Journal of Global Optimization
https://doi.org/10.1007/s10898-022-01131-5

Fractional 0–1 programming and submodularity

Shaoning Han[1] · Andrés Gómez[1] · Oleg A. Prokopyev[2]

Received: 24 October 2020 / Accepted: 9 January 2022

Abstract
In this note we study multiple-ratio fractional 0–1 programs, a broad class of $\mathcal{NP}$-hard combinatorial optimization problems. In particular, under some relatively mild assumptions we provide a complete characterization of the conditions, which ensure that a single-ratio function is submodular. Then we illustrate our theoretical results with the assortment optimization and facility location problems, and discuss practical situations that guarantee submodularity in the considered application settings. In such cases, near-optimal solutions for multiple-ratio fractional 0–1 programs can be found via simple greedy algorithms.

Keywords Fractional 0–1 programming · Hyperbolic 0–1 programming · Multiple ratios · Single ratio · Submodularity · Assortment optimization · Facility location · Greedy algorithm

1 Introduction

We consider a multiple-ratio *fractional* 0–1 program given by:

$$\max_{x\in\mathcal{F}} \sum_{k\in M} \frac{\sum_{i\in N} a_{ki}x_i}{b_{k0}+\sum_{i\in N} b_{ki}x_i}, \tag{1}$$

where $M = \{1,\ldots,m\}$, $N = \{1,\ldots,n\}$ and $\mathcal{F} := \{x \in \{0,1\}^n : Dx \le d\}$ for given $D \in \mathbb{R}^{q\times n}$ and $d \in \mathbb{R}^q$. Problem (1) is often referred to as a multiple-ratio *hyperbolic* 0–1 program. Problems of the form (1) can also be viewed as a class of set-function optimization problems that seek a subset S of N with its indicator variable $\mathbb{1}_S \in \mathbb{R}^n$, where the i-th element of $\mathbb{1}_S$ is 1 if and only if $i \in S$.

Throughout the note, we make the following assumptions:

A1: $b_{k0} + \sum_{i\in N} b_{ki}x_i > 0$ for all $k \in M$ and all $x \in \mathcal{F}\backslash\{0\}$.

A2: $a_{ki} \ge 0$, $b_{k0} \ge 0$ and $b_{ki} > 0$ for all $k \in M$ and $i \in N$.

A3: $\mathcal{F}$ is *downward closed*, i.e., if $S \in \mathcal{F}$ then $T \in \mathcal{F}$ for all $T \subseteq S$.

✉ Oleg A. Prokopyev
droleg@pitt.edu

1 Industrial and Systems Engineering, University of Southern California, Los Angeles, CA 90089, USA

2 Industrial Engineering, University of Pittsburgh, Pittsburgh, PA 15261, USA

Published online: 31 January 2022

Assumption **A1** is standard in fractional optimization [16, 17, 42]. In particular, if $b_{k0} > 0$, then it ensures that the denominator is strictly positive for each ratio and the objective function in (1) is well defined. Our main result (see Theorem 1 in Sect. 3) requires $b_{k0} > 0$ to hold. We provide additional discussions on (1) with $b_{k0} = 0$ in Sects. 3.4 and 4.2.

Assumption **A2** is not too restrictive as it naturally holds in many application settings, see examples in [17], including those considered in this note, see Sect. 4. For our results developed in this note we also require an additional relatively mild assumption **A3** on the structure of the feasible region, $\mathcal{F}$, in (1). We note that many types of feasible regions considered in the literature, such as $\mathcal{F} = 2^N$ (unconstrained problem), $\mathcal{F} = \{S \subseteq N : |S| \leq p\}$ for some positive integer p (cardinality constraint) and $\mathcal{F} = \left\{S \subseteq N : \sum_{i \in S} w_i \leq c\right\}$ for some weights $w \geq 0$ and $c \geq 0$ (capacity constraint) all satisfy assumption **A3**.

Applications of single- and multiple-ratio fractional 0–1 programs as in (1) appear in many diverse areas. For example, Méndez-Díaz et al. [43] discuss an assortment optimization problem under mixed multinomial logit choice models (MMNL). Tawarmalani et al. [58] consider a facility location problem, where a fixed number of facilities need to be located to service customers locations with the objective of maximizing a market share. Arora et al. [4] study a class of set covering problems in the context of airline crew scheduling that aim at covering all flights operated by an airline company. Furthermore, many combinatorial optimization problems can be formulated in the form (1) including the minimum fractional spanning tree problem [19, 59], the maximum mean-cut problem [32, 50] and the maximum clique ratio problem [54]. More application examples can be found in the studies by [14, 24, 37], the recent survey by Borrero et al. [17] and the references therein.

While in general problem (1) is $\mathcal{NP}$-hard even in the case of a single ratio [30, 48], single-ratio problems can be solved in polynomial time under **A1** and **A2** and additional assumptions, e.g., unconstrained problems [30], or problems where the convex hull of $\mathcal{F}$ is known [41, 52]. Furthermore, Rusmevichientong et al. [53] show that for the unconstrained multi-ratio problem, there is no approximation algorithm with polynomial running time that has an approximation factor better than $\mathcal{O}(1/m^{1-\delta})$ for any $\delta > 0$. Other related theoretical computational results are discussed in [48, 49].

Exact solution methods for (1) encompass mixed-integer programming reformulations [16, 24, 42], branch and bound algorithms [58], and other enumerative methods [17, 28, 29]. However, due to $\mathcal{NP}$-hardness of (1), these methods do not scale well when the size of the problem increases. Motivated by these computational complexity considerations, a number of studies rely on approximation schemes and heuristics for solving (1). Rusmevichientong et al. [51], Mittal and Schulz [44] and Désir et al. [22] all propose approximation algorithms for assortment optimization under the MMNL model when the number of customer segments, m, is fixed. Amiri et al. [3] develop a heuristic algorithm based on Lagrangian relaxation in the context of stochastic service systems. Prokopyev et al. [49] present a GRASP-based (Greedy Randomized Adaptive Search) heuristic for solving the cardinality constrained problems. Finally, simple greedy algorithms are also used in the literature [25, 34]. However, it is often not well understood when such algorithms perform well.

Contributions and outline The remainder of the note is organized as follows. In Sect. 2, we overview some necessary preliminaries and formulate our model (1) in terms of set functions.

In Sect. 3, we provide the main result of the note that characterizes the *submodularity* of a single ratio. Submodularity is often a key property for devising approximation algorithms [26, 46]. If the objective function can be identified as a submodular function, then simple greedy algorithms are capable of delivering high-quality solutions. In fact, it is possible to obtain

$(1-e^{-1})$-approximations under a variety of feasible regions—independently of the number of the ratios, m, involved–, thus improving over existing approximation methods for (1). We also discuss the connections between submodularity and monotonicity in the context of fractional 0–1 optimization.

In Sect. 4, we consider our theoretical results in the context of two applications—the assortment optimization and the p-choice facility location problems. For the assortment optimization problem, our results suggest that submodularity is linked to a phenomenon known as *cannibalization* [45], and naturally arises in several important scenarios. The results can also be applied in the case when there is a fixed cost associated with offering a product in the assortment [6, 35], which arises, for example, in online advertisement with costs-per-impression. For the p-choice facility location problem [58], we show how to reformulate the original problem in a desirable form that can be then exploited to benefit from the submodularity property. Finally, we conclude the note in Sect. 5.

2 Preliminaries

Next, we provide relevant background on submodular optimization, and discuss how our theoretical results can be used in practice; see also our brief discussion on applications in Sect. 4.

2.1 Notation

Let $a^k = (a_{ki})_{i \in N}$ and $b^k = (b_{ki})_{i \in N \cup \{0\}}$ for all $k \in M$, and for given $a^k \in \mathbb{R}^n$ and $b^k \in \mathbb{R}^{n+1}$, define

$$h(x; a^k, b^k) := \frac{\sum_{i \in N} a_{ki} x_i}{b_{k0} + \sum_{i \in N} b_{ki} x_i}.$$

Then equation (1) can be rewritten as

$$\max_{x \in \mathcal{F}} \sum_{k \in M} h(x; a^k, b^k). \tag{2}$$

This form appears in many applications such as the retail assortment and the p-choice facility location problems. Note that for each $x \in \{0, 1\}^n$, there is a unique set $S = \{i \in N : x_i = 1\} \subseteq N$, and conversely, each $S \subseteq N$ corresponds to an indicator vector $\mathbb{1}_S \in \{0, 1\}^n$. Thus, we can rewrite $h(x; a^k, b^k)$ as a set function

$$h(S; a^k, b^k) := h(\mathbb{1}_S; a^k, b^k),$$

and regard $\mathcal{F}$ as the domain of sets, i.e., $\mathcal{F} \subseteq 2^N$. Thereafter, we may use the vector form and the set form of (2) interchangeably for convenience. The main result of this note is a necessary and sufficient condition for the submodularity of each function $h(\cdot; a^k, b^k)$.

2.2 Submodularity and approximation algorithms

A set function $f : 2^N \to \mathbb{R}$ from the subsets of N to the real numbers is *submodular* over $\mathcal{F}$ if it exhibits diminishing returns, i.e., $f(S \cup \{i\}) - f(S) \geq f(T \cup \{i\}) - f(T)$ for all

$S \subseteq T \subseteq N \setminus \{i\}$ such that $T \cup \{i\} \in \mathcal{F}$. Equivalently, function f is submodular over $\mathcal{F}$ if

$$f(S \cup \{i, j\}) - f(S \cup \{j\}) \leq f(S \cup \{i\}) - f(S) \tag{3}$$

for all $S \subseteq N$ and $i, j \notin S$ such that $S \cup \{i, j\} \in \mathcal{F}$.

The greedy algorithm, see its pseudo-code in Algorithm 1, is a popular choice for tackling monotone submodular maximization problems because it is easy to implement and gives a constant-factor approximation in many cases. When the feasible region is a matroid, the greedy algorithm produces a solution with 1/2 approximation factor; see [26]. When the feasible region is given by a cardinality constraint, the approximation ratio can be improved to $(1 - e^{-1})$; see [46]. Other $(1 - e^{-1})$-approximation algorithms or near-optimal algorithms have also been provided for other classes of feasible regions over the years [18, 33, 55], for example, when $\mathcal{F}$ is defined with a single or multiple capacity constraints.

Algorithm 1 Greedy Algorithm for Submodular Function Maximization

Step 1. Set $S := \emptyset$.
Step 2. Set $A := \{\ell \in N \setminus S : \ S \cup \{\ell\} \in \mathcal{F}\}$.
Step 3. If $A \neq \emptyset$, set $\ell^* \in \arg\max_{\ell \in A} f(S \cup \{\ell\})$ and $S := S \cup \{\ell^*\}$. Go to *Step 2*.
Step 4. Return S.

We now discuss three settings in which the results of this note can be directly used to obtain approximation algorithms for fractional problems.

2.2.1 Constrained single-ratio problems

Consider a single-ratio instance of (1), i.e., $|M| = 1$, but the convex hull of $\mathcal{F}$ is not known, for example, when $\mathcal{F}$ is defined by multiple knapsack constraints, $\mathcal{F} = \left\{S \subseteq N : \sum_{i \in S} w_{\ell i} \leq c_\ell, \ \ell = 1, \ldots, m\right\}$. Clearly, this class of problems is NP-hard (recall our discussion in Sect. 1). Nonetheless, if function $h(\cdot; a^k, b^k)$ is monotone and submodular (see Proposition 1 in Sect. 3.2), then existing approximation algorithms can be used; see, e.g., [33].

2.2.2 Single ratio plus a linear function

Consider the special case of a two-ratio problem (1) where one of the ratios is a linear function, i.e.,

$$\max_{x \in \mathcal{F}} \sum_{i \in N} c_i x_i + \frac{\sum_{i \in N} a_i x_i}{b_0 + \sum_{i \in N} b_i x_i}. \tag{4}$$

Problem (4) is NP-hard and arises, for example, in assortment optimization problems with fixed costs of including an item into the assortment; see, e.g., [6, 35]. It can be easily checked [6] that if all $a_i/b_i = a_j/b_j$ for all $i \neq j$, in which case $\frac{\sum_{i \in N} a_i x_i}{b_0 + \sum_{i \in N} b_i x_i} = \lambda \frac{\sum_{i \in N} b_i x_i}{b_0 + \sum_{i \in N} b_i x_i}$ for some $\lambda > 0$, then problem (4) is a submodular optimization problem. In this note we provide more general conditions for verifying whether (4) is submodular; see, for example, Proposition 4 in Sect. 4.1.

2.2.3 Multiple ratios

The objective of the multiple-ratio problem (1) is submodular if all functions $h(\cdot; a^k, b^k)$ are submodular. While such conditions may appear to be quite restrictive, we show that several existing results concerning the tractability of assortment optimization problems correspond to submodularity of *each* ratio; for example, see Propositions 5 and 6 in Sect. 4.1.

2.3 Submodularity and cutting-plane methods

The submodularity results presented in this note can be systematically exploited to efficiently solve problems via mixed-integer programming solvers, even if the ratios in (1) are not submodular. Indeed, the convex envelope of a submodular function is described by its Lovász extension [39], which can be incorporated into branch-and-bound solvers via facet-defining valid inequalities; see [7, 8]. The concave envelope of a submodular function cannot be, in general, described efficiently, but valid inequalities to approximate it have been proposed nonetheless [1, 46].

Specifically, in the context of fractional optimization, given an arbitrary (and possibly non-submodular) ratio, Atamtürk and Narayanan [9] use the characterization of submodularity presented in this note to express it as a difference of submodular functions of the form

$$\frac{\sum_{i\in N} a_i x_i}{b_0 + \sum_{i\in N} b_i x_i} = \frac{\sum_{i\in N} (a_i + c_i) x_i}{b_0 + \sum_{i\in N} b_i x_i} - \frac{\sum_{i\in N} c_i x_i}{b_0 + \sum_{i\in N} b_i x_i}, \tag{5}$$

where both ratios are submodular. Then they use the aforementioned results to generate cutting planes corresponding to the convex and concave envelopes of each ratio, thus strengthening the mixed-integer programming formulations. We refer the reader to [9] for further details on the implementation of their method and corresponding computational results.

3 Submodularity of a single ratio and its implications

3.1 A necessary and sufficient condition

In this section, we give a necessary and sufficient condition for the submodularity of the function $h(\cdot)$, see Theorem 1. As a direct consequence, if $h(\cdot; a^k, b^k)$ satisfies the condition of Theorem 1 for every $k \in M$, then it follows that the fractional 0–1 program (2) admits a constant-factor approximation algorithm. For convenience, we drop the superscript k in a^k and b^k and use the notation $h(\cdot; a, b)$ throughout this section.

We first consider the case where $b_0 > 0$. The key result of this note is as follows:

Theorem 1 *If $b_0 > 0$, then function $h(\cdot; a, b)$ is submodular over $\mathcal{F}$ if and only if*

$$h(S \cup \{i\}; a, b) + h(S \cup \{j\}; a, b) \le \frac{a_i}{b_i} + \frac{a_j}{b_j} \tag{6}$$

for all $S \subseteq N$, and $i, j \notin S$ with $i \neq j$ such that $S \cup \{i\} \cup \{j\} \in \mathcal{F}$.

Proof Recall that assumption **A2** holds. Thus, the right-hand side of (6) is well-defined. Let $S \subseteq N$, let $i, j \notin S$ with $i \neq j$ satisfying $S \cup \{i\} \cup \{j\} \in \mathcal{F}$, and define $A_S = \sum_{j\in S} a_j$ and $B_S = b_0 + \sum_{j\in S} b_j$. Observe that $h(S; a, b) = A_S/B_S$. From (3) we find that $h(\cdot; a, b)$ is submodular if and only if

$$\frac{A_S + a_i + a_j}{B_S + b_i + b_j} - \frac{A_S + a_j}{B_S + b_j} \le \frac{A_S + a_i}{B_S + b_i} - \frac{A_S}{B_S}.$$

Multiplying both sides by $B_S(B_S + b_i + b_j)$, we get the equivalent condition

$$\begin{aligned}
& B_S\left(A_S + a_i + a_j\right) - B_S\left(1 + \frac{b_i}{B_S + b_j}\right)(A_S + a_j) \\
& \le B_S\left(1 + \frac{b_j}{B_S + b_i}\right)(A_S + a_i) - (B_S + b_i + b_j)A_S \\
\Leftrightarrow\ & a_i B_S - b_i B_S \frac{(A_S + a_j)}{B_S + b_j} \le a_i B_S - b_i A_S - b_j A_S + \frac{b_j}{B_S + b_i} B_S(A_S + a_i) \\
\Leftrightarrow\ & a_i B_S - b_i B_S \frac{(A_S + a_j)}{B_S + b_j} \le a_i B_S - b_i A_S + \frac{b_j}{B_S + b_i}\Big(B_S A_S + B_S a_i - (B_S + b_i)A_S\Big) \\
\Leftrightarrow\ & a_i B_S - b_i B_S \frac{(A_S + a_j)}{B_S + b_j} \le a_i B_S - b_i A_S + \frac{b_j}{B_S + b_i}(a_i B_S - b_i A_S).
\end{aligned}$$

Adding $b_i A_S + b_i B_S \frac{A_S + a_j}{B_S + b_j} - a_i B_S$ to both sides, we find

$$\begin{aligned}
& b_i A_S \le b_i B_S \frac{(A_S + a_j)}{B_S + b_j} + \frac{b_j}{B_S + b_i}(a_i B_S - b_i A_S) \\
\Leftrightarrow\ & b_i A_S\,(B_S + b_i)\left(B_S + b_j\right) \le b_i B_S(A_S + a_j)(B_S + b_i) + b_j(B_S + b_j)(a_i B_S - b_i A_S) \\
\Leftrightarrow\ & b_i A_S B_S^2 + b_i A_S B_S(b_i + b_j) + b_i^2 b_j A_S \\
& \le b_i A_S B_S^2 + b_i a_j B_S^2 + B_S A_S b_i^2 + a_j b_i^2 B_S + a_i b_j B_S^2 + a_i b_j^2 B_S - b_i b_j A_S B_S - b_i b_j^2 A_S.
\end{aligned}$$

After rearranging and canceling out some terms in the above expression, we obtain:

$$\begin{aligned}
& 2b_i b_j A_S B_S + b_i^2 b_j A_S + b_i b_j^2 A_S \le B_S^2(a_i b_j + a_j b_i) + a_j b_i^2 B_S + a_i b_j^2 B_S \\
\Leftrightarrow\ & b_i b_j A_S(b_i + b_j + 2B_S) \le b_i b_j\left(a_j/b_j B_S^2 + a_j/b_j b_i B_S + a_i/b_i B_S^2 + a_i/b_i b_j B_S\right) \\
\Leftrightarrow\ & A_S(b_i + b_j + 2B_S) \le a_j/b_j B_S^2 + a_j/b_j b_i B_S + a_i/b_i B_S^2 + a_i/b_i b_j B_S \\
\Leftrightarrow\ & (B_S + b_i)(A_S - a_j/b_j B_S) + (B_S + b_j)(A_S - a_i/b_i B_S) \le 0.
\end{aligned}$$

Finally, dividing by $(B_S + b_i)(B_S + b_j)$ and then adding $a_i/b_i + a_j/b_j$ on both sides, we get

$$\begin{aligned}
\Leftrightarrow\ & \left(\frac{A_S - a_j/b_j B_S}{B_S + b_j} + a_j/b_j\right) + \left(\frac{A_S - a_i/b_i B_S}{B_S + b_i} + a_i/b_i\right) \le a_i/b_i + a_j/b_j \\
\Leftrightarrow\ & \frac{A_S + a_j}{B_S + b_j} + \frac{A_S + a_i}{B_S + b_i} \le a_i/b_i + a_j/b_j,
\end{aligned}$$

which is precisely inequality (6). □

As we discuss next, submodularity is closely linked to monotonicity.

3.2 Monotonicity implies submodularity

The function $h(\cdot; a, b)$ is *monotone nondecreasing* if

$$h(S; a, b) \le h(S \cup \{j\}; a, b) \tag{7}$$

for every set S and $j \notin S$ such that $S \cup \{j\} \in \mathcal{F}$. Monotonicity is often a prerequisite for greedy algorithms, see, e.g., [46], to guarantee a constant approximation factor. Also, it arises naturally in many applications; see Sect. 4.1.1 for details. As we show next, monotonicity is a sufficient condition for submodularity.

Proposition 1 *If function $h(\cdot; a, b)$ is monotone nondecreasing, then $h(\cdot; a, b)$ is submodular.*

Proof Condition (7) is equivalent to

$$\begin{aligned}
&\frac{\sum_{i\in S} a_i}{b_0 + \sum_{i\in S} b_i} \le \frac{\sum_{i\in S} a_i + a_j}{b_0 + \sum_{i\in S} b_i + b_j} \\
\Leftrightarrow &\left(1 + \frac{b_j}{b_0 + \sum_{i\in S} b_i}\right) \sum_{i\in S} a_i \le \sum_{i\in S} a_i + a_j \\
\Leftrightarrow &\frac{\sum_{i\in S} a_i}{b_0 + \sum_{i\in S} b_i} \le \frac{a_j}{b_j} \\
\Leftrightarrow &h(S; a, b) \le \frac{a_j}{b_j}
\end{aligned} \tag{8}$$

for all S and $j \notin S$. Therefore, if $i, j \notin S$, then $h(S \cup \{i\}; a, b) \le a_i/b_i$ and $h(S \cup \{j\}; a, b) \le a_j/b_j$, and inequality (6) follows. □

Inequality (8) needs to hold for every combination of set S and element i for the function to be monotone. Note that $h(S \cup \{j\}; a, b)$ is the weighted average of $h(S; a, b)$ and a_j/b_j given by

$$h(S \cup \{j\}; a, b) = \left(\frac{b_0 + \sum_{i\in S} b_i}{b_0 + \sum_{i\in S} b_i + b_j}\right) \frac{\sum_{i\in S} a_i}{b_0 + \sum_{i\in S} b_i} + \left(\frac{b_j}{b_0 + \sum_{i\in S} b_i + b_j}\right) \frac{a_j}{b_j},$$

and hence, (8) is equivalent to $h(S \cup \{j\}; a, b) \le a_j/b_j$. Therefore, checking monotonicity can be done by verifying that

$$\frac{a_j}{b_j} \ge \max_{S\cup\{j\}\in\mathcal{F}} h(S \cup \{j\}; a, b) \quad \forall j \in N, \tag{9}$$

and the optimization problem on the right-hand side of (9) can be solved using existing algorithms for single-ratio fractional optimization; see [41, 50]. In fact, in some cases monotonicity can be verified without solving an optimization problem.

Corollary 1 *Function $h(\cdot; a, b)$ is monotone nondecreasing (and submodular) over 2^N if and only if*

$$\min_{i\in N} \frac{a_i}{b_i} \ge h(N; a, b).$$

Proof The forward direction follows directly from (9). For the backward direction, let $a^*/b^* = \min_{i\in N}\{a_i/b_i\}$, and then we find that

$$h(N; a, b) \le \frac{a^*}{b^*} \Leftrightarrow \frac{\sum_{i\in N} a_i}{b_0 + \sum_{i\in N} b_i} \le \frac{a^*}{b^*} \Leftrightarrow \sum_{i\in N} \left(\frac{a_i}{b_i} - \frac{a^*}{b^*}\right) b_i \le \frac{a^*}{b^*} b_0.$$

Since $a_i/b_i \ge a^*/b^*$, we find that $\sum_{i\in S}(a_i/b_i - a^*/b^*)b_i \le (a^*/b^*)\, b_0$ for any $S \subseteq N$, i.e., $h(S; a, b) \le a^*/b^*$ for any $S \subseteq N$. □

3.3 On non-monotone submodular functions

From Proposition 1, we know that monotonicity implies submodularity. In general, as Example 1 below shows, the converse does not hold.

Example 1 Assume we have three variables, i.e., $N = \{1, 2, 3\}$, with the setting $(a_1, a_2, a_3) = (3, 2, 1)$ and $(b_0, b_1, b_2, b_3) = (2, 1, 1, 1)$. Then from Theorem 1 we can verify that $h(\cdot; a, b)$ is submodular over 2^N: since $a_i/b_i + a_j/b_j \geq 3$ for any $i \neq j$ and, for any $S \subseteq N$, $h(S; a, b) \leq h(\{1, 2\}; a, b) = 5/4 \leq 3/2$, we find that inequality (6) holds. However, $h(\{3\}; a, b) = 1/3 < h(\{1, 2, 3\}; a, b) = 6/5 < h(\{1, 2\}; a, b) = 5/4$, and monotonicity does not hold. □

Nonetheless, if $h(\cdot; a, b)$ is submodular, then it is in fact very close to a nondecreasing function as shown in Proposition 2 below. In particular, if the decision variable with the smallest value a_i/b_i is fixed, then the resulting function is monotone.

Assume for the remainder of this section, without loss of generality, that $a_1/b_1 \geq a_2/b_2 \geq \cdots \geq a_n/b_n$. Define $\mathcal{F}_1 := \{S \in \mathcal{F} : n \in S\}$ and $\mathcal{F}_2 := \{S \in \mathcal{F} : n \notin S\}$.

Proposition 2 *If $h(\cdot; a, b)$ is submodular over $\mathcal{F}$, then the following holds:*

(i) function $h(\cdot; a, b)$ is monotone nondecreasing over $\mathcal{F}_1$;
(ii) for any $S \in \mathcal{F}_2$ and any $j \neq n$ such that $S \cup \{j\} \in \mathcal{F}$ and $S \cup \{n\} \in \mathcal{F}$, we have $h(S \cup \{j\}; a, b) \geq h(S; a, b)$.

Proof We first prove $h(\cdot; a, b)$ is monotone nondecreasing over $\mathcal{F}_1$ by contradiction. Assume there exists S and $j \neq n$ such that $n \notin S$ and $h(S \cup \{j, n\}; a, b) < h(S \cup \{n\}; a, b)$. Because $h(S \cup \{j, n\}; a, b)$ is a convex combination of $h(S \cup \{n\}; a, b)$ and a_j/b_j, we have $a_j/b_j < h(S \cup \{n\}; a, b)$. Since $a_n/b_n \leq a_j/b_j$, we find that $a_n/b_n < h(S \cup \{n\}; a, b)$. Note that

$$h(S \cup \{n\}; a, b) = \left(\frac{b_0 + \sum_{i\in S} b_i}{b_0 + b_n + \sum_{i\in S} b_i}\right) h(S; a, b) + \frac{b_n}{b_0 + b_n + \sum_{i\in S} b_i} \frac{a_n}{b_n} \tag{10}$$

is a convex combination of $h(S; a, b)$ and a_n/b_n, and since $a_n/b_n < h(S \cup \{n\}; a, b)$, it follows that $h(S \cup \{n\}; a, b) < h(S; a, b)$. By submodularity, $h(S \cup \{j, n\}; a, b) - h(S \cup \{j\}; a, b) \leq h(S \cup \{n\}; a, b) - h(S; a, b) < 0$, which indicates that $h(S \cup \{j, n\}; a, b) < h(S \cup \{j\}; a, b)$. Thus, $a_n/b_n < h(S \cup \{j\}; a, b)$. However, this implies $h(S \cup \{j\}; a, b) + h(S \cup \{n\}; a, b) > a_j/b_j + a_n/b_n$, which is a contradiction based on Theorem 1. Thus, (*i*) holds.

Next, we prove (*ii*) by contradiction. Assume there exists S and $j \neq n$ such that $n \notin S$ and $h(S \cup \{j\}; a, b) < h(S; a, b)$. Because $h(S \cup \{j\}; a, b)$ is the weighted average of a_j/b_j and $h(S; a, b)$, we have that $a_j/b_j < h(S \cup \{j\}; a, b) < h(S; a, b)$. Recall that $a_n/b_n \leq a_j/b_j$. Hence, $a_n/b_n < h(S; a, b)$, which implies $a_n/b_n < h(S \cup \{n\}; a, b)$—using similar arguments as in the proof of (*i*). Hence, $h(S \cup \{j\}; a, b) + h(S \cup \{n\}; a, b) > a_j/b_j + a_n/b_n$, which contradicts the submodularity of $h(\cdot; a, b)$. □

Corollary 2 *If either $\mathcal{F} = 2^N$ or $\mathcal{F} = \{S \subseteq N : |S| \leq p\}$ for any $p \in \{1, \ldots, n-1\}$, then submodularity of $h(\cdot; a, b)$ over $\mathcal{F}$ implies that $h(\cdot; a, b)$ is monotone nondecreasing over $\mathcal{F}_1$ and $\mathcal{F}_2$.*

Example 1 [Continued] Observe that $h(\cdot; a, b)$ is indeed monotone over $\mathcal{F}_1$, since $h(\{3\}; a, b) = 1/3$, $h(\{1, 3\}; a, b) = 1$, $h(\{2, 3\}; a, b) = 3/4$ and $h(\{1, 2, 3\}; a, b) = 6/5$. Similarly, we can verify that $h(\cdot; a, b)$ is monotone over $\mathcal{F}_2$ since $h(\emptyset; a, b) = 0$, $h(\{1\}; a, b) = 1$, $h(\{2\}; a, b) = 2/3$ and $h(\{1, 2\}; a, b) = 5/4$. □

3.4 On homogeneous fractional functions

In this section, we show that the assumption $b_0 > 0$ is indeed necessary in Theorem 1, as otherwise submodularity does not hold in most practical situations. Proposition 3 below formalizes this statement.

Proposition 3 *Assume $b_0 = 0$. If there exists a feasible set S such that there are at least three distinct values for a_i/b_i, $i \in S$, then $h(\cdot; a, b)$ is not submodular.*

Proof Assume without loss of generality that $a_1/b_1 < a_2/b_2 < a_3/b_3$. Then the following inequality

$$\begin{aligned}&\frac{b_1}{b_1+b_3}\left(\frac{a_3}{b_3}-\frac{a_1}{b_1}\right)+\frac{b_2}{b_2+b_3}\left(\frac{a_3}{b_3}-\frac{a_2}{b_2}\right)\\ &\geq \frac{b_1}{b_1+b_2+b_3}\left(\frac{a_3}{b_3}-\frac{a_1}{b_1}\right)+\frac{b_2}{b_1+b_2+b_3}\left(\frac{a_3}{b_3}-\frac{a_2}{b_2}\right).\end{aligned}$$

holds since denominators are greater on the right-hand side. Subtracting $2 \cdot (a_3/b_3)$ on both sides, we find that

$$-\frac{a_1+a_3}{b_1+b_3}-\frac{a_2+a_3}{b_2+b_3} \geq -\frac{a_1+a_2+a_3}{b_1+b_2+b_3}-\frac{a_3}{b_3},$$

which is equivalent to $h(\{1, 3\}; a, b) + h(\{2, 3\}; a, b) \leq h(\{1, 2, 3\}; a, b) + h(\{3\}; a, b)$, violating the definition of submodularity. □

3.5 Submodularity testing

In this section, we discuss how to verify whether $h(\cdot; a, b)$ is submodular over $\mathcal{F}$. By Theorem 1, to test for the submodularity, it suffices to compute

$$t_{ij} := \max_{\substack{S\cup\{i\}\cup\{j\}\in\mathcal{F}\\ i,j\notin S}} h(S\cup\{i\}; a, b) + h(S\cup\{j\}; a, b) \tag{11}$$

for each pair $\{i, j\}$, $i \neq j$, and check whether $t_{ij} \leq a_i/b_i + a_j/b_j$.

The maximization problem (11) involves an exponential number of candidates to be considered; hence, this problem is not trivial for a general feasible region $\mathcal{F}$. Nevertheless, for the unconstrained problems, i.e., $\mathcal{F} = 2^N$, using Algorithm 2 discussed below, submodularity testing can be achieved in polynomial time due to the connection between submodularity and monotonicity as outlined in Sects. 3.2 and 3.3.

Algorithm 2 Algorithm for submodularity testing with $\mathcal{F} = 2^N$

Step 1. Sort $\{a_i/b_i\}_{i=1}^n$ in the non-increasing order, i.e. $a_1/b_1 \geq a_2/b_2 \geq \cdots \geq a_n/b_n$.
Step 2. Compute and compare $h([n-1]; a, b)$ with a_{n-1}/b_{n-1}. If $h([n-1]; a, b) > a_{n-1}/b_{n-1}$, submodularity fails to hold and stop; otherwise, set $i := 1$ and go to *Step 3.*
Step 3. Set $t_{in} := h([n-1]; a, b) + h([n]\backslash\{i\}; a, b)$. If $t_{in} > a_i/b_i + a_n/b_n$, submodularity fails to hold and stop; otherwise, set $i := i + 1$. Next, if $i = n$, then go to *Step 4*; otherwise, go to *Step 3.*
Step 4. Return that submodularity holds.

Define $[k] := \{1, \ldots, k\}$ for any $k \in N$ and assume $\mathcal{F} = 2^N$. Recall that $\mathcal{F}_1 = \{S \in \mathcal{F} : n \in S\}$ and $\mathcal{F}_2 = \{S \in \mathcal{F} : n \notin S\}$. By Corollary 2, submodularity of $h(\cdot; a, b)$ must imply its monotonicity over $\mathcal{F}_1$ and $\mathcal{F}_2$. By Corollary 1 monotonicity over $\mathcal{F}_2$ is equivalent to $h([n-1]; a, b) \leq a_{n-1}/b_{n-1}$. Therefore, submodularity does not hold if $h([n-1]; a, b) > a_{n-1}/b_{n-1}$; see Step 2 of Algorithm 2.

Now assume $h([n-1]; a, b) \leq a_{n-1}/b_{n-1}$, and we need to verify monotonicity of $h(\cdot; a, b)$ over $\mathcal{F}_1$. Note that $h([n]; a, b)$ is the weighted average of $h([n-1]; a, b)$ and a_n/b_n; see equation (10) with $S = [n-1]$. Then it follows that $h([n]; a, b) \leq \max\{h([n-1]; a, b),\ a_n/b_n\} \leq a_{n-1}/b_{n-1}$, where the second inequality results from Step 1. Hence, we conclude that $h(\cdot; a, b)$ is monotone over $\mathcal{F}_1$ by Corollary 1.

From monotonicity of $h(\cdot; a, b)$ over $\mathcal{F}_1$ and $\mathcal{F}_2$, we find that

$$h(S \cup \{i\}; a, b) \leq \max\{h([n-1]; a, b),\ h([n]; a, b)\}$$

for all $S \subseteq N$ and $i \in N$. In view of (11), if $i \neq n$ and $j \neq n$,

$$t_{ij} \leq 2 \max\left\{h([n-1]; a, b),\ h([n]; a, b)\right\} \leq 2\frac{a_{n-1}}{b_{n-1}} \leq \frac{a_i}{b_i} + \frac{a_j}{b_j}.$$

If $j = n$, the optimal value of (11) is attained at $S = N \backslash \{i, n\}$ by monotonicity, which implies $t_{in} = h([n-1]; a, b) + h([n] \backslash \{i\}; a, b)$. This observation justifies Steps 3 and 4 of Algorithm 2 and concludes our discussion of the proposed approach.

4 Applications

In this section, we discuss the implications of our theoretical results in the context of the assortment optimization and the p-choice facility location problems.

4.1 Assortment optimization problem

In the assortment optimization problem, a firm offers a set of products to utility-maximizing customers. The goal of the firm is to choose an assortment of products that maximizes its expected revenue. It is a core revenue management problem pervasive in practice [56]. In this subsection, we mainly consider this problem under the mixed multinomial logit model (MMNL); see, e.g., [15, 40].

Formally, let N be the set of products that can be offered to customers. Denote by r_i the revenue perceived by the firm if a customer chooses product $i \in N$. Under the MMNL model, each product $i \in N$ is associated with a random weight $v_{ki} > 0$, and the no-purchase option is associated with weight $v_{k0} > 0$; these weights encode the relative preferences for the products by a customer of type $k \in M$, i.e., set M describes market segments.

Given the preference weights v^k, if assortment $S \subseteq N$ is offered, then the probability that a customer in $k \in M$ chooses product $i \in S$ is given by

$$q(i, S; v^k) = \frac{v_{ki}}{v_{k0} + \sum_{i \in S} v_{ki}}.$$

The conditional expected revenue from offering assortment $S \subseteq N$ is

$$r(S; v^k) = \sum_{i \in S} r_i q(i, S; v^k).$$

Taking the expectation over the random vector v^k, we formulate the assortment optimization problem under the MMNL model as

$$\max_{S\in\mathcal{F}} \mathbb{E}_v\left[r(S; v)\right] = \sum_{k\in M} p_k r(S; v^k), \tag{12}$$

where p_k is the probability of a customer to be in segment k and each realization of v can be interpreted as the preferences associated with a given customer of customer segment. We assume that the support of v is finite. Hence, (12) can be posed in the form of (1), where $a_{ki} = p_k r_i v_{ki}$, $b_{ki} = v_{ki}$ and $b_{k0} = v_{k0}$ for all $k \in M$ and $i \in N$. Thus, $a_{ki}/b_{ki} = p_k r_i$.

Finally, we note that $p_k \geq 0$ for each $k \in M$. Hence, for submodularity of the objective function in (12) it is sufficient to consider the single-ratio functions $r(\cdot; v^k), k \in M$. Therefore, in our discussion below when applying the results of Theorem 1 and Corollary 1 (with ratio a_i/b_i), the multiplier p_k can be dropped from consideration.

4.1.1 Cannibalization and submodularity

Intuitively, in retail assortment problems, monotonicity of the revenue function implies that there is limited cannibalization, i.e., the introduction of a new product i (when feasible) always increases the expected revenue perceived by the firm—despite that the revenue obtained from previously offered products in S might decrease slightly. To be more specific, this limited cannibalization phenomenon arises in online advertising: the probability that a given customer clicks on an ad is often quite low, and the advertiser usually profits from offering more ads within the limited number of spots on the webpage.

Let $r_{\min} = \min_{i\in N} r_i$ and $r_{\max} = \max_{i\in N} r_i$. By Proposition 1 and Corollary 1, we obtain the following results in terms of revenue functions immediately.

Corollary 3 *If function $r(\cdot; v)$ is monotone nondecreasing, then $r(\cdot; v)$ is submodular.*

Corollary 4 *Function $r(\cdot; v)$ is monotone nondecreasing (and submodular) over 2^N if and only if $r_{min} \geq r(N; v)$.*

4.1.2 Revenue spread, no-purchase probability and submodularity

When the revenues r of all products are identical, assortment optimization problems are known to be submodular maximization problems [6, 21]. Intuitively, one would expect that if the revenues are sufficiently close (but not identical), then submodularity should be preserved. Proposition 4 formalizes this intuition: if the gap between the largest and the smallest revenues is bounded above by the odds of no-purchase, then the function is nondecreasing and submodular.

Proposition 4 *If*

$$\frac{r_{\max} - r_{\min}}{r_{\min}} \leq \min_{S\in\mathcal{F}} \frac{1 - q(S; v)}{q(S; v)}, \tag{13}$$

then $r(\cdot; v)$ is nondecreasing and submodular, where $r_{\max}$ and $r_{\min}$ are the largest and smallest revenues, respectively, and $q(S; v) = \sum_{i\in S} q(i, S; v)$ is the probability that an item is purchased.

Proof Equation (13) can be rewritten as $r_{\max} q(S; v) \leq r_{\min}$ for all $S \in \mathcal{F}$. Since for any S and $i \notin S$ it follows that $r(S; v) \leq r_{\max} q(S; v) \leq r_{\min} \leq r_i$, we find that (8) is satisfied and the function $r(\cdot; v)$ is monotone submodular. □

Proposition 4 provides us with additional intuition on the industries in which the expected revenues are submodular functions of the assortment offered. In the online advertisement, where the revenues obtained from clicks are usually similar and the odds of no-purchase are high, we would expect to obtain submodular revenue functions. In a *monopoly*, the firm offering the assortment would have a large flexibility in setting prices (resulting in a large revenue spread) and the odds of no-purchase would be low (due to the lack of competing alternatives), resulting in a revenue function that is not submodular. In contrast, in a *competitive market*, the odds of no-purchase would be larger and firms have little or no control over prices (and if the values r_i are interpreted as profits instead of revenues, the spread would typically be low), resulting in submodular revenue functions.

From Proposition 4 we also gain insights on the differences between revenue management in the airline and hospitality industries, two industries that are often treated as equivalent in the literature [57]. In the hospitality industries, no-purchase odds can be high as shown by the relatively low occupancy rates—66.1% in the US [31] in 2018; in addition, revenue differences between products are often due to ancillary charges (e.g., breakfast, non-refundable, long stay), which account for a small portion of the baseline price for a room. In such circumstances we would expect revenue functions to be submodular and simple greedy heuristics to perform well. In contrast, in the airline industries no-purchase odds are often smaller—the load factor was 86.1% in the US in 2018 [27]—, and air fares can change dramatically depending on the conditions. Thus, in the airline industry we would expect to encounter non-submodular revenue functions, and simple heuristics may be inadequate.

4.1.3 On the greedy algorithm and revenue-ordered assortments

Revenue-ordered assortments are optimal for unconstrained assortment optimization under the MNL model, and tend to perform well in practice [56]. Berbeglia and Joret [13] study the revenue-ordered assortments under the general discrete choice model and prove performance guarantees.

Proposition 5 (Berbeglia and Joret [13]) *Revenue-ordered assortments are a* $\frac{1}{1+\log\left(\frac{r_{\max}}{r_{\min}}\right)}$*-approximation for the unconstrained assortment optimization problem under the MMNL choice model, where* $r_{\max}$ *and* $r_{\min}$ *are the largest and smallest revenues, respectively.*

Thus, the quality of revenue-ordered assortments depend on the ratio r_{max}/r_{min}; in particular, if $r_{max}/r_{min} = 1$, then the revenue-ordered assortments strategy delivers an optimal solution, and the guarantee degrades as the value of the ratio increases.

From Proposition 4, we can also obtain guarantees depending on the ratio r_{max}/r_{min}. Define:

$$\alpha(S) = \max_{k\in M} q(S; v^k) = \max_{k\in M} \sum_{i\in S} q(i, S; v^k) \tag{14}$$

as the maximum probability that a customer from any segment purchases an item when assortment S is offered.

Proposition 6 *If* $\mathcal{F} = \{S : |S| \leq p\}$ *for some positive integer* p *and* $r_{max}/r_{min} \leq 1 + \frac{1-\alpha(S)}{\alpha(S)}$ *for all* $S \in \mathcal{F}$*, then Algorithm 1 delivers a* $(1 - 1/e)$*-optimal solution for the assortment optimization problem under the MMNL choice model.*

Unlike Proposition 5, we impose a condition on the ratio r_{max}/r_{min} in Proposition 6; however, if such condition is satisfied, then we obtain an approximation guarantee of $(1 -$

$1/e) \approx 0.63$ for the more general assortment optimization problem under a cardinality constraint.

Finally, we also point out that Rusmevichientong et al. [53] prove that if customers are *value conscious*, i.e., $v_1 \leq v_2 \leq \ldots \leq v_n$ and $r_1 v_1 \geq r_2 v_2 \geq \ldots \geq r_n v_n$ for all realizations of v, then the revenue ordered assortments are optimal for the unconstrained and cardinality constrained cases. It is easy to check that in this case the solutions obtained from the greedy algorithm correspond precisely with the revenue ordered assortments. Thus, Algorithm 1 delivers optimal solutions as well.

Remark 1 We comment that our results so far are only applicable to those logit models with linear fractional objectives. When the revenue function involves a ratio of nonlinear functions, e.g., the nested logit models in assortment optimization [2], how to test and exploit the submodularity is still an open problem and needs more work to be done in the future.

4.2 *p*-choice facility location problem

Facility location problems deal with deciding where to locate facilities across a finite set of feasible points, taking into account the needs of customers to be served in such a way that a given economic index is optimized [10]. Submodularity often arises in facility location problems. For example, Benati [11] considers the maximum facility location problem with random utilities (MCFLRU); see also [12, 38]. Since in the MCFLRU all ratios a_i/b_i are identical, submodularity follows directly from Proposition 4. Dam et al. [20] study the maximum capture problem in facility location under random utility models, where the objective function is a sum of the multiplicative inverses of a nonlinear choice probability generating function (CPGF). The authors [20] show that if the CPGF is increasing and submodular, the resulting problem is a submodular maximization problem. We also refer the reader to [5, 23, 36, 47] for additional studies on the applications of submodularity to facility location problems.

In this subsection, we consider a particular class of facility location problems with a fractional 0–1 objective function, referred to as the p-choice facility location problem, which is considered in [58]. In the p-choice facility location problem, a decision-maker has to decide where to locate p facilities in n possible locations to service m demand points, in order to maximize the market share.

Formally, let $d_k > 0$ be the demand at customer location $k \in M = \{1, \ldots, m\}$, and $v_{ki} > 0$ be the utility of location i to customers at k. Let $S \subseteq N := \{1, \ldots, n\}$, $|S| = p$, be the set of facilities chosen by the decision-maker. It is assumed that the market share provided by facility $j \in S$ with respect to demand point k is given by:

$$d_k \frac{v_{kj}}{\sum_{i \in S} v_{ki}}.$$

Let $w_i > 0$ be some weight parameter that represents the importance of locating facility in location $i \in N$. Then the problem of determining the set of facility locations S that maximizes the weighted market share can be formulated as:

$$\max_{|S|=p} \sum_{i \in S} w_i \sum_{k \in M} d_k \frac{v_{ki}}{\sum_{j \in S} v_{kj}},$$

which can be reorganized as

$$\max_{|S|=p} \sum_{k\in M} d_k \frac{\sum_{i\in S} v_{ki} w_i}{\sum_{i\in S} v_{ki}}. \tag{15}$$

Clearly, the model in (15) can be formulated as a fractional 0–1 program given by (1).

Note that from Proposition 3, the objective function in (15) is, in general, not submodular since it is homogeneous. Nonetheless, exploiting the equality constraint, we can convert the objective function to a non-homogeneous one. Define $v^k_{\min} = \delta \cdot \min_{i\in N}\{v_{ki}\}$ for some fixed $\delta \in (0, 1)$. For any feasible solution S, where $|S| = p$, we also have that:

$$\sum_{i\in S} v_{ki} = \sum_{i\in S} v^k_{\min} + \sum_{i\in S}(v_{ki} - v^k_{\min}) = pv^k_{\min} + \sum_{i\in S}(v_{ki} - v^k_{\min}).$$

As a result, (15) can be equivalently stated as:

$$\max_{|S|=p} \sum_{k\in M} d_k \frac{\sum_{i\in S} v_{ki} w_i}{pv^k_{\min} + \sum_{i\in S}(v_{ki} - v^k_{\min})}, \tag{16}$$

where $v_{ki} - v^k_{\min} > 0$ and $v^k_{\min} > 0$ for all $i \in N$ and $k \in M$ by our construction procedure.

Recall our discussion on the links between monotonicity and submodularity in Sect. 3.2. Applying inequality (9), we find that a given ratio k in the objective function of (16) is monotone nondecreasing over set $\mathcal{F} := \{S \subseteq N : |S| \leq p\}$ if

$$\min_{i\in N} \frac{v_{ki} w_i}{v_{ki} - v^k_{\min}} \geq \max_{|S|\leq p} \frac{\sum_{i\in S} v_{ki} w_i}{pv^k_{\min} + \sum_{i\in S}(v_{ki} - v^k_{\min})}. \tag{17}$$

Hence, if (17) holds for all ratios $k \in M$, then the feasibility set in (16) can be relaxed to $|S| \leq p$. Consequently, assumption **A3** is satisfied and (16) reduces to the maximization problem of a submodular function by Proposition 1.

The right-hand side of (17) can be interpreted as the best average revenue weighted by market share, or simply the best total revenue that can be obtained from customer segment k. The intuition for (17) to hold in the p-choice facility location problem is rather similar to our observations in the assortment optimization problem. Indeed, it is easy to verify, for example, that if all locations have the same utilities and weights, i.e., $v_{ki} = v^k$ and $w_i = w$ for all $i \in N$ and some v^k and w, then (17) holds. Moreover, from (17) we obtain the following sufficient condition.

Proposition 7 *Let $w_{\max}$ and $w_{\min}$ be the maximum and minimum weights, and let $v^k_{\max}$ be the maximum utility associated with customer segment k. If*

$$\frac{w_{\min}}{w_{\max}} + 1 \geq \frac{v^k_{\max}}{v^k_{\min}}, \tag{18}$$

then the revenue of customer segment k is submodular.

Proof Observe that since $w_i \geq w_{\min}$ and $\frac{v_{ki}}{v_{ki}-v^k_{\min}} \geq \frac{v^k_{\max}}{v^k_{\max}-v^k_{\min}}$, we find that

$$\frac{v_{ki} w_i}{v_{ki} - v^k_{\min}} \geq \frac{v^k_{\max}}{v^k_{\max} - v^k_{\min}} w_{\min}.$$

Moreover, we also find that

$$\max_{|S|\leq p} \frac{\sum_{i\in S} v_{ki} w_i}{pv^k_{\min} + \sum_{i\in S}(v_{ki} - v^k_{\min})} \leq \max_{|S|\leq p} \frac{w_{\max} \sum_{i\in S} v_{ki}}{pv^k_{\min}} \leq \frac{w_{\max} v^k_{\max}}{v^k_{\min}}.$$

After rearranging terms corresponding to the sufficient condition

$$\frac{v_{\max}^k}{v_{\max}^k - v_{\min}^k} w_{\min} \geq \frac{w_{\max} v_{\max}^k}{v_{\min}^k},$$

we obtain precisely (18). □

Simply speaking, if the considered facility locations are sufficiently similar with respect to their utilities, i.e., $\frac{v_{\max}^k}{v_{\min}^k} \approx 1$, then ratios in (16) are submodular. Submodularity may be preserved for larger spread of utilities, provided that the weights are sufficiently close. If all the considered facility locations are sufficiently similar with respect to their utilities and weights, then (15) can be reduced to maximizing a submodular function; consequently, high-quality solutions can be obtained by a greedy approach, e.g., Algorithm 1.

4.3 On minimization problems

In this note we focus on identifying submodularity in maximization problems, in which case greedy algorithms can be used to obtain near optimal solutions. However, submodularity can be exploited in minimization problems as well. Indeed, the epigraph of a submodular set function is described by its Lovász extension [39], which can be used to improve mixed-integer programming formulations via cutting planes. Moreover, even if a given ratio is not submodular, the results presented in this note can be used to decompose any ratio into two components such that one of which is submodular (and strengthening can be done using the submodular component); see, e.g., [9].

5 Conclusion

In this note we explore submodularity of the objective function for a broad class of fractional 0–1 programs with multiple-ratios. Under some mild assumptions, we derive the necessary and sufficient condition for a single ratio of two linear functions to be submodular. Therefore, if the derived condition holds for every considered single-ratio function, then simple greedy algorithms can be used to deliver good quality solutions for multiple-ratio fractional 0–1 programs. Finally, we also illustrate applicability of our results in the context of the assortment optimization and facility location problems.

Acknowledgements This note is based upon work supported by the National Science Foundation under Grant No. 1818700. The authors would like to thank the Associate Editor and two anonymous referees for their constructive and helpful comments.

References

1. Ahmed, S., Atamtürk, A.: Maximizing a class of submodular utility functions. Math. Program. **128**(1–2), 149–169 (2011)
2. Alfandari, L., Hassanzadeh, A., Ljubić, I.: An exact method for assortment optimization under the nested logit model. Eur. J. Oper. Res. **291**(3), 830–845 (2021)
3. Amiri, A., Rolland, E., Barkhi, R.: Bandwidth packing with queuing delay costs: Bounding and heuristic solution procedures. Eur. J. Oper. Res. **112**(3), 635–645 (1999)
4. Arora, S., Puri, M., Swarup, K.: The set covering problem with linear fractional functional. Indian J. Pure Appl. Math. **8**(5), 578–588 (1977)

5. Atamtürk, A., Berenguer, G., Shen, Z.-J.: A conic integer programming approach to stochastic joint location-inventory problems. Oper. Res. **60**(2), 366–381 (2012)
6. Atamtürk, A., Gómez, A.: Maximizing a class of utility functions over the vertices of a polytope. Oper. Res. **65**(2), 433–445 (2017)
7. Atamtürk, A., Gómez, A.: Submodularity in conic quadratic mixed 0–1 optimization. Oper. Res. **68**(2), 609–630 (2020)
8. Atamtürk, A., Narayanan, V.: Polymatroids and mean-risk minimization in discrete optimization. Oper. Res. Lett. **36**(5), 618–622 (2008)
9. Atamtürk, A., Narayanan, V.: Submodular function minimization and polarity. Math. Program. 1–11 (2021)
10. Barros, A.I.: *Discrete and Fractional Programming Techniques for Location Models*, vol. 3. Springer Science & Business Media (2013)
11. Benati, S.: Submodularity in competitive location problems. Ricerca Oper. (1996)
12. Benati, S., Hansen, P.: The maximum capture problem with random utilities: problem formulation and algorithms. Eur. J. Oper. Res. **143**(3), 518–530 (2002)
13. Berbeglia, G., Joret, G.: Assortment optimisation under a general discrete choice model: a tight analysis of revenue-ordered assortments. In: Proceedings of the 2017 ACM Conference on Economics and Computation, EC '17, pp. 345–346. New York, NY, USA. ACM (2017)
14. Bertsimas, D., Korolko, N., Weinstein, A.M.: Identifying exceptional responders in randomized trials: an optimization approach. INFORMS J. Optim. **1**(3), 187–199 (2019)
15. Bonnet, C., Simioni, M.: Assessing consumer response to protected designation of origin labelling: a mixed multinomial logit approach. Eur. Rev. Agric. Econ. **28**(4), 433–449 (2001)
16. Borrero, J.S., Gillen, C., Prokopyev, O.A.: A simple technique to improve linearized reformulations of fractional (hyperbolic) 0–1 programming problems. Oper. Res. Lett. **44**(4), 479–486 (2016)
17. Borrero, J.S., Gillen, C., Prokopyev, O.A.: Fractional 0–1 programming: applications and algorithms. J. Global Optim. **69**(1), 255–282 (2017)
18. Calinescu, G., Chekuri, C., Pál, M., Vondrák, J.: Maximizing a monotone submodular function subject to a matroid constraint. SIAM J. Comput. **40**(6), 1740–1766 (2011)
19. Chandrasekaran, R.: Minimal ratio spanning trees. Networks **7**(4), 335–342 (1977)
20. Dam, T.T., Ta, T.A., Mai, T.: Submodularity and local search approaches for maximum capture problems under generalized extreme value models. Eur. J. Oper. Res. (2021)
21. Désir, A., Goyal, V., Segev, D., Ye, C.: Capacity constrained assortment optimization under the Markov chain based choice model. Working paper, Columbia University, New York, NY (2015). http://dx.doi.org/10.2139/ssrn.2626484
22. Désir, A., Goyal, V., Zhang, J.: Near-optimal algorithms for capacity constrained assortment optimization. Working paper, Columbia University, NY (2014). http://dx.doi.org/10.2139/ssrn.2543309
23. Du, D., Lu, R., Xu, D.: A primal-dual approximation algorithm for the facility location problem with submodular penalties. Algorithmica **63**(1–2), 191–200 (2012)
24. Elhedhli, S.: Exact solution of a class of nonlinear knapsack problems. Oper. Res. Lett. **33**(6), 615–624 (2005)
25. Feldman, J., Topaloglu, H.: Bounding optimal expected revenues for assortment optimization under mixtures of multinomial logits. Prod. Oper. Manag. **24**(10), 1598–1620 (2015)
26. Fisher, M.L., Nemhauser, G.L., Wolsey, L.A.: An analysis of approximations for maximizing submodular set functions–II. In: Balinski, M.L., Hoffman, A.J. (eds.) Polyhedral Combinatorics, pp. 73–87. Springer, Berlin (1978)
27. Goldstein, M.: Meet the most crowded airlines: Load factor hits all-time high (2018). https://www.forbes.com/sites/michaelgoldstein/2018/07/09/meet-the-most-crowded-airlines-load-factor-hits-all-time-high/#90f0b4354fbd. Accessed 12 April 2019 (Forbes)
28. Granot, D., Granot, F.: On solving fractional (0, 1) programs by implicit enumeration. INFOR Inform. Syst. Oper. Res. **14**(3), 241–249 (1976)
29. Hansen, P., De Aragão, M.V.P., Ribeiro, C.C.: Boolean query optimization and the 0–1 hyperbolic sum problem. Ann. Math. Artif. Intell. **1**(1–4), 97–109 (1990)
30. Hansen, P., De Aragão, M.V.P., Ribeiro, C.C.: Hyperbolic 0–1 programming and query optimization in information retrieval. Math. Program. **52**(1–3), 255–263 (1991)
31. Hoisington, A.: Hotel Management (2018). https://www.hotelmanagement.net/own/occupancy-hits-30-year-high-u-s. Accessed 12 April 2019
32. Iwano, K., Misono, S., Tezuka, S., Fujishige, S.: A new scaling algorithm for the maximum mean cut problem. Algorithmica **11**(3), 243–255 (1994)

33. Kulik, A., Shachnai, H., Tamir, T.: Maximizing submodular set functions subject to multiple linear constraints. In: Mathieu, C. (ed.), Proceedings of the Twentieth Annual ACM-SIAM Symposium on Discrete Algorithms, pp. 545–554. Society for Industrial and Applied Mathematics (2009)
34. Kunnumkal, S.: On upper bounds for assortment optimization under the mixture of multinomial logit models. Oper. Res. Lett. **43**(2), 189–194 (2015)
35. Kunnumkal, S., Martínez-de Albéniz, V.: Tractable approximations for assortment planning with product costs. Oper. Res. **67**(2), 436–452 (2019)
36. Li, Y., Du, D., Xiu, N., Xu, D.: Improved approximation algorithms for the facility location problems with linear/submodular penalties. Algorithmica **73**(2), 460–482 (2015)
37. Lin, Y.H., Tian, Q.: Exact approaches for competitive facility location with discrete attractiveness. Optim. Lett. **15**, 377–389 (2021)
38. Ljubić, I., Moreno, E.: Outer approximation and submodular cuts for maximum capture facility location problems with random utilities. Eur. J. Oper. Res. **266**(1), 46–56 (2018)
39. Lovász, L. (1983). Submodular functions and convexity. In *Mathematical programming the state of the art*, pages 235–257. Springer
40. McFadden, D., Train, K.: Mixed MNL models for discrete response. J. Appl. Economet. **15**(5), 447–470 (2000)
41. Megiddo, N., et al.: Combinatorial optimization with rational objective functions. Math. Oper. Res. **4**(4), 414–424 (1979)
42. Mehmanchi, E., Gómez, A., Prokopyev, O.A.: Fractional 0–1 programs: links between mixed-integer linear and conic quadratic formulations. J. Global Optim. **75**(2), 273–339 (2019)
43. Méndez-Díaz, I., Miranda-Bront, J.J., Vulcano, G., Zabala, P.: A branch-and-cut algorithm for the latent-class logit assortment problem. Discret. Appl. Math. **164**, 246–263 (2014)
44. Mittal, S., Schulz, A.S.: A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one. Oper. Res. **61**(2), 386–397 (2013)
45. Moorthy, K.S., Png, I.P.: Market segmentation, cannibalization, and the timing of product introductions. Manage. Sci. **38**(3), 345–359 (1992)
46. Nemhauser, G.L., Wolsey, L.A., Fisher, M.L.: An analysis of approximations for maximizing submodular set functions-I. Math. Program. **14**(1), 265–294 (1978)
47. Ortiz-Astorquiza, C., Contreras, I., Laporte, G.: Formulations and approximation algorithms for multilevel uncapacitated facility location. INFORMS J. Comput. **29**(4), 767–779 (2017)
48. Prokopyev, O.A., Huang, H.-X., Pardalos, P.M.: On complexity of unconstrained hyperbolic 0–1 programming problems. Oper. Res. Lett. **33**(3), 312–318 (2005)
49. Prokopyev, O.A., Meneses, C., Oliveira, C.A., Pardalos, P.M.: On multiple-ratio hyperbolic 0–1 programming problems. Pacific J. Optim. **1**(2), 327–345 (2005)
50. Radzik, T.: Fractional combinatorial optimization. In: Du, D.-Z., Pardalos, P.M. (eds.) Handbook of Combinatorial Optimization, pp. 429–478. Springer-Verlag, New York (1998)
51. Rusmevichientong, P., Shen, Z.-J.M., Shmoys, D.B.: A PTAS for capacitated sum-of-ratios optimization. Oper. Res. Lett. **37**(4), 230–238 (2009)
52. Rusmevichientong, P., Shen, Z.-J.M., Shmoys, D.B.: Dynamic assortment optimization with a multinomial logit choice model and capacity constraint. Oper. Res. **58**(6), 1666–1680 (2010)
53. Rusmevichientong, P., Shmoys, D., Tong, C., Topaloglu, H.: Assortment optimization under the multinomial logit model with random choice parameters. Prod. Oper. Manag. **23**(11), 2023–2039 (2014)
54. Sethuraman, S., Butenko, S.: The maximum ratio clique problem. CMS **12**(1), 197–218 (2015)
55. Sviridenko, M.: A note on maximizing a submodular set function subject to a knapsack constraint. Oper. Res. Lett. **32**(1), 41–43 (2004)
56. Talluri, K., van Ryzin, G.: Revenue management under a general discrete choice model of consumer behavior. Manage. Sci. **50**(1), 15–33 (2004)
57. Talluri, K.T., van Ryzin, G.J.: The Theory and Practice of Revenue Management, vol. 68. Springer Science & Business Media (2006)
58. Tawarmalani, M., Ahmed, S., Sahinidis, N.V.: Global optimization of 0–1 hyperbolic programs. J. Global Optim. **24**(4), 385–416 (2002)
59. Ursulenko, O., Butenko, S., Prokopyev, O.A.: A global optimization algorithm for solving the minimum multiple ratio spanning tree problem. J. Global Optim. **56**(3), 1029–1043 (2013)