take positive values for all $x \in \mathcal{F}$. We observe that the optimal objective value of the problem above can be obtained by solving $\min\{t : t \geq f(x)/g(x) \; \forall \, x \in \mathcal{F}\}$, which is, in turn, equivalent to

$$\begin{aligned} \min \quad & t \\ \text{st} \quad & f(x) - t\, g(x) \leq 0 \quad \forall \, x \in \mathcal{F}, \end{aligned} \tag{7}$$

where the only decision variable is t. One strategy to solve problem (7) is to use iterative constraint generation; see Ruszczynski (2006). In particular, at any iteration k of constraint generation, we solve a master problem, which has the same form as problem (7), but includes only a subset $\mathcal{F}^k$ of the constraints. Solving the master problem $\min\{t : f(x) - t\, g(x) \leq 0 \; \forall \, x \in \mathcal{F}^k\}$ yields the solution t^k. To see if the solution t^k violates any of the constraints in problem (7), we can solve the constraint generation subproblem $\max_{x \in \mathcal{F}} f(x) - t^k\, g(x)$ to get the solution x^k. If $f(x^k) - t^k\, g(x^k) > 0$, then the constraint $f(x^k) - t\, g(x^k) \leq 0$ in problem (7) is violated by the solution t^k. We add this constraint to the master problem to obtain the subset of constraints $\mathcal{F}^{k+1} = \mathcal{F}^k \cup \{x^k\}$ and resolve the master problem at the next iteration $k+1$. If, however, $f(x^k) - t^k\, g(x^k) \leq 0$, then none of the constraints in problem (7) is violated by the solution t^k, in which case, t^k is the optimal solution to problem (7) and we can stop. Since problems (6) and (7) are equivalent to each other, this procedure yields the optimal solution to problem (6). So, the discussion in this paragraph shows that we can solve problem (6) efficiently as long as we can solve the constraint generation subproblem efficiently.

Interestingly, if $f(\cdot)$ and $g(\cdot)$ are linear functions, then we can carry out constraint generation implicitly, without solving a separate constraint generation subproblem. This gives us a direct approach to solve problem (6). To see this, we define $\pi(t) = \max_{x \in \mathcal{F}} f(x) - t\, g(x)$. Assuming that $f(\cdot)$ and $g(\cdot)$ are linear with $f(x) = f_0 + \sum_{j \in N} f_j\, x_j$ and $g(x) = g_0 + \sum_{j \in N} g_j\, x_j$, we have

$$\begin{aligned} \pi(t) = \max \quad & \sum_{j \in N} (f_j - t\, g_j)\, x_j + f_0 - t\, g_0 \\ \text{st} \quad & \sum_{j \in N} a_{ij}\, x_j \leq b_i \qquad \forall \, i \in M \\ & 0 \leq x_j \leq 1 \qquad \forall \, j \in N, \end{aligned} \tag{8}$$

where we drop the integrality requirements on the decision variables x since the constraint matrix is totally unimodular. When the constraint matrix is not totally unimodular, the optimal objective value of the problem above only provides an upper bound on $\pi(t)$. For a fixed value of t, the problem above is a bounded and feasible linear program. Therefore, we can compute $\pi(t)$ by solving the dual of this problem. Associating the dual variables $\{\mu_i : i \in M\}$ and $\{\eta_j : j \in N\}$ with the two sets of constraints above, we write the dual of the problem above as

$$
\begin{aligned}
\pi(t) = \min \quad & \sum_{i \in M} b_i\, \mu_i + \sum_{j \in N} \eta_j + f_0 - t\, g_0 && && (9)\\
\text{st} \quad & \sum_{i \in M} a_{ij}\, \mu_i + \eta_j \geq f_j - t\, g_j && \forall\, j \in N\\
& \mu_i,\ \eta_j \geq 0 && \forall\, i \in M,\ j \in N.
\end{aligned}
$$

By the definition of $\pi(\cdot)$, problem (7) is equivalent to the problem $\min\{t : \pi(t) \leq 0\}$. To write this last problem more explicitly, we can use the definition of $\pi(t)$ given in problem (9) to get

$$
\begin{aligned}
\min \quad & t && && (10)\\
\text{st} \quad & \sum_{i \in M} b_i\, \mu_i + \sum_{j \in N} \eta_j + f_0 - t\, g_0 \leq 0\\
& \sum_{i \in M} a_{ij}\, \mu_i + \eta_j \geq f_j - t\, g_j && \forall\, j \in N\\
& \mu_i,\ \eta_j \geq 0,\ t \text{ is free} && \forall\, i \in M,\ j \in N,
\end{aligned}
$$

which is a linear program. The first constraint ensures that $\pi(t) \leq 0$, whereas the second set of constraints ensure that for a given value of t, the value of $\pi(t)$ is computed correctly through the decision variables $\{\mu_i : i \in M\}$ and $\{\eta_j : j \in N\}$. By the discussion in this paragraph, problem (10) is equivalent to problem (7), which is, in turn, equivalent to problem (6). Therefore, if $f(\cdot)$ and $g(\cdot)$ are linear functions and A is a totally unimodular matrix, then we can solve the linear program given in problem (10) to directly obtain the optimal objective value of problem (6). To reconcile the discussion so far in this section with the development in Section 1, if $f(\cdot)$ and $g(\cdot)$ are linear functions of the form $f(x) = \sum_{j \in N} r_j\, v_j\, x_j$ and $g(x) = 1 + \sum_{j \in N} v_j\, x_j$, then the optimal objective value of problem (10) corresponds to the optimal objective value of problem (2). Also, if we use $f_0 = 0$, $f_j = r_j\, v_j$, $g_0 = 1$ and $g_j = v_j$ in problem (10) and write the dual of problem (10), then we immediately obtain problem (3). Thus, the discussion in this section gives an alternative, more constructive approach to show the equivalence between problems (2) and (3).

Considering more general cases, when A is not a totally unimodular matrix but $f(\cdot)$ and $g(\cdot)$ are linear functions, the optimal objective value of problem (8) provides only an upper bound on $\pi(t)$. Using this observation, it is not too difficult to show that the optimal objective value of problem (10) is an upper bound on the optimal objective value of problem (6) when A is not necessarily totally unimodular. Such an upper bound on the optimal objective value of problem (6) may be useful when we use a heuristic to find a potentially good solution to problem (6) and we want to assess the optimality gap of this solution. This observation also implies that the optimal objective value of problem (3) is an upper bound on the optimal objective value of problem (2) when A is not necessarily totally unimodular.

When $f(\cdot)$ and $g(\cdot)$ are not linear, we can potentially obtain approximate solutions to problem (6) by solving problem (7) through iterative constraint generation. For example, if $f(\cdot)$ is concave and $g(\cdot)$ is convex, then to check whether the solution t^k to the master problem at iteration k violates any of the constraints in problem (7), we need to solve $\max_{x\in\mathcal{F}} f(x) - t^k\, g(x)$. Assuming without loss of generality that $t^k \geq 0$, the objective function of this problem is concave. Thus, we can at least solve the continuous relaxation of this problem in a tractable fashion to obtain an upper bound. Using such upper bounds in the constraint generation strategy ultimately provides an upper bound on the optimal objective value of problem (7). So, we can obtain upper bounds on the optimal objective value of problem (6) when $f(\cdot)$ is concave and $g(\cdot)$ is convex.

When $f(\cdot)$ is convex and $g(\cdot)$ is concave, the continuous relaxation of the problem $\max_{x\in\mathcal{F}} f(x) - t^k\, g(x)$ maximizes a convex function over a convex region. When $f(\cdot)$ and $g(\cdot)$ are both concave, the continuous relaxation of the problem $\max_{x\in\mathcal{F}} f(x) - t^k\, g(x)$ can still be cast as maximizing a convex function over a convex region. To see this, we use the additional decision variable z to write the continuous relaxation of the last problem as

$$\begin{aligned} \max \quad & z - t^k\, g(x) && (11)\\ \text{st} \quad & \sum_{j\in N} a_{ij}\, x_j \leq b_i && \forall\, i \in M\\ & f(x) \geq z \\ & 0 \leq x_j \leq 1,\ \ z \text{ is free} && \forall\, j \in N. \end{aligned}$$

Since $f(\cdot)$ is concave, the feasible region of the problem above is convex, whereas since $g(\cdot)$ is concave, the objective function is convex. At the optimal solution, the second constraint is satisfied at equality. When $f(\cdot)$ and $g(\cdot)$ are both convex, we can use a similar trick to cast the continuous relaxation of the problem $\max_{x\in\mathcal{F}} f(x) - t^k\, g(x)$ as maximizing a convex function over a convex region, but we use the objective function $f(x) - t^k\, z$ and the constraint $g(x) \leq z$.

Therefore, if $f(\cdot)$ is concave and $g(\cdot)$ is convex, then we can obtain an upper bound on the optimal objective value of the problem $\max_{x\in\mathcal{F}} f(x) - t^k\, g(x)$, which, in turn, yields an upper bound on the optimal objective value of problem (6) through constraint generation. The remaining cases are difficult and the continuous relaxation of the problem $\max_{x\in\mathcal{F}} f(x) - t^k\, g(x)$ can be cast as maximizing a convex function over a convex region. Algorithms have been proposed for maximizing a convex function over a convex region, which may come useful when generating approximate solutions to problem (6); see Hoffman (1981) and Vu Thieu et al. (1983).

4 Conclusions

In this paper, we showed how to use linear programs to formulate and solve certain assortment optimization problems with totally unimodular constraint structures. We demonstrated a number of practically useful assortment and pricing problems that can be formulated as an assortment optimization problem with a totally unimodular constraint matrix. There are two interesting avenues to pursue for future research. First, it is worthwhile to investigate other assortment and pricing problems that can be cast within the totally unimodular constraint framework studied in this paper. Second, we showed that all of our results hold under the general attraction model, which is an extension of MNL. It would be useful to come up with other choice models for which we can reduce the corresponding assortment optimization problem to a linear program. One immediate candidate is the nested logit model, but due to the so called dissimilarity parameters in this choice model, if we follow the strategy used in this paper, then we end up with an intractable nonconvex program, rather than a linear program.

References

Ben-Akiva, M. and Lerman, S. (1994), *Discrete Choice Analysis: Theory and Application to Travel Demand*, MIT Press in Transportation Studies, Cambridge, MA.

Bront, J. J. M., Mendez-Diaz, I. and Vulcano, G. (2009), 'A column generation algorithm for choice-based network revenue management', *Operations Research* **57**(3), 769–784.

Chen, K. D. and Hausman, W. H. (2000), 'Technical note: Mathematical properties of the optimal product line selection problem using choice-based conjoint analysis', *Management Science* **46**(2), 327–332.

Davis, J., Gallego, G. and Topaloglu, H. (2011), Assortment optimization under variants of the nested logit model, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

Dong, L., Kouvelis, P. and Tian, Z. (2009), 'Dynamic pricing and inventory control of substitute products', *Manufacturing & Service Operations Management* **11**(2), 317–339.

Gallego, G., Iyengar, G., Phillips, R. and Dubey, A. (2004), Managing flexible products on a network, Computational Optimization Research Center Technical Report TR-2004-01, Columbia University.

Gallego, G., Ratliff, R. and Shebalov, S. (2011), A general attraction model and an efficient formulation for the network revenue management problem, Technical report, Columbia University, New York, NY.

Gallego, G. and Stefanescu, C. (2009), Upgrades, upsells and pricing in revenue management, Technical report, Columbia University.

Gallego, G. and Topaloglu, H. (2012), Constrained assortment optimization for the nested logit model, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

Gallego, G. and Wang, R. (2011), Multi-product price optimization and competition under the nested attraction model, Technical report, Columbia University.

Hanson, W. and Martin, K. (1996), 'Optimizing multinomial logit profit functions', *Management Science* **42**(7), 992–1003.

Hausman, J. and McFadden, D. (1984), 'Specification tests for the multinomial logit model', *Econometrica* **52**(5), 1219–1240.

Heller, I. (1957), 'On linear systems with integral valued solutions', *Pacific Journal of Mathematics* **7**(3), 1351–1364.

Heller, I. and Hoffman, A. J. (1962), 'On unimodular matrices', *Pacific Journal of Mathematics* **12**(4), 1321–1327.

Hoffman, K. L. (1981), 'A method for globally minimizing concave functions over convex sets', *mathematical Programming* **20**, 22–32.

Kunnumkal, S. and Topaloglu, H. (2008), 'A refined deterministic linear program for the network revenue management problem with customer choice behavior', *Naval Research Logistics Quarterly* **55**(6), 563–580.

Li, H. and Huh, W. T. (2011), 'Pricing multiple products with the multinomial logit and nested models: Concavity and implications', *Manufacturing & Service Operations Management* **13**(4), 549–563.

Liu, Q. and van Ryzin, G. (2008), 'On the choice-based linear programming model for network revenue management', *Manufacturing & Service Operations Management* **10**(2), 288–310.

Luce, R. D. (1959), *Individual Choice Behavior: A Theoretical Analysis*, Wiley, New York, NY.

McFadden, D. (1974), Conditional logit analysis of qualitative choice behavior, *in* P. Zarembka, ed., '*Frontiers in Economics*', Academic Press, pp. 105–142.

Meissner, J. and Strauss, A. K. (2012), 'Network revenue management with inventory-sensitive bid prices and customer choice', *European Journal of Operational Research* **216**, 459–468.

Meissner, J., Strauss, A. and Talluri, K. (2012), 'An enhanced concave program relaxation for choice network revenue management', *Production and Operations Management* (to appear).

Mendez-Diaz, I., Bront, J. J. M., Vulcano, G. and Zabala, P. (2010), 'A branch-and-cut algorithm for the latent-class logit assortment problem', *Discrete Applied Mathematics* **36**, 383–390.

Nemhauser, G. and Wolsey, L. (1988), *Integer and Combinatorial Optimization*, John Wiley & Sons, Inc., Chichester.

Rusmevchientong, P., Van Roy, B. and Glynn, P. (2006), 'A nonparametric approach to multiproduct pricing', *Operations Research* **54**(1), 82–98.

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

Rusmevichientong, P., Shen, Z.-J. M. and Shmoys, D. B. (2010), 'Dynamic assortment optimization with a multinomial logit choice model and capacity constraint', *Operations Research* **58**(6), 1666–1680.

Rusmevichientong, P., Shmoys, D. B. and Topaloglu, H. (2010), Assortment optimization with mixtures of logits, Technical report, Cornell University, School of Operations Research and Information Engineering.
Available at `http://legacy.orie.cornell.edu/~huseyin/publications/publications.html`.

Ruszczynski, A. (2006), *Nonlinear Optimization*, Princeton University Press, Princeton, New Jersey.

Song, J.-S. and Xue, Z. (2007), Demand management and inventory control for substitutable products, Technical report, Duke University, Durham, NC.

Talluri, K. (2011), A randomized concave programming method for choice network revenue management, Technical report, Universitat Pompeu Fabra, Barcelona, Spain.

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

Train, K. (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press, Cambridge, UK.

Vu Thieu, T., Tam, B. T. and Ban, V. T. (1983), ‘An outer approximation method for globally minimizing a concave function over a compact convex set’, *Acta Mathematica Vietnamica* **8**(1), 21–40.

Wang, R. (2012), ‘Capacitated assortment and price optimization under the multinomial logit model’, *Operations Research Letters* **40**, 492–497.

Zhang, D. and Adelman, D. (2009), ‘An approximate dynamic programming approach to network revenue management with customer choice’, *Transportation Science* **42**(3), 381–394.

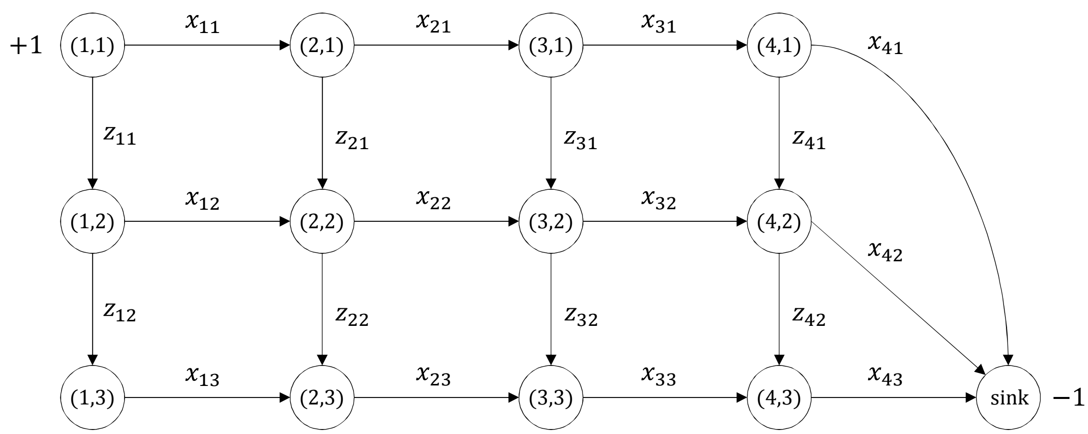

Figure 1: Shortest path problem in the quality consistent pricing setting.