OPERATIONS RESEARCH
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0030-364X | EISSN 1526-5463 | 00 | 0000 | 0001

INFORMS
DOI 10.1287/xxxx.0000.0000

On the Tightness of an LP Relaxation for Rational Optimization and its Applications

---

We consider the problem of optimizing a linear rational function subject to totally unimodular (TU) constraints over $\{0,1\}$ variables. We work with the LP relaxation of this problem and prove that its extreme points are integral thereby showing that the LP relaxation is "tight." We also consider a more general variant of this problem by allowing additional (not totally unimodular) constraints. By extending structural insights about extreme points we present a polynomial time approximation scheme (PTAS) for the general problem. Examples of such settings in the context of assortment optimization with multinomial logit choice (MNL) model are also discussed along with numerical simulations.

1. Introduction

We consider the problem of optimizing a linear rational function over $\{0,1\}$ variables subject to totally unimodular (TU) constraints, i.e.

$$\begin{aligned} \text{maximize} \quad & \frac{a_0 + \sum_{i=1}^n a_i x_i}{c_0 + \sum_{j=1}^n c_j x_j} \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\ & \mathbf{x} \in \{0,1\}^n, \end{aligned} \tag{1}$$

where $\mathbf{A}$ is a TU matrix, $\mathbf{b} \in Z^m$, $c_i \ge 0$ for all i. Many combinatorial optimization problems like minimum mean cycle, minimum ratio shortest path and assortment optimization over an MNL choice model involve optimizing a rational objective. Davis et al. (2014) consider a special case of our problem in the context of assortment optimization. In particular, they consider an MNL choice model with TU constraints and provide an algorithm to obtain the optimal solution. Megiddo

(1979) consider a similar problem of optimizing a rational objective over specific combinatorial sets. In that work, they provide an algorithm to compute the optimal solution of the following problem

$$\begin{aligned} \text{maximize} \quad & \frac{a_0 + \sum_{i=1}^{n} a_i x_i}{c_0 + \sum_{j=1}^{n} c_j x_j} \\ \text{subject to} \quad & \mathbf{x} \in D, \end{aligned} \tag{2}$$

where D is a combinatorial set, such that the following problem can be solved efficiently using a combinatorial algorithm (only involving addition, subtraction or comparison),

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^{n} d_i x_i \\ \text{subject to} \quad & \mathbf{x} \in D. \end{aligned} \tag{3}$$

Hashizume et al. (1987), Correa et al. (2010) extend the work of Megiddo (1979) to provide approximation algorithms for problems similar to (2) assuming there are efficient algorithms to compute approximate solutions for problem (3). Mittal and Schulz (2013) provides a fully polynomial time approximation scheme (FPTAS) for optimizing a rational objective over a polytope and their technique can be extended to the case of a combinatorial set, if the extreme points of the polytope are feasible in the combinatorial set.

Our Contributions. We reformulate the rational optimization problem (1) as an integer program and show that the LP relaxation of this integer program is tight. We would like to note that, Davis et al. (2014) takes a similar approach, but their results do not discuss the structure of optimal solutions of the reformulation. Instead they obtain the optimal solution by solving another linear program using the optimal value obtained from the relaxation. In contrast, we provide structural results on the LP reformulation to establish that the relaxation is tight and the optimal solution(s) of the relaxation solve the rational optimization problem (1). We work with this structural result to obtain a polynomial time approximate scheme (PTAS) for the rational optimization problem (1) in the presence of an additional constraint which will be motivated in what follows. We extend our PTAS approach to an important class of applications of recent interest, a problem referred to as "joint assortment and display optimization with capacity constraint". To the best of our knowledge, this is the first approximation algorithm for this application domain for which it is known that obtaining an exact solution is "hard".

Notation. We use the following notations in this paper. We use bold font to denote all vectors and matrices. For any matrix $\mathbf{Y} \in R^{m\times n}$ and index set $T \subset \{1, 2, \ldots, m\}$, $\mathbf{Y}(T)$ denotes the submatrix corresponding to rows T. The identity matrix is denoted as $\mathbf{I}$ and $\mathbf{e}$ denotes the vector of all ones of appropriate dimension. All vectors are column vectors. For any $n \in \mathcal{N}$, $[n]$ denotes the set $\{1, 2, \ldots, n\}$.

Outline. The rest of the paper is organized as follows. In Section 2, we prove tightness of the LP relaxation. In Section 3, we present applications of the rational optimization problem (1) including "cardinality constrained assortment optimization" and "joint assortment and display optimization with cardinality constraints." In Section 4, we present extensions to more general constraint sets and applications thereof. Section 5 contains some numerical illustrations.

2. Rational Optimization: LP relaxation

We consider the rational optimization problem (1), Substituting

$$p_0 = \frac{1}{c_0 + \sum_{j=1}^n c_j x_j}, \quad p_i = x_i p_0,$$

we have the following equivalent reformulation of the rational optimization problem (1) with linear objective function.

$$\begin{aligned} \underset{(\mathbf{p}, p_0)}{\text{maximize}} \quad & \sum_{i=0}^n a_i p_i \\ \text{subject to} \quad & \mathbf{A}\mathbf{p} \le p_0 \mathbf{b} \\ & \sum_{i=0}^n c_i p_i = 1 \\ & p_i \in \{0, p_0\} \ \forall \ i \in \{1, 2, \cdots, n\} \\ & p_0 \ge 0. \end{aligned} \tag{4}$$

Note that in the above the reformulation is not a mixed integer program but can be easily reformulated as the following mixed integer program, by rewriting the constraints $p_i \in \{0, p_0\}$ for all i as

$$\begin{aligned} & p_i \le x_i \ \forall \ i \in \{1, 2, \cdots, n\} \\ & p_i + (1 - x_i) \ge p_0 \ \forall \ i \in \{1, 2, \cdots, n\} \\ & x_i \in \{0, 1\} \ \forall \ i \in \{1, 2, \cdots, n\}. \end{aligned} \tag{5}$$

2.1. Tightness of the LP relaxation

We consider the following LP relaxation for (4).

$$\begin{aligned} z_{\mathsf{LP}} = \max_{(\mathbf{p}, p_0)} & \sum_{i=1}^n a_i p_i \\ & \mathbf{A}\mathbf{p} \le p_0 \mathbf{b} \\ & \sum_{j=0}^n c_j p_j = 1 \\ & 0 \le p_i \le p_0, \ \forall i = 1, \ldots, n, \\ & p_0 \ge 0. \end{aligned} \tag{6}$$

where we relax the constraints $p_i \in \{0, p_0\}$ to $0 \le p_i \le p_0$ for all $i = 1, \ldots, n$. Let $\mathcal{P}$ be the polytope defined by the constraints in the above LP formulation, i.e.,

$$\mathcal{P} = \left\{ (\mathbf{p}, p_0) \in \mathbb{R}_+^n \times \mathbb{R}_+ \;\middle|\; \mathbf{A}\mathbf{p} \le p_0 \mathbf{b},\ \mathbf{c}^T \mathbf{p} + c_0 p_0 = 1,\ 0 \le p_i \le p_0, \forall i \right\}. \tag{7}$$

We show that all extreme points of $\mathcal{P}$ are "integral," i.e., for any extreme point $(\mathbf{p}, p_0) \in \mathcal{P}$, $p_i \in \{0, p_0\}$ for all $i = 1, \ldots, n$. In particular, we have the following theorem.

THEOREM 1. *Every extreme point $(\mathbf{p}, p_0)$ of the polytope $\mathcal{P}$ is such that $p_i \in \{0, p_0\}$ for all $i = 1, \ldots, n$.*

We will prove Theorem 1 by establishing a correspondence between extreme points of $\mathcal{P}$ and $\mathcal{Q}$, where

$$\mathcal{Q} = \{ \mathbf{x} \mid \mathbf{A}\mathbf{x} \le \mathbf{b},\ 0 \le x_i \le 1 \ \text{ for all } i = 1, 2, \ldots, n \},$$

is the polytope corresponding to relaxed constraints of the rational optimization problem (1).

LEMMA 1. *If $(\mathbf{p}, p_0)$ is an extreme point of $\mathcal{P}$, then $\mathbf{x} = \frac{\mathbf{p}}{p_0}$ is an extreme point of $\mathcal{Q}$. Conversely, if $\mathbf{x}$ is an extreme point of $\mathcal{Q}$, then $(\mathbf{p}, p_0)$ where*

$$p_0 = \frac{1}{(c_0 + \mathbf{c}'\mathbf{x})},\ \mathbf{p} = p_0 \mathbf{x}$$

is an extreme point of $\mathcal{P}$.

PROOF. Note that for every extreme point of $\mathcal{P}$, there must be $n+1$ linearly independent and active constraints. Let $(\mathbf{p}, p_0)$ be an extreme point of $\mathcal{P}$ and define

$$\begin{aligned} S_0 &= \{ i \mid p_i = 0 \},\ S_1 = \{ i \mid p_i = p_0 \}, \\ T &= \{ i \mid \sum_{j=1}^{n} a_{ij} p_j = b_i p_0 \} \\ k &= |S_0| + |S_1| + |T|. \end{aligned} \tag{8}$$

We claim that $k \ge n$. This follows by observing that we have $|S_0| + |S_1|$ linearly independent and active constraints from the constraint set $S_0 \cup S_1$, $|T|$ active constraints from the constraint set T and one active constraint from the constraint $\sum_{i=0}^{n} c_i p_i = 1$. Hence the total number of linearly independent and active constraints at $(\mathbf{p}, p_0)$ is at most $k+1$.

Without loss of generality we can assume that $k = n$; since $k > n$ implies that $|S_0| + |S_1| + |T| + 1 > n + 1$, making some constraints in T redundant.

Define

$$\mathbf{B}_p = \begin{bmatrix} \mathbf{A}(T) & -\mathbf{b}(T) \\ \mathbf{I}(S_0) & \mathbf{0} \\ \mathbf{I}(S_1) & -\mathbf{e} \\ \mathbf{c}' & c_0 \end{bmatrix},\quad \mathbf{B}_x = \begin{bmatrix} \mathbf{A}(T) \\ \mathbf{I}(S_0) \\ \mathbf{I}(S_1) \end{bmatrix},\quad \mathbf{b}_x = \begin{bmatrix} \mathbf{b}(T) \\ \mathbf{0} \\ \mathbf{e} \end{bmatrix}, \tag{9}$$

Note that $\mathbf{B}_p$ is the basis matrix corresponding to the extreme point $(\mathbf{p}, p_0)$. Hence, $\mathbf{B}_p$ is full rank. For the sake of contradiction, assume that $\mathbf{B}_x$ is not full rank. There there exists $\boldsymbol{\lambda} \in R^n, \boldsymbol{\lambda} \neq \mathbf{0}$ such that $\boldsymbol{\lambda}'\mathbf{B}_x = \mathbf{0}$, then we have

$$\begin{bmatrix}\boldsymbol{\lambda}' & 0\end{bmatrix} \mathbf{B}_p = \begin{bmatrix}\boldsymbol{\lambda}'\mathbf{B}_x & -\boldsymbol{\lambda}'\mathbf{b}_x\end{bmatrix} = \begin{bmatrix}\mathbf{0} & -\boldsymbol{\lambda}'\mathbf{b}_x\end{bmatrix},$$

which implies

$$\begin{bmatrix}\boldsymbol{\lambda}' & 0\end{bmatrix} \mathbf{B}_p \begin{bmatrix}\mathbf{p} \\ p_0\end{bmatrix} = -p_0\boldsymbol{\lambda}'\mathbf{b}_x,$$

Since $\mathbf{B}_p$ is a full rank, we have $\boldsymbol{\lambda}'\mathbf{b}_x \neq 0$ and $p_0 \neq 0$, contradicting that,

$$\mathbf{B}_p \begin{bmatrix}\mathbf{p} \\ p_0\end{bmatrix} = \begin{bmatrix}\mathbf{0} \\ 1\end{bmatrix}.$$

Hence, $\mathbf{B}_x$ is a full rank.

Clearly $\mathbf{x} = \mathbf{p}/p_0$ is a feasible point in $\mathcal{Q}$ and solves the system of linear equations $\mathbf{B}_x\mathbf{x} = \mathbf{b}_x$. Hence $\mathbf{x}$ is the basic feasible solution corresponding to the basis matrix $\mathbf{B}_x$.

Now we prove the converse. Consider $\mathbf{x}$, any extreme point of $\mathcal{Q}$. Let

$$p_0 = \frac{1}{c_0 + \mathbf{c}'\mathbf{x}}, \;\; \mathbf{p} = p_0\mathbf{x}.$$

Clearly $(\mathbf{p}, p_0) \in \mathcal{P}$. We define the quantities S_0, S_1, T, k as in (8) and $\mathbf{B}_p, \mathbf{B}_x, \mathbf{b}_x$ as in (9). Using similar arguments, claim that without loss of generality we can take $k = n$. Since $\mathbf{x}$ is a basic feasible solution corresponding to the basis $\mathbf{B}_x$, $\mathbf{B}_x$ is full rank.

For the sake of contradiction, suppose $\mathbf{B}_p$ is not full rank. Then there exists $\boldsymbol{\lambda} \in R^{n+1}, \boldsymbol{\lambda} \neq \mathbf{0}$ such that $\boldsymbol{\lambda}'\mathbf{B}_p = \mathbf{0}$. Therefore,

$$\boldsymbol{\lambda}'\mathbf{B}_p \begin{bmatrix}\mathbf{p} \\ p_0\end{bmatrix} = 0, \text{ which implies } (\boldsymbol{\lambda}([n]))'(\mathbf{B}_x\mathbf{p} + p_0\mathbf{b}_x) + \lambda_{n+1}(\mathbf{c}'\mathbf{p} + c_0p_0) = 0.$$

Since $\mathbf{B}_x\mathbf{x} = \mathbf{b}_x$, we have $\mathbf{B}_x\mathbf{p} + p_0\mathbf{b}_x = \mathbf{0}$ and $\lambda_{n+1} = 0$. Note that,

$$\boldsymbol{\lambda}'\mathbf{B}_p = \begin{bmatrix}\boldsymbol{\lambda}([n])'\mathbf{B}_x + \lambda_{n+1}\mathbf{c}' & \boldsymbol{\lambda}([n])'\mathbf{b}_x + \lambda_{n+1}c_0\end{bmatrix} \text{ and } \boldsymbol{\lambda}'\mathbf{B}_p = \mathbf{0}$$

Therefore $\boldsymbol{\lambda}([n])'\mathbf{B}_x = \mathbf{0}$, contradicting the fact that $\mathbf{B}_x$ is full rank. Hence, $\mathbf{B}_p$ is a full rank matrix and $(\mathbf{p}, p_0)$ is the basic feasible solution corresponding to the basis matrix $\mathbf{B}_p$. This completes the proof. □

Theorem 1 follows from Lemma 1 and the fact that any extreme point $\mathbf{x}$ of $\mathcal{Q}$ is integral, i.e. $x_i \in \{0, 1\}$.

We emphasize that although Davis et al. (2014) takes a similar approach of reformulating the rational optimization problem (1) as LP relaxation (6), their result does not focus on the optimal solutions of the relaxation. Instead they obtain the optimal solution by solving another linear program using the optimal value of LP relaxation (6). Our Theorem 1, apart from establishing that the relaxation LP relaxation (6) is tight, also proves that the optimal solution of LP relaxation (6) is the same as the MIP reformulation (4) and hence it suffices to solve the relaxation.

3. Applications: Assortment Optimization over MNL

In this section, we present specific cases of assortment optimization problems referred to as "cardinality constrained assortment optimization" and "joint assortment and display optimization with cardinality constraints," as an application of the rational optimization problem (1). We employ Theorem 1 to obtain its optimal solutions. Davis et al. (2014) presents many other applications related to assortment optimization and pricing problems under MNL that can be formulated within the framework of the rational optimization problem (1). Later in Section 4, we further consider a more general version of "joint assortment and display optimization with cardinality constraints," where apart from TU constraints, we allow an additional capacity constraint. Before describing the optimization problem, we first give some brief background on assortment optimization.

In retail settings, an assortment of products selected by a retailer for display has significant impact on revenues. Assuming a specific choice model for substitution among products, the assortment optimization problem attempts to find the optimal subset of products satisfying various constraints (budgetary, space); Kok et al. (2003) provides a detailed review of assortment optimization problems. The Multinomial Logit (MNL) model, owing to its tractability, is a popular and well studied choice model for assortment selection problems. Talluri and van ryzin (2004), Rusmeivichientong et al. (2006), Desir and Goyal (2014) and Davis et al. (2014) consider variants of the assortment optimization problem under an MNL choice model. Assortment optimization under MNL choice framework is an important class of problems that involves optimizing a rational objective over $\{0,1\}$ variables subject to various constraints.

3.1. Assortment Optimization with Cardinality Constraint

We consider an assortment optimization problem, where the retailer needs to select a subset of products to offer to customers who make their selection according to the MNL choice model. The objective is to compute an optimal assortment to maximize the expected revenue such that the total number of products selected does not exceed some upper bound.

We now formulate this problem as a combinatorial optimization of a rational objective over TU constraints. Let n be the total number of products, the product assortment is represented by the vector $\mathbf{x} \in \{0,1\}^n$, where $x_i = 1$ implies product i is selected and $x_i = 0$ implies product i is discarded. Typically, under the MNL framework, product i is characterized by its attractiveness parameter (mean utility) v_i. The probability of buying product i, for a given assortment $\mathbf{x}$ this is given by,

$$p_i = \frac{v_i x_i}{v_0 + \sum_{i=1}^{n} v_i x_i}$$

If r_i is the marginal profit for product i, then the expected revenue when the product assortment is $\mathbf{x}$ is given by:

$$R(\mathbf{x}) = \frac{\sum_{i=1}^{n} r_i v_i x_i}{v_0 + \sum_{i=1}^{n} v_i x_i}.$$

Therefore, the optimization problem of maximizing the revenue subject to cardinality constraint is,

$$\begin{aligned} \underset{\mathbf{x}\in\{0,1\}^n}{\text{maximize}} \quad & R(\mathbf{x}) = \frac{\sum_{i=1}^{n} r_i v_i x_i}{v_0 + \sum_{i=1}^{n} v_i x_i} \\ \text{subject to} \quad & \sum_{i=1}^{n} x_i \leq K \\ & x_i \in \{0,1\} \ \forall\ i. \end{aligned} \tag{10}$$

3.2. Joint Assortment and Display Optimization with Cardinality Constraints

Here we consider a joint assortment and display optimization problem, where the retailer needs to select the subset of products to offer and also decide on the display segment. This problem arises in retailing and online advertising where the display slot of the product/ad affects the choice probability. In particular, we consider a model with m display segments and each segment has an upper bound on the number of products that can be displayed. The customers choose in accordance with an MNL model, where the purchasing probability of each offered product also depends on its display segment. The objective is to compute an optimal assortment together with the optimal display segment for each offered product such that the cardinality constraints for each segment are satisfied and the expected revenue is maximized.

We now formulate this as a combinatorial optimization of a rational objective over TU constraints. Let n be the total number of products and m be the number of display segments. Display segment j can accommodate at most K_j products for each j (hereafter refered to as "cardinality" constraints). We assume that every product can only be displayed in one of the available display segments. Product offer decisions are denoted by $x_{ij} \in \{0,1\}$, which will be the decision variables in our optimization problem:

$$x_{ij} = \begin{cases} 1 & \text{if product } i \text{ is displayed in slot } j \\ 0 & \text{otherwise.} \end{cases}$$

The product assortment and their display slots are represented by $n \times m$ matrix $\mathbf{X}$. Here $\sum_{j=1}^{m} x_{ij} = 0$ implies that product i is not displayed in any segment.

Apart from the usual attractiveness parameter (mean utility) v_i, for each product i we introduce an additional display parameter β_j for all j and assume that the overall attractiveness parameter for a product i displayed in slot j is $\beta_j v_i$. Thus, if r_i is the marginal profit for product i, then the expected revenue when the product assortment and display arrangement is $\mathbf{X}$ is given by:

$$R(\mathbf{X}) = \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} r_i v_i \beta_j x_{ij}}{v_0 + \sum_{i=1}^{n} \sum_{j=1}^{m} v_i \beta_j x_{ij}}.$$

Let $\hat{v}_{ij} = v_i\beta_j$, the optimization problem of maximizing expected revenue can be formulated as:

$$
\begin{aligned}
\underset{\mathbf{X}\in\{0,1\}^{n\times m}}{\text{maximize}} \quad & R(\mathbf{X}) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{m} r_i \hat{v}_{ij} x_{ij}}{v_0 + \sum_{i=1}^{n}\sum_{j=1}^{m} \hat{v}_{ij} x_{ij}} \\
\text{subject to} \quad & \mathbf{C_i}: \quad \sum_{j=1}^{m} x_{ij} \le 1,\ i = 1,\dots,n \\
& \mathbf{C_j}: \quad \sum_{i=1}^{n} x_{ij} \le K_j,\ j = 1,\dots,m \\
& \qquad\quad x_{ij} \in \{0,1\},\ \ i = 1,\dots,n,\ j = 1,\dots,m.
\end{aligned}
\tag{11}
$$

Constraints $\{\mathbf{C_i}\}$ enforce that every product can be displayed only in one of the display segments, while constraints $\{\mathbf{C_j}\}$ enforce the "cardinality" constraints in each segment. The constraints in problem (11) are identical to the constraints in a transportation problem and hence are TU, making problem (11) a special case of the rational optimization problem (1).

4. Extension to More General Constraints

In this section, we consider a more general variant of the rational optimization problem (1), where constraints are not necessarily TU. In particular, we consider the following problem where we have a set of TU constraints and one additional constraint such that the overall constraints are not TU:

$$
\begin{aligned}
\text{maximize} \quad & \frac{a_0 + \sum_{i=1}^{n} a_i x_i}{c_0 + \sum_{j=1}^{n} c_j x_j} \\
\text{subject to} \quad & \mathbf{A}\mathbf{x} \le \mathbf{b} \\
& \boldsymbol{\alpha}^T \mathbf{x} \le \gamma \\
& \mathbf{x} \in \{0,1\}^n,
\end{aligned}
\tag{12}
$$

where $\mathbf{A}$ is a $\{0,1\}^{m\times n}$ TU matrix, $\mathbf{b} \in Z^m,\ c_i \ge 0$ and $\alpha_i \ge 0$ for all i. Let

$$
\begin{aligned}
&\mathcal{Q} = \{\mathbf{x} \mid A\mathbf{x} \le \mathbf{b},\ 0 \le x_i \le 1 \ \text{ for all } i = 1,2,\dots,n\} \\
&\hat{\mathcal{Q}} = \left\{\mathbf{x} \in \mathcal{Q} \;\middle|\; \boldsymbol{\alpha}^T\mathbf{x} \le \gamma\right\},
\end{aligned}
$$

be the polytopes corresponding to the relaxations of (1) and (12) respectively.

Similar to our approach in Section 2, we consider the following LP relaxation for (12),

$$
\begin{aligned}
\underset{(\mathbf{p},p_0)}{\text{maximize}} \quad & \sum_{i=0}^{n} a_i p_i \\
\text{subject to} \quad & (\mathbf{p}, p_0) \in \mathcal{P} \\
& \boldsymbol{\alpha}^T \mathbf{p} \le p_0 \gamma.
\end{aligned}
\tag{13}
$$

where $\mathcal{P}$ is as defined in (7), the polytope corresponding to the LP relaxation of the rational optimization problem (1) . We have,

$$
\hat{\mathcal{P}} = \left\{(\mathbf{p}, p_0) \in \mathcal{P} \;\middle|\; \boldsymbol{\alpha}^T\mathbf{p} \le p_0\gamma \ \right\},
$$

as the polytope corresponding to the LP relaxation of the problem (12).

Since constraints in (12) are not TU, the LP relaxation (13) may not be tight. In this section, we present a polynomial time approximation scheme (PTAS) for (12) under some assumptions on $\mathcal{Q}$. In other words, for a fixed ϵ, we compute an $(1-\epsilon)$-approximation for (12) in time $O(n^{1/\epsilon})$. Our PTAS is based on the following structure of extreme points of (13).

Observe that the polytope $\hat{\mathcal{Q}}$ (respectively $\hat{\mathcal{P}}$) is the intersection of the polytope $\mathcal{Q}$ (respectively $\mathcal{P}$) and the hyperplane $\boldsymbol{\alpha}^T \mathbf{x} \le \gamma$ (respectively $\boldsymbol{\alpha}^T \mathbf{p} \le p_0 \gamma$). Hence, any extreme point of $\hat{\mathcal{Q}}$ (respectively $\hat{\mathcal{P}}$) is either an extreme point of $\mathcal{Q}$ or a convex combination of two adjacent extreme points of $\mathcal{Q}$ (respectively $\mathcal{P}$). Therefore, if two adjacent extreme points of $\mathcal{Q}$ "differ" only in a small number of components, then the number of "fractional components" in any extreme point of $\hat{\mathcal{Q}}$ and $\hat{\mathcal{P}}$ is small. We obtain an approximate solution for (12) by ignoring a small number of "fractional components" from the optimal solution of (13). Specifically, for any two extreme points $\mathbf{x}_1, \mathbf{x}_2$ of $\mathcal{Q}$, define

$$\begin{aligned} d(\mathbf{x}_1, \mathbf{x}_2) &= |\{i \mid x_{1i} \neq x_{2i}\}| \\ d(\mathcal{Q}) &= \max\{d(\mathbf{x}_1, \mathbf{x}_2) \mid \mathbf{x}_1,\ \mathbf{x}_2 \text{ are adjacent extreme points of } \mathcal{Q}\}. \end{aligned}$$

Here $d(\mathcal{Q})$ denotes the maximum number of components by which the two adjacent extreme points of $\mathcal{Q}$ can differ. If $d(\mathcal{Q}) \le \ell$, then the number of fractional components for any extreme point of $\hat{\mathcal{Q}}$ is atmost ℓ. From Lemma 1, we know that there is a correspondence between extreme points of $\mathcal{P}$ and $\mathcal{Q}$. A similar correspondence also holds for extreme points of $\hat{\mathcal{P}}$ and $\hat{\mathcal{Q}}$ can be shown. Hence, the number of "fractional components" in any extreme point of $\hat{\mathcal{P}}$ is also bounded by ℓ. In particular, for any extreme point $(\mathbf{p}, p_0)$ of $\hat{\mathcal{P}}$, let

$$\mathcal{F}((\mathbf{p}, p_0)) = \{i \ge 1 \mid 0 < p_i < p_0\},$$

denote the set of "fractional components" in $(\mathbf{p}, p_0)$. We have the following result,

Corollary 1. *If $d(\mathcal{Q}) \le \ell$, then the number of fractional components for any extreme point $(\mathbf{p}, p_0)$ of $\hat{\mathcal{P}}$ is bounded by ℓ, i.e. $|\mathcal{F}((\mathbf{p}, p_0))| \le \ell$.*

4.1. PTAS Sketch

In the context of Lemma 1, whenever we refer a solution $(\mathbf{p}, p_0)$ as optimal (feasible) to problem (12) in the rest of the paper, it should be interpreted as optimality (feasibility) of $\mathbf{x} = \mathbf{p}/p_0$.

We consider the case where $d(\mathcal{Q})$ is a constant (say ℓ). From Corollary 1, we know that anl extreme point to (13) has at most ℓ "fractional" variables. A simple idea to make the optimal solution of (13) feasible for (12) is to ignore the "fractional variables, i.e., let $(\mathbf{p}, p_0) \in \hat{\mathcal{P}}$ be an optimal extreme point of (13). Define $(\hat{\mathbf{p}}, \hat{p}_0)$ as

$$\hat{p}_i = \begin{cases} 0 & \text{if } p_i < p_0 \\ \hat{p}_0 & \text{otherwise} \end{cases}$$

where

$$\hat{p}_0 = \frac{1}{c_0 + \sum_{i:\hat{p}_i \neq 0} c_i}.$$

Observe that we are ignoring at most ℓ variables of $(\mathbf{p}, p_0)$. If the variables we ignored are such that

$$a_i p_i \leq \frac{\epsilon}{\ell} R^* \quad \forall\ p_i < p_0, \tag{14}$$

then we have

$$\sum_{i \in \mathcal{F}(\mathbf{p}, p_0)} a_i p_i \leq \epsilon R^*,$$

which implies

$$(1-\epsilon) R^* \leq \sum_{i \notin \mathcal{F}(\mathbf{p}, p_0)} a_i p_i = \sum_i a_i \hat{p}_i,$$

and $(\hat{\mathbf{p}}, \hat{p}_0)$ is an $(1-\epsilon)$-approximate solution for (12). Note that in $(\mathbf{p}, p_0)$ there can be at most $\left\lceil \frac{\ell}{\epsilon} \right\rceil$ variables such that $a_i p_i > \frac{\epsilon}{\ell} R^*$. Therefore, to ensure (14) we guess the top $\left\lceil \frac{\ell}{\epsilon} \right\rceil$ variables contributing to the objective in (12), set those variables $p_i = p_0$ and solve the resulting linear program.

In Step 1 of Algorithm 1, we guess all set of solutions that contain at most $\lceil \frac{\ell}{\epsilon} \rceil$ positive variables. In Steps 3-6, we consider all subset of solutions that have strictly less than $\lceil \frac{\ell}{\epsilon} \rceil$ positive variables and compute the objective value for those subset of solutions that are consistent with the constraints. In Step-7, we consider all subset of solutions that have exactly $\lceil \frac{\ell}{\epsilon} \rceil$ positive variables. In Steps 8-9, we guess the top $\lceil \frac{\ell}{\epsilon} \rceil$ variables contributing to the objective and we solve the linear program z_{LP} by setting $p_i = p_0$ for those variables variables i. In Steps 10-12, we ignore the "fractional variables." In Step 16, we pick the solution corresponding to the maximum objective value among the considered subset of solutions. Theorem 2 establishes the validity of Algorithm 1.

THEOREM 2. *Let $d(\mathcal{Q}) \leq \ell$ and $(\hat{\mathbf{p}}, \hat{p}_0)$ be the solution obtained by Algorithm 1. Then $\sum_{i=0}^{n} a_i \hat{p}_i > (1-\epsilon) R^*$, where R^* is the optimal value of (12).*

PROOF. Let $(\mathbf{p}^*, p_0^*)$ be an optimal solution to (12), define S as:

$$S = \{ i \geq 1 \mid p_i^* > 0 \}.$$

Algorithm 1 PTAS for (12)

1: Set $\mathcal{S} = \left\{ S_t \subset \{1,2,\ldots,n\} \;\middle|\; |S_t| \le \left\lceil \frac{\ell}{\epsilon} \right\rceil \right\}$.

2: **for** $S_t \in \mathcal{S}$ **do**

3: **if** $|S_t| < \left\lceil \frac{\ell}{\epsilon} \right\rceil$ **then**

4: Obtain $(\hat{\mathbf{p}}_t, p_{t0})$ as follows: $\hat{p}_{t0} = \dfrac{1}{c_0 + \sum_{i \in S_t} c_i}$

$$\hat{p}_{ti} = \begin{cases} \hat{p}_{t0} & \text{if } i \in S_t \\ 0 & \text{otherwise} \end{cases}$$

5: **if** $(\hat{\mathbf{p}}_t, p_{t0})$ is feasible in (12) **then** Set $R_t = \sum_{i=0}^{n} a_i \hat{p}_{ti}$

6: **end if**

7: **else** Set $Q_t = \{i \in \{1,2,\ldots,n\} \mid i \notin S_t \text{ and } \exists\, j \in S_t \text{ such that } a_j \le a_i\}$

8: Consider the linear program (13) with additional constraints $p_i = p_0$ for all $i \in S_t$ and $p_i = 0$ for all $i \in Q_t$. Let the modified linear program be denoted by $z_{\mathsf{LP}}(t)$.

9: **if** $z_{\mathsf{LP}}(t)$ is feasible **then** Set $(\mathbf{p}_t^*, p_{t0}^*)$ as the optimal extreme point of $z_{\mathsf{LP}}(t)$.

10: Set $\hat{S}_t = \{i \mid p_{ti}^* = p_{t0}^*\}$

11: Obtain $(\hat{\mathbf{p}}_t, p_{t0})$ as follows: $\hat{p}_{t0} = \dfrac{1}{c_0 + \sum_{i \in S_t} c_i}$

$$\hat{p}_{ti} = \begin{cases} \hat{p}_{t0} & \text{if } i \in S_t \\ 0 & \text{otherwise} \end{cases}$$

12: Set $R_t = \sum_{i=0}^{n} a_i \hat{p}_{ti}$

13: **end if**

14: **end if**

15: **end for**

16: Set $t^* = \arg\max_t R_t$;

17: Output $(\hat{\mathbf{p}}, \hat{p}_0) = (\hat{\mathbf{p}}_{t^*}, \hat{p}_{t^*0})$

In Steps 3-6 of the algorithm we consider all the solutions that have strictly less than $\lceil \frac{\ell}{\epsilon} \rceil$. Hence, without loss of generality assume that $|S| \ge \left\lceil \frac{\ell}{\epsilon} \right\rceil$.

Now, without loss of generality, assume that

$$S = \{1,2,\ldots,k\}, \text{ for some } k \ge \left\lceil \frac{\ell}{\epsilon} \right\rceil \text{ and}$$
$$a_k \le a_{k-1} \le \cdots \le a_1$$
$$S_1 = \{1,2,\ldots,k^*\}, \text{ where } k^* = \left\lceil \frac{\ell}{\epsilon} \right\rceil.$$

Note that $p_1^* = p_2^* = \cdots = p_k^* = p_0^*$. Therefore,

$$a_k p_k^* \le a_{k-1} p_{k-1}^* \le \cdots \le a_1 p_1^*, \text{ which implies } a_{k^*} p_{k^*}^* < \frac{\epsilon}{\ell} R^*.$$

Now consider a feasible point of (12), $(\mathbf{p}_1, p_{10})$ defined as

$$p_{10} = \frac{1}{c_0 + \sum_{i \in S_1} \hat{c}_i}, \quad p_{1i} = \begin{cases} p_{10} & \text{if } i \in S_1 \\ 0 & \text{otherwise.} \end{cases} \text{ implying } p^*_{1i} < p_{1i} \text{ for all } i \in S_1$$

and since $(\mathbf{p}_1, p_{10})$ is a feasible point to (12), it follows that

$$\sum_{i \in S_1} a_i p_{1i} = \sum_{i=1}^{n} a_i p_{1i} < R^* \text{ which implies } \sum_{i \in S_1} a_i p^*_{1i} \le R^*$$

By construction of $z_{\mathsf{LP}}(1)$, we must also have $p^*_{1i} = 0$ for every $i > k^*$ and $a_{k^*} \le a_i$, implying

$$a_i p^*_{1i} < a_{k^*} p^*_{k^*} < \frac{\epsilon}{\ell} R^* \quad \text{for all } i > k^*.$$

Observe that $z_{\mathsf{LP}}(1) \ge R^*$ and the variables i in the extreme point $(\mathbf{p}^*_1, p^*_{10})$ that can be "fractional" are $i > k^*$. Therefore,

$$a_i p^*_{1i} < \frac{\epsilon}{\ell} R^* \ \forall\ i \ \in \mathcal{F}(\mathbf{p}^*_1, p^*_{10}).$$

Thus by Lemma 1 it follows that

$$\sum_{i \in \mathcal{F}(p^*)} a_i p^*_i(1) < \epsilon R^* \text{ which implies } (1-\epsilon) R^* \le z_{\mathsf{LP}}(1) - \epsilon R^* < \sum_{i=0}^{n} a_i \hat{p}_i(1).$$

□

4.2. Examples of $\mathcal{Q}$ with small $d(\mathcal{Q})$

So far, we have assumed that $d(\mathcal{Q}) \le \ell$ and restricted our attention to specific instances of the rational optimization problem (1) that satisfy this criteria. There are large classes of problems that can be formulated in the framework of the rational optimization problem (1) and also satisfy our assumption that $d(\mathcal{Q})$ is small. In this section, we revisit applications discussed in Section 3 and establish that $d(\mathcal{Q})$ is indeed small, enabling our PTAS approach to solve the more generic version of these problems.

Assortment Optimization with Cardinality Constraint: The polytope $\mathcal{Q}$ corresponding to the feasible region of cardinality constrained assortment optimization problem (10) is

$$\mathcal{Q} = \left\{ \mathbf{x} \;\middle|\; \sum_{i=1}^{n} x_i \le K,\ 0 \le x_i \le 1\ ,\ i = 1, 2, \ldots, n \right\}.$$

Note that the polytope $\mathcal{Q}$ is the intersection of the n-dimensional hypercube and the hyperplane $\sum_{i=1}^{n} x_i \le K$. We know that every extreme point $\mathbf{x}$ of $\mathcal{Q}$ is such that $\mathbf{x} \in \{0,1\}^n$ and every pair of adjacent extreme points in the n-dimensional hypercube only differ in two components. Hence, we have the following result

Lemma 2. *For $\mathcal{Q}$ corresponding to the cardinality constrained assortment optimization problem* (10)*, we have $d(\mathcal{Q})=2$.*

Joint Assortment and Display Optimization with Cardinality Constraints: The polytope $\mathcal{Q}$ corresponding to the feasible region of cardinality constrained joint assortment and display optimization problem (11) is

$$\mathcal{Q}=\left\{\mathbf{X}\;\middle|\;\sum_{j=1}^{m}x_{ij}\leq 1\,\forall i,\;\sum_{i=1}^{n}x_{ij}\leq K_j\,\forall j,\;0\leq x_{ij}\leq 1\;,\;i=1,2,\ldots,n,\;j=1,2,\ldots,m\right\}.$$

The constraints in problem (11) are the same as the transportation problem, the number of variables that are different in two adjacent extreme points of the LP relaxation of problem (11) is bounded by the maximum cycle length in the corresponding transportation network. Since the transportation network is a bipartite graph, the maximum cycle length cannot exceed twice the number of nodes in either of the partitions. Hence, we have the following result,

Lemma 3. *For $\mathcal{Q}$ corresponding to feasible region of cardinality constrained joint assortment and display optimization problem* (11)*, we have $d(\mathcal{Q})\leq 2m$, where m is the number of display segments.*

Theorem-2 and the Lemmas 2,3 establishes that there exists a PTAS for the above applications in the presence of an additional constraint.

5. A Computational Study

In this section, we study the computational performance of our PTAS algorithm for rational optimization over a TU constraint set with one additional constraint. In particular, we consider the "Cardinality constrained joint assortment and display optimization" problem where there is a capacity constraint in addition to the display constraints. Each item has capacity c_i and there is a bound C on the total capacity of items selected. The problem formulation is shown below.

$$\begin{aligned}
\underset{\mathbf{X}\in\{0,1\}^{n\times m}}{\text{maximize}}\quad & R(\mathbf{X})\\
\text{subject to}\quad & \mathbf{C_i}:\;\sum_{j=1}^{m}x_{ij}\leq 1,\,\forall i\,;\;\mathbf{C_j}:\;\sum_{j=1}^{m}x_{ij}\leq K_j;\\
& \sum_{i=1}^{n}\sum_{j=1}^{m}c_ix_{ij}\leq C\,;\quad x_{ij}\in\{0,1\}\,\forall\, i,\,j
\end{aligned}\tag{15}$$

It is to be noted that the problem (15) is NP hard and the existing techniques in the assortment optimization literature cannot be easily extended to solve this problem.

n(products)	m(segments)	z_{PTAS}/z_{LP}	Time for PTAS(secs)
10	2	0.9408	0.653
50	2	0.996	261.876
50	3	0.947	3606.466
100	2	0.994	2886.35
100	3	0.869	3648.761

Table 1 PTAS Performance, $\epsilon = 0.8$

5.1. Experimental Setup

To evaluate the performance of our PTAS algorithm we perform 5 experiments by varying the number of products ($n \in \{10, 50, 100\}$) and the number of display segments ($m \in \{2, 3\}$). For each experiment, we generate 10 random instances of problem (15). The parameters $\mathbf{v}, \mathbf{c}$ and $\mathbf{r}$ are chosen as uniform random numbers between 0 and 1, as the scale of these parameters does not change the optimal solution. For every instance, we solve the corresponding LP relaxation and implement a slightly modified version of the PTAS algorithm. All implementations have been done using Gurobi libraries in C++. In the modified version of PTAS, we enforce a time limit on the running time of the algorithm. Specifically, we restrict the time spent in guessing the top variables (steps 8-9 in Algorithm 1) to one hour. Although Lemma-3 bounds the number of fractional variables to $2m$, based on empirical observations, we relaxed the bound to m in order to decrease the number of computations. Hence, we only considered subsets of size not exceeding $\lceil \frac{m}{\epsilon} \rceil$ instead of the theoretically correct $\lceil \frac{2m}{\epsilon} \rceil$. To avoid trivial cases, the value of the capacity bound C is chosen appropriately to ensure that the additional capacity constraint is tight and the optimal solution of LP relaxation has atleast $\lceil m/\epsilon \rceil$ positive components.

5.2. Results

Table 1 summarizes performance for our PTAS approach. For each experiment, we report two quantities of interest namely i) the average ratio of approximate value obtained by the PTAS method and the LP solution (z_{PTAS}/z_{LP}) and ii) the average running time of the PTAS method. It is important to note that the LP solution (i.e. optimal solution to LP relaxation of (15)) is clearly an upper bound to the optimal solution to (15) itself and hence the ratio z_{PTAS}/z_{LP} is a conservative measure of PTAS performance. Even though we fixed $\epsilon = 0.8$, which theoretically guarantees only a 0.2 approximation, the approximate optimal value is on an average about 85% of the optimal value. This suggests that one can use a higher value of ϵ to avoid large computations and still obtain a reasonable approximation.

References

J. Correa, C. Fernandes and Y. Wakabayashi. 2010. Approximating a class of combinatorial problems with rational objective function. *Mathematical Programming*.

J. Davis, G. Gallego and H. Topaloglu. 2014. Assortment planning under the multinomial logit model with totally unimodular constraint structures. *Working Paper*.

A. Desir and V. Goyal. 2014. An FPTAS for capacity constrained assortment optimization. *Working Paper*.

S. Hashizume, M. Fukushima, N. Katoh and T. Ibaraki. 1987. Approximation algorithms for combinatorial fractional programming problems. *Mathematical Programming*.

A.G. Kok, M.L. Fisher and R. Vaidyanathan. 2003. Assortment planning: Review of literature and industry practice. *Retail Supply Chain Management*.

N. Megiddo, 1979. Combinatorial optimization with rational objective functions. *Mathematics of Operations Research*.

S. Mittal and A. Schulz. 2013. An FPTAS for optimizing a class of low-rank functions over a polytope. *Mathematical Programming*.

P. Rusmeivichientong, Z-J.M. Shen and D.B. Shmoys. 2006. A PTAS for capacitated sum of ratios optimization. *Operations Research Letters*.

K. Talluri, G.V. Ryzin. 2004. Revenue management under a general discrete choice model of consumer behaviour. *Management Science*.