FULL LENGTH PAPER

Series B

Submodular maximization of concave utility functions composed with a set-union operator with applications to maximal covering location problems

Stefano Coniglio[1,2] · **Fabio Furini**[3] · **Ivana Ljubić**[4]

Abstract

We study a family of discrete optimization problems asking for the maximization of the expected value of a concave, strictly increasing, and differentiable function composed with a set-union operator. The expected value is computed with respect to a set of coefficients taking values from a discrete set of scenarios. The function models the utility function of the decision maker, while the set-union operator models a covering relationship between two ground sets, a set of items and a set of metaitems. This problem generalizes the problem introduced by Ahmed S, Atamtürk A (Mathematical programming 128(1-2):149–169, 2011), and it can be modeled as a mixed integer nonlinear program involving binary decision variables associated with the items and metaitems. Its goal is to find a subset of metaitems that maximizes the total utility corresponding to the items it covers. It has applications to, among others, maximal covering location, and influence maximization problems. In the paper, we propose a double-hypograph decomposition which allows for projecting out the variables associated with the items by separately exploiting the structural properties of the utility function and of the set-union operator. Thanks to it, the utility function is linearized via an exact outer-approximation technique, whereas the set-union operator is linearized

---

✉ Stefano Coniglio
   s.coniglio@soton.ac.uk; stefano.coniglio@unibg.it

   Fabio Furini
   fabio.furini@uniroma1.it

   Ivana Ljubić
   ivana.ljubic@essec.edu

[1] Mathematical Sciences, University of Southampton, Southampton, UK

[2] Department of Economics, University of Bergamo, Bergamo, Italy

[3] Department of Computer, Control and Management Engineering "Antonio Ruberti", Sapienza University of Rome, Rome, Italy

[4] Information Systems, Decision Sciences and Statistics Department, ESSEC Business School of Paris, Cergy-Pontoise, France

in two ways: either (i) via a reformulation based on submodular cuts, or (ii) via a Benders decomposition. We analyze from a theoretical perspective the strength of the inequalities of the resulting reformulations, and embed them into two branch-and-cut algorithms. We also show how to extend our reformulations to the case where the utility function is not necessarily increasing. We then experimentally compare our algorithms *inter se*, to a standard reformulation based on submodular cuts, to a state-of-the-art global-optimization solver, and to the greedy algorithm for the maximization of a submodular function. The results reveal that, on our testbed, the method based on combining an outer approximation with Benders cuts significantly outperforms the other ones.

Keywords Submodular maximization · Branch-and-Cut · Benders decomposition · Stochastic maximal covering location problems · Influence maximization

Mathematics Subject Classification 90C57 · 90C35 · 90C30 · 90C27 · 90C11

1 Introduction

The problem of maximizing a concave, strictly increasing[1], and differentiable utility function $f : \mathbb{R} \to \mathbb{R}$ over a discrete set plays an important role in many applications. Examples can be found in, among others, investment problems with discrete choices such as infrastructure projects, venture capital, and private equity deals (Ahmed and Atamtürk [2]), project selection (Klastorin [21], Mehrez and Sinuany-Stern [27], Weingartner [36]), competitive facility location (Berman and Krass [4], Aboolian et al. [1], Ljubić and Moreno [25]), and combinatorial auctions (Feige et al. [11], Lehmann et al. [23]). In such utility-maximization problems, the function f typically models the decision maker's risk-averse attitude. When some of the coefficients of f are subject to uncertainty, the tools of stochastic optimization are used to maximize it in expectation over a typically discrete set of scenarios.

Together with his coauthors, Shabbir Ahmed made substantial contributions in advancing the state-of-the-art of mathematical programming methods for solving this challenging set of problems (Ahmed and Atamtürk [2], Yu and Ahmed [38, 39]). In this article, we study optimization problems in which the expected value of the function f composed with a set union operator is maximized over a discrete feasible region. The expected value of the objective function is computed w.r.t. a set of coefficients taking values from a discrete set of scenarios. In this way, we generalize the family of problems addressed in the three aforementioned articles.

1.1 The problem introduced in Ahmed and Atamtürk [2]

Let N be a ground set of n items and M a finite set of m scenarios, each occurring with a strictly positive probability π_i, $i \in M$, and let $a_{ij} \in \mathbb{R}_+$, $i \in M$, $j \in N$, and

---

[1] We call a function $f : \mathbb{R} \to \mathbb{R}$ increasing if $f(x) \geq f(y)$ whenever $x \geq y$ and strictly increasing if $f(x) > f(y)$ whenever $x > y$.

$d_i \in \mathbb{R}_+$, $i \in M$, be nonnegative real numbers. Letting the variable $x_j \in \{0, 1\}^n$ be equal to 1 if and only if item $j \in N$ is chosen, the problem tackled in Ahmed and Atamtürk [2] can be cast as the following Mixed Integer Nonlinear Program (MINLP):

$$\max_{x \in \{0,1\}^n} \left\{ \sum_{i \in M} \pi_i \, f\left(\sum_{j \in N} a_{ij} \, x_j + d_i \right) : x \in X \right\}. \tag{1}$$

Here, X is a polyhedron encapsulating the constraints of the problem and $X \cap \{0, 1\}^n$ is the discrete set of feasible solutions. From a stochastic-optimization perspective, Problem (1) asks for maximizing

$$\mathbb{E}_\xi \left[f\left(\sum_{j \in N} a_j(\xi) \, x_j + d(\xi) \right) \right]$$

subject to $x \in X \cap \{0, 1\}^n$, where ξ is a random variable and the discrete realizations of $a_j(\xi)$, $j \in N$, and $d(\xi)$ are, respectively, $a_{1j}, \ldots, a_{mj}$ and $d_1, \ldots, d_m$, and occur with strictly positive probabilities $\pi_1, \ldots, \pi_m$.

The objective function of Problem (1) can be interpreted as a linear combination of m set functions $h_i : 2^N \to \mathbb{R}_+$ with weights π_i, $i \in M$. For each $i \in M$, h_i is defined as the composition of f with an affine set function $q_i : 2^N \to \mathbb{R}$ that maps subsets of N into the reals. Formally:

$$h_i(S) := f\bigg(\underbrace{\sum_{j \in S} a_{ij} + d_i}_{q_i(S)} \bigg), \quad i \in M, S \subseteq N.$$

It is not difficult to see that the function h_i, $i \in M$, is submodular, see Ahmed and Atamtürk [2], Yu and Ahmed [39], and so is the whole objective function of the problem. Thanks to this, Problem (1) can be reformulated as a Mixed Integer Linear Program (MILP) by exploiting a reformulation of Nemhauser et al. [29] featuring an exponential number of linear inequalities, one per subset S of the ground set N. Ahmed and Atamtürk [2] showed how to tighten such inequalities by a sequence-independent sequential lifting procedure. In a follow-up article, Yu and Ahmed [39] further improved the strength of such inequalities for the case where the set of discrete choices X is subject to a single knapsack constraint. Further results on the nature of the exact lifting function adopted in Ahmed and Atamtürk [2] are reported in Shi et al. [33].

The specific function f considered in Ahmed and Atamtürk [2] and Yu and Ahmed [39] is the negative exponential function

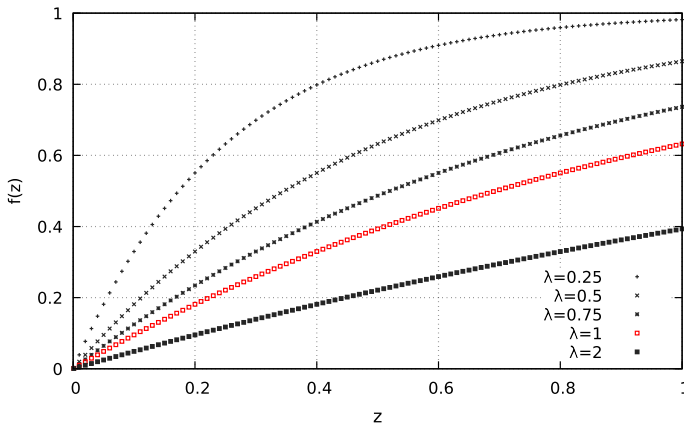$$f(z) = 1 - \exp\left(-\frac{z}{\lambda} \right), \tag{f}$$

Fig. 1 The utility function $f(z) = 1 - \exp(-\frac{z}{\lambda})$ graphed for different values of λ.

which is frequently used to model the behavior of risk-averse decision makers. Here, $\lambda > 0$ represents the risk-tolerance parameter and, the larger the value of λ and, thus, the closer f is to being linear, the less risk-averse the decision maker is. An illustration of such a function for different values of λ is provided in Fig. 1.

1.2 A generalization of Problem (1)

In this work, we study a generalization of Problem (1) in which, besides the set of items N, we are given an additional ground set $\hat{N}$ of $\hat{n}$ *metaitems*. We assume that the two ground sets N and $\hat{N}$ are linked by a *covering relationship* modeled by a bipartite graph $G = (\hat{N} \cup N, E)$ where, for each $j \in N$ and $\ell \in \hat{N}$, $\{j, \ell\} \in E$ if and only if item j is covered by metaitem ℓ. For each item $j \in N$, we define by $\hat{N}(j) \subseteq \hat{N}$ the set of metaitems of $\hat{N}$ by which j is covered and, for each $\ell \in \hat{N}$, we define by $N(\ell) \subseteq N$ the set of items that are covered by ℓ.

The key feature of this generalized problem is that, in order to access an item j of N and benefit from its profit, the decision maker has to select at least one of the metaitems $\ell \in \hat{N}$ covering it. An illustration of an instance of the problem is reported in Fig. 2.

After introducing a binary variable $y_\ell \in \{0, 1\}$ for each $\ell \in \hat{N}$, equal to 1 if and only if metaitem ℓ is chosen, the problem can be cast as the following MINLP:

$$\max_{\substack{x \in \{0,1\}^n \\ y \in Y \cap \{0,1\}^{\hat{n}}}} \sum_{i \in M} \pi_i \, f\left(\sum_{j \in N} a_{ij} \, x_j + d_i\right) \tag{2a}$$

$$\text{s.t.} \quad x_j \leq \sum_{\ell \in \hat{N}(j)} y_\ell, \qquad j \in N \tag{2b}$$

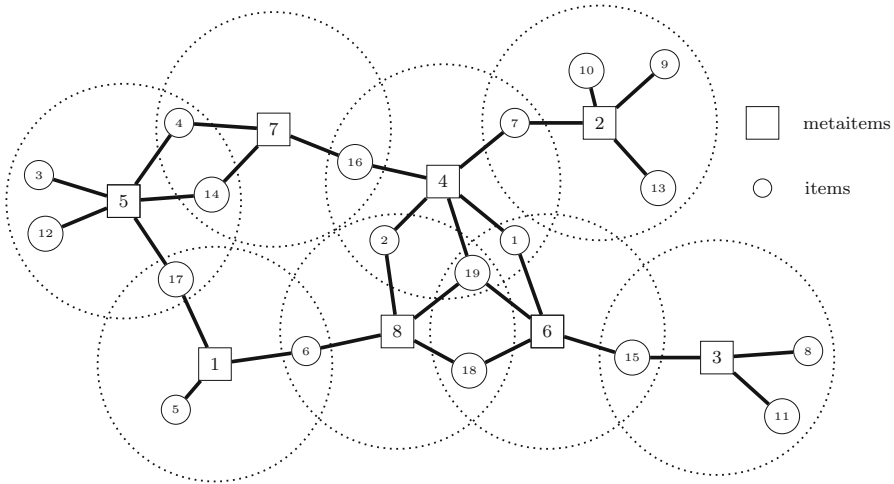$$y_\ell \leq x_j, \qquad \ell \in \hat{N}, \, j \in N(\ell), \tag{2c}$$

Fig. 2 An instance with $n = 19$ items (denoted by circles) and $\hat{n} = 8$ metaitems (denoted by squares). In the instance, the covering relationship between a metaitem $\ell \in \hat{N}$ and an item $j \in N$ (corresponding to the edge $\{\ell, j\}$) is induced by the "covering radius" represented by the dotted circles. This way, the sets $N(\ell)$, $\ell \in \hat{N}$, are determined by the discs centered on the metaitems; for, e.g., $\ell = 1$, we have $N(1) = \{5, 6, 17\}$.

where Y is a polyhedron encapsulating the constraints of the problem and $Y \cap \{0, 1\}^{\hat{n}}$ is the discrete set of feasible choices of metaitems. The parameters π, a, and d have the same meaning as in Problem (1), i.e., M contains the discrete-scenario realizations of the uncertain parameters a and d, and the objective function corresponds to the expected value of f over such realizations. For each $j \in N$, Constraints (2b) and (2c) imply that an item $j \in N$ is taken if and only if at least a metaitem ℓ in $\hat{N}$ is chosen from $\hat{N}(j) \subseteq \hat{N}$. Since the objective function of the problem is strictly increasing, Constraints (2c) are automatically satisfied in any optimal solution and can be dropped. Besides being binary, no further constraints are assumed for x.

To see that Problem (2) generalizes Problem (1), it suffices to observe that the latter is obtained from the former by defining $\hat{N}$ as a copy of N and defining the edge set E as a matching; this way, $x = y$ is imposed and the constraints $x \in X$ are translated into $y \in Y$.

We define the *set-union operator* expressing the covering relationship between $\hat{N}$ and N as

$$N(\hat{S}) := \bigcup_{\ell \in \hat{S}} N(\ell), \qquad \hat{S} \subseteq \hat{N}. \tag{3}$$

With it, we can rewrite the objective function of Problem (2) as a linear combination with weights π_i, $i \in M$, of m set functions $\hat{h}_i : 2^{\hat{N}} \to \mathbb{R}_+$, each of which is defined as the composition of f with a set function $\hat{q}_i : 2^{\hat{N}} \to \mathbb{R}_+$ that maps the choice of metaitems $\hat{S}$ into the value of the items that $\hat{S}$ covers. For each $i \in M$, $\hat{h}_i$ is defined as

$$\hat{h}_i(\hat{S}) := f\Big(\underbrace{\sum_{j \in N(\hat{S})} a_{ij} + d_i}_{\hat{q}_i(\hat{S})} \Big)$$

and $\hat{q}_i$ is defined as

$$\hat{q}_i(\hat{S}) := \sum_{j \in N(\hat{S})} a_{ij} + d_i, \qquad i \in M, \hat{S} \subseteq \hat{N}. \tag{4}$$

The function $\hat{q}_i$ frequently occurs in problems featuring a *set covering* component and it has been heavily studied in the combinatorial optimization literature—see, e.g., Schrijver [32], pag. 768.

Due to d_i and a_{ij} being nonnegative for all $i \in M$, $j \in N$, $\hat{q}_i$ is submodular. Thanks to f being strictly increasing and concave, it is not difficult to see that its composition with $\hat{q}_i$ (i.e., $\hat{h}_i$) is submodular as well.

Problem (2) is $\mathcal{NP}$-hard even in the simple setting where Y is defined by a single cardinality constraint and f is the trivial identity function $f(z) := z$. This is because, letting $m = 1$, $a_{1j} = 1$ for all $j \in N$, $d_1 = 0$, and $Y = \{\hat{S} \subseteq \hat{N} : |\hat{S}| \leq k\}$, Problem (2) admits a solution of value n if and only if there is a feasible solution to the $\mathcal{NP}$-complete problem of covering N with at most k subsets from $\{N(\ell)\}_{\ell \in \hat{N}}$.

1.3 Applications of Problem (2)

Problem (2) arises in different applications. An important one is found in social-network analysis when looking for "key players" or "influencers" to quickly distribute a piece of information through the network (Güney et al. [16], Wu and Küçükyavuz [37], Kempe et al. [20], Karimi et al. [19]). Both of the widely adopted independent cascade (IC) and linear threshold (LT) propagation models (based on probabilistic networks and random sampling) described in Kempe et al. [20] fit into the framework of Problem (2). In IC, the influence propagates on the graph according to a specified edge-propagation probability. In LT, each node of the graph is influenced only if the fraction of the node's neighbors that has become active (weighted by the corresponding influence weight) exceeds a node-specific stochastic threshold. Given a graph $\tilde{G} = (\tilde{V}, \tilde{E})$, m so-called *live-edge graphs* $\tilde{G}_i = (\tilde{V}, \tilde{E}_i)$, $i \in M := \{1, \ldots, m\}$, are created by sampling the edges. For each $i \in M$, π_i is defined as probability of sampling graph i. The sampling process and, accordingly, the live-edge graphs one obtains, depend on the propagation mechanism in use.[2] A bipartite graph $G_i = (N \cup \hat{N}, E_i)$ is then constructed for each $i \in M$. In it, the set of items N is equal to $\tilde{V}$ (therefore, N contains all the users of the social network) and the set of metaitems $\hat{N}$ is equal to the set of influencers in $\tilde{V}$. For each user-influencer pair j, ℓ, E_i contains the edge $\{j, \ell\}$ if and only if there is a path between ℓ and j in $\tilde{G}_i$. Thus, for each $i \in M$ and $\ell \in \hat{N}$, $N_i(\ell) \subseteq N$ is the subset of users that can be reached by influencer ℓ in the live-edge graph $\tilde{G}_i$ due to belonging to the same connected component as ℓ (see also Wu and Küçükyavuz [37] for further details). The goal is then to choose a subset

---

[2] In the IC case, the live-edge graph contains all the edges of G that are declared *live* by flipping a coin of bias equal to the edge-propagation probability—see Theorem 2.2 of Kempe et al. [20]. In the LT case, the live-edge graph contains all the edges that are declared *live* by picking, for each vertex, at most one of its incoming edges with a probability proportional to the corresponding influence weight—see Theorem 2.5 of Kempe et al. [20].

of influencers so as to maximize the expected number of users that can be reached. Modeling such a problem as an instance of Problem (2) requires a slight modification of the set-union operator as well as of the function $\hat{q}_i$ defined in (3)–(4) to make them scenario-dependent:

$$N_i(\hat{S}) := \bigcup_{\ell \in \hat{S}} N_i(\ell), \qquad i \in M, \hat{S} \subseteq \hat{N},$$

$$\hat{q}_i(\hat{S}) := \sum_{j \in N_i(\hat{S})} a_{ij} + d_i, \quad i \in M, \hat{S} \subseteq \hat{N}.$$

For each $i \in M$, $j \in N$, a_{ij} is set to 1. Without loss of generality, all the results presented in this paper can be applied to such a scenario-dependent case as well. To simplify the presentation, we will restrict ourselves to the original definitions given in (3)–(4). We remark that, while in the standard applications of influence maximization referenced above the function f is the identity function, the methods we propose in this paper allow for taking a more general risk-averse utility function into account to measure the total influence in the network.

A second application arises in marketing problems. Let M be the set of products that a marketer can advertise to a set N of potential customers, and let $\hat{N}$ be a set of marketing campaigns, where each campaign $\ell \in \hat{N}$ can only reach a subset $N(\ell)$ of customers. In such a case, the value of a_{ij}, $i \in M$, $j \in N$, corresponds to customer j's demand of product i. Letting π_i be a weight measuring the relevance of a product $i \in M$, the utility function f expresses the decreasing marginal utility of serving an additional unit of customer demand.

A third application arises in a class of stochastic competitive facility location problems with uncertain customer demands. In the deterministic competitive facility location problem (see Ljubić and Moreno [25], Berman and Krass [4], Aboolian et al. [1] and further references therein), a subset of facilities has to be opened subject to a budget constraint in such a way that the customer demand that is satisfied is maximized. In this application, the set $\hat{N}$ corresponds to a set of potential facility locations and the customers are represented by the set N. When adopting, e.g., the multinomial logit model, one can show that the market share function satisfies the properties of function f. In the stochastic setting, the expected market share is then calculated over the set M of possible scenario realizations, each with probability π_i, $i \in M$, where a_{ij} corresponds to the demand of customer $j \in N$ under scenario $i \in M$.

1.4 Contribution and outline of the paper

Throughout the paper, we focus on the case with $n \gg \hat{n}$, i.e., with a number of items orders of magnitude larger than the number of metaitems, which holds in many applications, including those we introduced before. The main contribution of the paper is an exact method for solving Problem (2) based on a double-hypograph decomposition by which the many variables associated with the item set N are projected out. Such a decomposition exploits the structural properties of the utility function f and of the set-union operator. In it, f is linearized via an outer-approximation method, whereas the set-union operator is linearized in two ways: either (i) via a reformulation based

on submodular cuts or (ii) via a Benders decomposition. In particular, we show that, assuming f can be computed in constant time, the inequalities arising from the outer approximation combined with the Benders decomposition can be separated in linear time even for fractional points. We compare the strength of the two resulting mixed integer linear programming reformulations from a theoretical perspective, and also show how to extend them to the case where the utility function is not necessarily increasing. Lastly, we embed the reformulations in two branch-and-cut algorithms, the most efficient of which, according to our computational experiments, allows for solving to optimality instances with up to $n = 20,000$, $\hat{n} = 60$, and $m = 100$ in a short amount of computing time.

The paper is organized as follows. After recalling some background notions on submodularity in Sect. 2, we introduce our decomposition approach and our two reformulations in Sect. 3. The strength of the inequalities of these two reformulations is compared in Sect. 4. Section 5 presents an extension of our method to the case where the utility function is not necessarily increasing. In Sect. 6, we demonstrate via computational experiments the advantages offered by the algorithms we propose. Concluding remarks are reported in Sect. 7. In the Appendix, we also investigate the construction of worst-case instances for the case where Problem (2) is approximately solved via the classical greedy algorithm proposed in Nemhauser et al. [29] for maximizing an increasing submodular function subject to a k-cardinality constraint.

2 Preliminaries

In this section, we briefly summarize the basic terminology and notions used in optimization problems involving submodular functions as well as the key result of Nemhauser et al. [29] that leads to a direct MILP reformulation of Problem (2).

Let $h : 2^N \to \mathbb{R}$ be a generic set function defined for some ground set N. The associated set function $\varrho_j^h(S) := h(S \cup \{j\}) - h(S)$, $j \in N$, $S \subseteq N$, is called the *marginal contribution* of j with respect to S. We call the function h *increasing* if $\varrho_j^h(S) \geq 0$ for all $j \in N$, $S \subseteq N$ (in the literature, such a function is sometimes called *nondecreasing*), and *submodular* if $\varrho_j^h(S) \geq \varrho_j^h(T)$ for all $S, T \subseteq N$ with $S \subseteq T$ and $j \in N$. Throughout the paper, we report the name of the submodular function to which a marginal contribution is referred as a superscript (as in ϱ_j^h).

Every submodular set function h enjoys the following properties:

Proposition 1 (Nemhauser et al. [29]) *If h is submodular, then:*

$$h(T) \leq h(S) - \sum_{j \in S \setminus T} \varrho_j^h\big(N \setminus \{j\}\big) + \sum_{j \in T \setminus S} \varrho_j^h(S), \qquad S, T \subseteq N,$$

$$h(T) \leq h(S) - \sum_{j \in S \setminus T} \varrho_j^h\big(S \setminus \{j\}\big) + \sum_{j \in T \setminus S} \varrho_j^h(\emptyset), \qquad S, T \subseteq N.$$

By applying Proposition 1 to each function $\hat{h}_i$, $i \in M$, featured in Problem (2), we straightforwardly obtain the following MILP reformulation, which we refer to as the *Reformulation based on Submodular Cuts*, or Ref(SC):

$$
\begin{aligned}
\left(\text{Ref(SC)}\right) \max_{\substack{w \in \mathbb{R}^m \\ y \in Y \cap \{0,1\}^{\hat{n}}}} \quad & \sum_{i \in M} \pi_i \, w_i \\
\text{s.t.} \quad w_i \leq {}& \hat{h}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} \varrho_\ell^{\hat{h}_i}(\hat{S}) \, y_\ell \\
& - \sum_{\ell \in \hat{S}} \varrho_\ell^{\hat{h}_i}\left(\hat{N} \setminus \{\ell\}\right)(1 - y_\ell), \quad \hat{S} \subseteq \hat{N}, i \in M \quad (\text{SC}_1^{\hat{h}}) \\
w_i \leq {}& \hat{h}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} \varrho_\ell^{\hat{h}_i}(\emptyset) \, y_\ell \\
& - \sum_{\ell \in \hat{S}} \varrho_\ell^{\hat{h}_i}\left(\hat{S} \setminus \{\ell\}\right)(1 - y_\ell), \quad \hat{S} \subseteq \hat{N}, i \in M. \quad (\text{SC}_2^{\hat{h}})
\end{aligned}
$$

For the correctness of the reformulation, we refer the reader to Nemhauser et al. [29].

An alternative correct reformulation is obtained by replacing Constraints ($\text{SC}_1^{\hat{h}}$) by

$$
w_i \leq \hat{h}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} \varrho_\ell^{\hat{h}_i}(\hat{S}) \, y_\ell, \qquad \hat{S} \subseteq \hat{N}, i \in M. \qquad (\text{SC}_3^{\hat{h}})
$$

Such constraints are dominated by Constraints ($\text{SC}_1^{\hat{h}}$) if there is an item $j \in N$ which is covered by a single metaitem $\ell \in \hat{N}$ (i.e., such that $|\hat{N}(j)| = 1$) and $a_{ij} > 0$, but as strong otherwise. This is because, by definition of $\hat{h}_i$, $\varrho_\ell^{\hat{h}_i}(\hat{N} \setminus \{\ell\}) = 0$ unless there is an item $j \in N$ such that $\hat{N}(j) = \{\ell\}$ and $a_{ij} > 0$. Constraints ($\text{SC}_3^{\hat{h}}$) will be useful in the analysis carried out in Sect. 4.

The main advantage of Ref(SC) is that it is linear and it solely works in the y space. Similar reformulations have been extensively used in many works involving submodular functions. In spite of this, they are known to be rather weak ((Ahmed and Atamtürk [2], Ljubić and Moreno [25], Nemhauser et al. [28]). This is confirmed by the computational experiments we report in Sect. 6.

3 Decomposition methods and MILP reformulations for Problem (2)

In this section, we propose a decomposition approach for Problem (2) that relies on the structural properties of f as well as on those of the set-union operator underlying $\hat{q}_i$, $i \in M$. We then combine this decomposition with an outer approximation of f (see Duran and Grossmann [9], Fletcher and Leyffer [13] for general references on outer approximations) and a linear approximation of the hypograph of $\hat{q}_i$, $i \in M$, based

on either submodular cuts or Benders cuts. This results in two MILP reformulations that only work in the y space. As we will see, both reformulations are solved much more efficiently by a branch-and-cut method than Ref(SC).

3.1 Double-hypograph decomposition

The key idea of our decomposition approach is reformulating Problem (2) in such a way that the function $\hat{h}_i$, $i \in M$, is decomposed into its two parts f and $\hat{q}_i$.

Letting y be the characteristic vector of $\hat{S} \subseteq \hat{N}$, we rewrite the set function $\hat{q}_i(\hat{S})$, $i \in M$, as the following function $\hat{\mathcal{Q}}_i(y) : \{0, 1\}^{\hat{n}} \to \mathbb{Z}_+$:

$$
\hat{\mathcal{Q}}_i(y) := \max_{x \in \{0,1\}^n} \left\{ \sum_{j \in N} a_{ij} \, x_j + d_i : \quad x_j \leq \sum_{\ell \in \hat{N}(j)} y_\ell, \quad j \in N \right\}, \quad i \in M. \tag{VF}
$$

$\hat{\mathcal{Q}}_i(y)$ can be interpreted as the value function of the subproblem of computing $\hat{q}_i(\hat{S})$ for a subset of metaitems $\hat{S} \subseteq \hat{N}$ with incidence vector $y \in Y \cap \{0, 1\}^{\hat{n}}$. Such a function is studied, e.g., by Cordeau et al. [6] in the context of maximal-covering location problems.

The problem of computing $\hat{\mathcal{Q}}_i(y)$ can be solved in linear time $O(|E|)$ by letting $x_j^* := 1$ for each $j \in N$ with $\sum_{\ell \in \hat{N}(j)} y_\ell \geq 1$ and letting $x_j^* := 0$ otherwise. Moreover, as the problem only features upper bounds on the x variables (equal to $\min\{1, \sum_{\ell \in \hat{N}(j)} y_\ell\}$ for each $j \in N$), its Linear Programming (LP) relaxation is integer. This property will be useful later.

Let $\eta_i, w_i, i \in M$ be two auxiliary variables. We introduce the following *double hypograph* decomposition in which two hypograph reformulations are applied in sequence:

$$
\max_{\substack{w, \eta \in \mathbb{R}^m \\ y \in Y \cap \{0,1\}^{\hat{n}}}} \sum_{i \in M} \pi_i \, w_i
$$

$$
\text{s.t.} \quad w_i \leq f(\eta_i), \qquad i \in M \tag{HYPO$_1$}
$$

$$
\eta_i \leq \hat{\mathcal{Q}}_i(y), \qquad i \in M. \tag{HYPO$_2$}
$$

This decomposition is correct as, in any optimal solution, $w_i = f(\eta_i)$ holds due to $\pi_i > 0$, $i \in M$, and $\eta_i = \hat{\mathcal{Q}}_i(y)$, $i \in M$, holds due to f being strictly increasing.

Starting from this decomposition, in the following subsections we propose two MILP reformulations for Problem (2) that belong to the (w, y) space and feature $O(\hat{n} + m)$ variables. Both reformulations are obtained by projecting out the n decision variables x associated with the items $j \in N$ as well as the m auxiliary variables $\eta_i, i \in M$.

3.2 Projecting out the η variables via an outer approximation

The auxiliary variable $\eta_i, i \in M$, can be projected out via Fourier-Motzkin elimination after applying an outer-approximation technique to Constraints (HYPO$_1$).

Proposition 2 *Constraints (HYPO$_1$)–(HYPO$_2$) are equivalent to*

$$w_i \le f(p) - f'(p)\, p + f'(p)\, \hat{\mathcal{Q}}_i(y), \qquad i \in M, p \in \left[0, \sum_{j \in N} a_{ij} + d_i \right]. \quad \text{(OA)}$$

Proof Since f is concave, for each $i \in M$ the set of pairs $(w_i, \eta_i) \in \mathbb{R}^2$ that satisfy Constraint (HYPO$_1$) forms a convex set. By means of an outer-approximation technique, we can therefore restate Constraints (HYPO$_1$) as:

$$w_i \le \big(f(p) - f'(p)\, p\big) + f'(p)\, \eta_i, \qquad i \in M, p \in \left[0, \sum_{j \in N} a_{ij} + d_i \right], \quad (5)$$

where f' is the first derivative of f and $[0, \sum_{j \in N} a_{ij} + d_i]$ is a superset of the set of values that η_i (and, therefore, $\hat{\mathcal{Q}}_i$ and $\hat{q}_i$) can take. As $f' > 0$ due to f being strictly increasing, we can combine Constraints (5) and (HYPO$_2$) using Fourier-Motzkin elimination to project out the η_i variables. This results in Constraints (OA). □

In the following two subsections, we discuss two ways of obtaining MILP reformulations of Problem (2) starting from Constraints (OA). The first one exploits the submodularity of $\hat{q}_i$, $i \in M$. The second one relies on the integrality property of the LP relaxation of (VF), and it exploits LP duality in a Benders-cuts fashion. The idea is to combine Constraints (OA) with a finite collection of affine functions yielding a relaxation of the hypograph of, respectively, $\hat{q}_i$ and $\hat{\mathcal{Q}}_i$, $i \in M$, that is tight at every integer point.

3.3 Reformulation based on outer approximation plus submodular cuts: Ref(OA+SC)

As $\hat{q}_i$ is submodular, the following constraints are valid due to Proposition 1:

$$\hat{\mathcal{Q}}_i(y) \le \hat{q}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} \varrho_\ell^{\hat{q}_i}(\hat{S})\, y_\ell - \sum_{\ell \in \hat{S}} \varrho_\ell^{\hat{q}_i}\big(\hat{N} \setminus \{\ell\}\big)(1 - y_\ell), \qquad \hat{S} \subseteq \hat{N}, \quad (\text{SC}_1^{\hat{q}})$$

$$\hat{\mathcal{Q}}_i(y) \le \hat{q}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} \varrho_\ell^{\hat{q}_i}(\emptyset)\, y_\ell - \sum_{\ell \in \hat{S}} \varrho_\ell^{\hat{q}_i}\big(\hat{S} \setminus \{\ell\}\big)(1 - y_\ell), \qquad \hat{S} \subseteq \hat{N}. \quad (\text{SC}_2^{\hat{q}})$$

Combining these constraints with Constraints (OA) of Proposition 2, we obtain the following constraints for each choice of $\hat{S} \subseteq \hat{N}$, $i \in M$, and $p \in [0, \sum_{j \in N} a_{ij} + d_i]$:

$$w_i \le f(p) - f'(p)\, p + f'(p)\, \hat{q}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} f'(p)\, \varrho_\ell^{\hat{q}_i}(\hat{S})\, y_\ell$$

$$- \sum_{\ell \in \hat{S}} f'(p)\, \varrho_\ell^{\hat{q}_i}\big(\hat{N} \setminus \{\ell\}\big)(1 - y_\ell),$$

$$w_i \le f(p) - f'(p)\, p + f'(p)\, \hat{q}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} f'(p)\, \varrho_\ell^{\hat{q}_i}(\emptyset)\, y_\ell$$

$$- \sum_{\ell \in \hat{S}} f'(p)\, \varrho_\ell^{\hat{q}_i}\big(\hat{S} \setminus \{\ell\}\big)\, (1 - y_\ell).$$

By restricting ourselves to $p = \hat{q}_i(\hat{S})$ (which suffices to obtain a correct reformulation—see Proposition 3), we obtain the following MILP reformulation of Problem (2), to which we refer as the *Reformulation based on Outer Approximation plus Submodular Cuts*, or Ref(OA+SC):

$$\big(\text{Ref(OA+SC)}\big) \quad \max_{\substack{w \in \mathbb{R}^m \\ y \in Y \cap \{0,1\}^{\hat{n}}}} \quad \sum_{i \in M} \pi_i\, w_i$$

$$\text{s.t.} \quad w_i \le f\big(\hat{q}_i(\hat{S})\big) + \sum_{\ell \in \hat{N} \setminus \hat{S}} f'\big(\hat{q}_i(\hat{S})\big)\, \varrho_\ell^{\hat{q}_i}(\hat{S})\, y_\ell$$

$$- \sum_{\ell \in \hat{S}} f'\big(\hat{q}_i(\hat{S})\big)\, \varrho_\ell^{\hat{q}_i}\big(\hat{N} \setminus \{\ell\}\big)\, (1 - y_\ell), \qquad \hat{S} \subseteq \hat{N}, i \in M \quad \text{(OA+SC}_1\text{)}$$

$$w_i \le f\big(\hat{q}_i(\hat{S})\big) + \sum_{\ell \in \hat{N} \setminus \hat{S}} f'\big(\hat{q}_i(\hat{S})\big)\, \varrho_\ell^{\hat{q}_i}(\emptyset)\, y_\ell$$

$$- \sum_{\ell \in \hat{S}} f'\big(\hat{q}_i(\hat{S})\big)\, \varrho_\ell^{\hat{q}_i}\big(\hat{S} \setminus \{\ell\}\big)\, (1 - y_\ell), \qquad \hat{S} \subseteq \hat{N}, i \in M. \quad \text{(OA+SC}_2\text{)}$$

Proposition 3 *Ref(OA+SC) is valid.*

Proof Let $y^* \in Y \cap \{0,1\}^{\hat{n}}$ be a feasible solution to Problem (2). Letting $\hat{S}^* := \{\ell \in \hat{N} : y_j^* = 1\}$ for each $i \in M$, the pair of Constraints (OA+SC$_1$) and (OA+SC$_2$) corresponding to $\hat{S} = \hat{S}^*$ boil down to $w_i \le f(\hat{q}_i(\hat{S}^*))$, thereby guaranteeing (due to $\pi_i > 0, i \in M$) that $\sum_{i \in M} \pi_i w_i = \sum_{i \in M} \pi_i f(\hat{q}_i(\hat{S}^*))$ holds in any optimal solution. The latter coincides with the objective function of Problem (2). $\square$

When solving the Ref(OA+SC) with a branch-and-cut method, to obtain a correct algorithm it suffices to check for the existence of violated constraints among (OA+SC$_1$) and (OA+SC$_2$) after a solution (w^*, y^*) with y^* integer is found (thus carrying out a so-called *lazy cut separation*). Such solutions can be separated very efficiently:

Proposition 4 *Assume that f and f' can be computed in constant time. Given a pair (w^*, y^*) with $w \in \mathbb{R}^m$ and $y^* \in \{0,1\}^{\hat{n}}$, the separation problem asking for a constraint among (OA+SC$_1$)–(OA+SC$_2$) that is violated by (w^*, y^*) can be solved in linear time $O(|E|)$ for each $i \in M$.*

Proof As observed in the proof of Proposition 3, Constraints (OA+SC$_1$)–(OA+SC$_2$) impose $w_i \le f(\hat{q}_i(\hat{S}))$ for every set $\hat{S} \subseteq \hat{N}$. Let $\hat{S}^* := \{\ell \in \hat{N} : y_\ell^* = 1\}$. If (w^*, y^*) is

infeasible, there is some $i \in M$ such that the two constraints in (OA+SC$_1$)–(OA+SC$_2$) corresponding to $i \in M$ and $\hat{S} = \hat{S}^*$ are violated. To solve the separation problem, we only need to compute the coefficients of such constraints. Since $\hat{q}_i$ can be computed in linear time by construction and f and f' can be computed in constant time by assumption, all these coefficients can be computed in linear time $O(|E|)$. Moreover, the coefficients $\varrho_{\ell}^{\hat{q}_i}(\hat{N} \setminus \{\ell\})$, $\ell \in \hat{S}$, and $\varrho_{\ell}^{\hat{q}_i}(\emptyset)$, $\ell \in \hat{N} \setminus \hat{S}$, can be precomputed since they do not depend on $\hat{S}^*$.   □

3.4 Reformulation based on outer approximation plus Benders cuts: Ref(OA+BC)

Let us focus on the subproblem underlying the value function $\hat{\mathcal{Q}}_i$, $i \in M$, defined in (VF). Assuming $y \in Y \cap \{0, 1\}^{\hat{n}}$, every optimal solution to the LP relaxation of the subproblem of computing $\hat{\mathcal{Q}}_i$ in which $x \in \{0, 1\}^n$ is replaced by $x \in [0, 1]^n$ can be assumed to be binary without loss of generality. As the subproblem is bounded, we can consider its LP dual. Let δ and σ be the dual variables associated with, respectively, the constraints linking the x and the y variables and the unit upper-bound constraints on x. Let

$$\mathcal{P}_i := \left\{ (\delta, \sigma) \in \mathbb{R}_+^{n+n} : \ \delta_j + \sigma_j \geq a_{ij}, \ \ j \in N \right\}, \quad i \in M,$$

be the polyhedron of the dual feasible solutions to (VF) (for simplicity, we drop the index i when referring to (δ, σ)). By LP duality, the following holds:

$$\hat{\mathcal{Q}}_i(y) = d_i + \min \left\{ \sum_{j \in N} \sigma_j + \sum_{j \in N} \delta_j \left(\sum_{\ell \in \hat{N}(j)} y_\ell \right) : (\delta, \sigma) \in \mathcal{P}_i \right\}. \tag{BS}$$

In the Benders decomposition literature, (BS) is called the *Benders subproblem*. Letting $\mathcal{P}_i^e$, $i \in M$, denote the set of extreme points of $\mathcal{P}_i$, we derive the following *Benders cuts*:

$$\hat{\mathcal{Q}}_i(y) \leq d_i + \sum_{j \in N} \tilde{\sigma}_j + \sum_{j \in N} \tilde{\delta}_j \left(\sum_{\ell \in \hat{N}(j)} y_\ell \right), \quad i \in M, (\tilde{\delta}, \tilde{\sigma}) \in \mathcal{P}_i^e. \tag{BC}$$

Thanks to Proposition 2 and by restricting ourselves to $p = \hat{\mathcal{Q}}_i(\hat{y})$ for all the not necessarily continuous $\hat{y} \in Y$ (this suffices to obtain a correct reformulation—see Proposition 5), we obtain the following alternative MILP reformulation of Problem (2), which we refer to as the *Reformulation based on Outer Approximation plus Benders Cuts*, or Ref(OA+BC):

$$\big(\text{Ref(OA+BC)}\big) \quad \max_{\substack{w \in \mathbb{R}^m \\ y \in Y \cap \{0,1\}^{\hat{n}}}} \sum_{i \in M} \pi_i \, w_i$$

$$\text{s.t.} \quad w_i \leq f(\hat{\mathcal{Q}}_i(\hat{y})) - f'(\hat{\mathcal{Q}}_i(\hat{y})) \, \hat{\mathcal{Q}}_i(\hat{y})$$

$$+ f'(\hat{\mathcal{Q}}_i(\hat{y})) \, d_i + \sum_{j \in N} f'(\hat{\mathcal{Q}}_i(\hat{y})) \, \tilde{\sigma}_j$$

$$+ \sum_{j \in N} f'(\hat{\mathcal{Q}}_i(\hat{y})) \, \tilde{\delta}_j \bigg(\sum_{\ell \in \hat{N}(j)} y_\ell \bigg), \quad i \in M, (\tilde{\delta}, \tilde{\sigma}) \in \mathcal{P}_i^e, \hat{y} \in Y. \quad \text{(OA+BC)}$$

For all $i \in M$ and $(\tilde{\delta}, \tilde{\sigma}) \in \mathcal{P}_i^e$, if $\hat{y}$ is integer, (OA+BC) coincides with the constraint

$$w_i \leq f(\hat{q}_i(\hat{S})) - f'(\hat{q}_i(\hat{S})) \, \hat{q}_i(\hat{S}) + f'(\hat{q}_i(\hat{S})) \, d_i + \sum_{j \in N} f'(\hat{q}_i(\hat{S})) \, \tilde{\sigma}_j$$

$$+ \sum_{j \in N} f'(\hat{q}_i(\hat{S})) \, \tilde{\delta}_j \bigg(\sum_{\ell \in \hat{N}(j)} y_\ell \bigg),$$

where $\hat{S} := \{\ell \in \hat{N} : \hat{y}_\ell = 1\}$. Notice that, even if we restrict ourselves to $\hat{y} \in Y \cap \{0,1\}^{\hat{n}}$ (i.e., to $\hat{S} \in \hat{N}$), we have a valid formulation. This is shown in the following proposition:

Proposition 5 *Ref(OA+BC) is valid even if we restrict $\hat{y}$ to belong to $Y \cap \{0,1\}^{\hat{n}}$.*

Proof Let $y^* \in Y$ be an optimal integer solution to Problem (2). Pick an optimal pair $(\sigma^*, \delta^*) \in \mathcal{P}_i^e$ satisfying $d_i + \sum_{j \in N} \sigma_j^* + \sum_{j \in N} \delta_j^* \big(\sum_{\ell \in \hat{N}(j)} y_\ell^* \big) = \hat{\mathcal{Q}}_i(y^*)$. For each $i \in M$, it is easy to verify that the instance of Constraints (OA+BC) corresponding to $\hat{y} = y^*$ and $(\tilde{\sigma}, \tilde{\delta}) = (\sigma^*, \delta^*)$ boils down to $w_i \leq f(\hat{\mathcal{Q}}_i(y^*))$. In particular, due to $\pi_i > 0, i \in M$, this guarantees that $\sum_{i \in M} \pi_i w_i = \sum_{i \in M} \pi_i f(\hat{\mathcal{Q}}_i(y^*))$. $\square$

For a given (not necessarily integer) $y^* \in [0,1]^{\hat{n}}$, optimal solutions to (BS) can be computed in closed form according to the following lemma.

Lemma 1 *A pair $(\tilde{\delta}, \tilde{\sigma}) \in \mathbb{R}^{n+n}$ is an optimal solution to (BS) if and only if it satisfies the following conditions:*

$$\begin{array}{lll} \tilde{\delta}_j = a_{ij}, & \tilde{\sigma}_j = 0, & j \in N : \sum_{\ell \in \hat{N}(j)} y_\ell < 1 \\ \tilde{\delta}_j = 0, & \tilde{\sigma}_j = a_{ij}, & j \in N : \sum_{\ell \in \hat{N}(j)} y_\ell > 1 \\ \tilde{\delta}_j = \gamma_j a_{ij}, & \tilde{\sigma}_j = (1-\gamma_j)a_{ij}, \text{ with } \gamma_j \in [0,1], & j \in N : \sum_{\ell \in \hat{N}(j)} y_\ell = 1. \end{array}$$

Proof Problem (BS) decomposes into n subproblems, one per item $j \in N$. Each subproblem asks for a value for δ_j and $\sigma_j \delta_j + \sigma_j \geq a_{ij}$ at minimum cost $(\sum_{\ell \in \hat{N}(j)} y_\ell)\delta_j + \sigma_j$. We refer to it as $(\tilde{\delta}, \tilde{\sigma})$. If $\sum_{\ell \in \hat{N}(j)} y_\ell < 1$, the coefficient of δ_j is smaller than the coefficient of σ_j and, thus, we have $\tilde{\delta}_j = a_{ij}$ and $\tilde{\sigma}_j = 0$. If $\sum_{\ell \in \hat{N}(j)} y_\ell > 1$, the coefficient of δ_j is strictly larger than the coefficient of σ_j and, thus, we have $\tilde{\delta}_j = 0$ and $\tilde{\sigma}_j = a_{ij}$. If $\sum_{\ell \in \hat{N}(j)} y_\ell = 1$, the coefficients of δ_j and σ_j are identical and $(\tilde{\sigma}, \tilde{\delta})$ with $\tilde{\delta}_j = \gamma_j a_{ij}$ and $\tilde{\sigma}_j = (1-\gamma_j)a_{ij}$ is optimal for every $\gamma_j \in [0,1]$. $\square$

The third case in the lemma arises when the Benders subproblem is (dual) degenerate and admits multiple optimal solutions. Thanks to this, Lemma 1 generalizes the results given in [Cordeau et al. [6], Propositions 1, 2, and 4].

Since any dual pair $(\tilde{\delta}, \tilde{\sigma})$ with $\gamma_j \in (0, 1)$ for some $j \in N$ can be obtained as a convex combination of two other dual pairs $(\tilde{\delta}^1, \tilde{\sigma}^1)$, $(\tilde{\delta}^2, \tilde{\sigma}^2)$ (one corresponding to $\gamma_j = 1$ and one to $\gamma_j = 0$), the Benders cut corresponding to $(\tilde{\delta}, \tilde{\sigma})$ is implied by the Benders cuts derived from $(\tilde{\delta}^1, \tilde{\sigma}^1)$ and $(\tilde{\delta}^2, \tilde{\sigma}^2)$. Thus, in Lemma 1 we can w.l.o.g. restrict ourselves to $\gamma_j \in \{0, 1\}$, $j \in N$.

Thanks to Lemma 1, the following holds:

Proposition 6 *Assume that f and f' can be computed in constant time. Given a pair (w^*, y^*) with $w^* \in \mathbb{R}^m$ and $y^* \in [0, 1]^{\hat{n}}$ (not necessarily integer), the separation problem asking for a constraint among (OA+BC) that is violated by (w^*, y^*) can be solved in linear time $O(|E|)$ for each $i \in M$.*

Proof Due to Proposition 5, (w^*, y^*) is infeasible for a given $i \in M$ if and only if $w_i^* > f(\hat{\mathcal{Q}}_i(y^*))$, which implies that w_i^* is strictly larger than the outer approximation of f at $\hat{\mathcal{Q}}_i(y^*)$. Thanks to Lemma 1, the value of the dual multipliers (δ^*, σ^*) satisfying $d_i + \sum_{j \in N} \sigma_j^* + \sum_{j \in N} \delta_j^* \left(\sum_{\ell \in \hat{N}(j)} y_\ell^* \right) = \hat{\mathcal{Q}}(y^*)$ can be computed in linear time $O(|E|)$. Since $\hat{q}_i$, f, and f' can be computed in linear time, the claim follows. □

The main advantage of Ref(OA+BC) over Ref(OA+SC) is that, as we show in the following section, Constraints (OA+BC) subsume Constraints (OA+SC$_1$)–(OA+SC$_2$). Moreover, if we assume that f and f' can be computed in constant time, the separation problem for Constraints (OA+BC) can be solved in linear time even when y^* is fractional, a property which is not enjoyed by Constraints (OA+SC$_1$)–(OA+SC$_2$). From a branch-and-cut perspective, this allows for the efficient separation of Constraints (OA+BC) not just when a solution (w^*, y^*) with an integer y^* is found, but also at each node of the branch-and-bound tree to produce tighter bounds.

4 On the strength of the new inequalities

In this section, we provide a theoretical comparison of the relative strength of the inequalities featured in Ref(SC), Ref(OA+SC), and Ref(OA+BC). Throughout it, we will compare pairs of constraints of different families corresponding to the same subset $\hat{S} \subseteq \hat{N}$ and to the same scenario $i \in M$. When comparing two constraints, we say that one dominates the other one if the coefficients of the former are at least as tight as those of the latter, and that the domination is strict if at least one of the coefficients of the former is strictly tighter.

Let N_1 be the set of items in N that can be covered by a single metaitem $\ell \in \hat{N}$, namely, $N_1 := \{j \in N : |\hat{N}(j)| = 1\}$, and let $\ell(j)$ be such a metaitem. Similarly, for every subset of metaitems $\hat{S} \subseteq \hat{N}$, let $N_1(\hat{S})$ be the set of items in N which can be covered by a single metaitem $\ell \in \hat{S}$, namely, $N_1(\hat{S}) := \{j \in N : |\hat{S} \cap \hat{N}(j)| = 1\}$. The definition of N_1 will be useful in the following analysis.

4.1 Comparison between the inequalities in Ref(SC) and Ref(OA+SC)

In this subsection, we make a pairwise comparison between ($SC_1^{\hat{h}}$) and ($OA+SC_1$) and between ($SC_2^{\hat{h}}$) and ($OA+SC_2$). After that, we show that ($OA+SC_1$) and ($OA+SC_2$) do not dominate one another.

4.1.1 Comparison between ($SC_1^{\hat{h}}$) and ($OA+SC_1$)

Sufficient conditions showing that there are instances in which ($SC_1^{\hat{h}}$) and ($OA+SC_1$) do not dominate one another are reported in the following proposition, as well as a sufficient condition subject to which ($SC_1^{\hat{h}}$) dominates ($OA+SC_1$).

Proposition 7 *For each $i \in M$, $\hat{S} \subseteq \hat{N}$, we have:*

1. *If $\varrho_\ell^{\hat{q}}(\hat{S}) > 0$ for some $\ell \in \hat{N} \setminus \hat{S}$ and $\hat{q}_i(\hat{N} \setminus \{\ell'\}) \geq \hat{q}_i(\hat{S})$ and $\varrho_{\ell'}^{\hat{q}_i}(\hat{N} \setminus \{\ell'\}) > 0$ for some $\ell' \in \hat{S}$, ($SC_1^{\hat{h}}$) and ($OA+SC_1$) do not dominate one another.*
2. *If $\hat{q}_i(\hat{S}) = \hat{q}_i(\hat{N})$, ($SC_1^{\hat{h}}$) dominates ($OA+SC_1$).*

The proof is reported in Appendix A. The appendix also reports a numerical illustration of the no domination result between the two constraints.

4.1.2 Comparison between ($SC_2^{\hat{h}}$) and ($OA+SC_2$)

Sufficient conditions showing that there are instances in which ($SC_2^{\hat{h}}$) and ($OA+SC_2$) do not dominate one another are reported in the following proposition. The proposition also provides sufficient conditions subject to which ($SC_2^{\hat{h}}$) dominates ($OA+SC_2$) and *vice versa*.

Proposition 8 *For each $i \in M$ and $\hat{S} \subseteq \hat{N}$, we have:*

1. *If $\hat{q}_i(\hat{S}) \geq \hat{q}_i(\{\ell\})$ and $\varrho_\ell^{\hat{q}_i}(\emptyset) > 0$ for some $\ell \in \hat{N} \setminus \hat{S}$ and $\varrho_{\ell'}^{\hat{q}}(\hat{S} \setminus \{\ell'\}) > 0$ for some $\ell' \in \hat{S}$, ($SC_2^{\hat{h}}$) and ($OA+SC_2$) do not dominate one another.*
2. *If $\hat{q}_i(\hat{S}) \geq \hat{q}_i(\{\ell\})$ for all $\ell \in \hat{N} \setminus \hat{S}$ and $\varrho_{\ell'}^{\hat{q}}(\hat{S} \setminus \{\ell'\}) = 0$ for all $\ell' \in \hat{S}$, then ($OA+SC_2$) dominates ($SC_2^{\hat{h}}$).*
3. *If $\hat{q}_i(\hat{S}) = \hat{q}_i(\emptyset)$, then ($SC_2^{\hat{h}}$) dominates ($OA+SC_2$).*

The proof is reported in Appendix B, together with a numerical illustration of the no domination result between the two constraints.

4.1.3 Comparison between ($OA+SC_1$) and ($OA+SC_2$)

One may wonder whether the two constraints featured in Ref(OA+SC) dominate one another. We answer the question in the negative:

Remark 1 There is at least an instance of Problem (2) where ($OA+SC_1$) and ($OA+SC_2$) do not dominate one another.

To see this, consider an instance with $N := \{1, \ldots, 6\}$, $\hat{N} := \{1, \ldots, 5\}$, $G = (N \cup \hat{N}, E)$ with $N(1) = \{2\}$, $N(2) = \{2, 3\}$, $N(3) = \{3\}$, $N(4) = \{1, 4\}$, $N(5) = \{2, 3, 5, 6\}$, $m = 1$, $f(z) = 1 - \exp(-\frac{1}{\lambda}z)$ with $\lambda = 10$, and $a = (0.7, 0.6, 0.5, 0.4, 0.3, 0.2)$. Let $\hat{S} = \{1, 2\}$. (OA+SC$_1$) and (OA+SC$_2$) read:

$$w \leq 0.105 - 0.000(1 - y_1) - 0.000(1 - y_2) + 0.000\,y_3 + 0.099\,y_4 + 0.045\,y_5$$
$$w \leq 0.105 - 0.000(1 - y_1) - \mathbf{0.044}(1 - y_2) + \underline{0.045}\,y_3 + 0.099\,y_4 + \underline{0.144}\,y_5.$$

As the latter is tighter at the coefficient of $1 - y_2$ (in boldface) and weaker at the coefficients of y_3 and y_5 (underlined), the two inequalities do not dominate one another. Notice that, as the OA part of the two constraints is identical, the same remark applies to (SC$_1^{\hat{h}}$) and (SC$_2^{\hat{h}}$).

4.2 Comparison between the inequalities in Ref(OA+SC) and Ref(OA+BC)

As the outer-approximation part of the constraints featured in Ref(OA+SC) and Ref(OA+BC) is identical, comparing Ref(OA+SC) to Ref(OA+BC) requires understanding the relationship between the submodular cuts (SC$_1^{\hat{q}}$), (SC$_2^{\hat{q}}$) and the Benders cut (BC).

4.2.1 Comparison between (OA+SC$_1$) and (OA+BC)

To simplify the analysis, we first consider the following additional submodular cut:

$$\hat{Q}_i(y) \leq \hat{q}_i(\hat{S}) + \sum_{\ell \in \hat{N} \setminus \hat{S}} \varrho_\ell^{\hat{q}_i}(\hat{S})\, y_\ell, \quad i \in M, \hat{S} \subseteq \hat{N}. \tag{SC$_3^{\hat{q}}$}$$

Similarly to what we observed in Sect. 2 when comparing (SC$_3^{\hat{h}}$) to (SC$_1^{\hat{h}}$), we have that, for every $i \in M$ and $\hat{S} \subseteq \hat{N}$, (SC$_3^{\hat{q}}$) and (SC$_1^{\hat{q}}$) are equally strong if $\varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\}) = 0$ for all $\ell \in \hat{S}$, i.e., unless $N_1 \neq \emptyset$ and there is an item $j \in N$ which is covered by a single metaitem $\ell \in \hat{N}$ (i.e., such that $|\hat{N}(j)| = 1$) with $a_{ij} > 0$; in such a case, (SC$_1^{\hat{q}}$) strictly dominates (SC$_3^{\hat{q}}$). We establish the following:

Proposition 9 *For each pair of $i \in M$ and $\hat{S} \subseteq \hat{N}$, the submodular cut (SC$_3^{\hat{q}}$) coincides with the Benders cut (BC) obtained with dual multipliers defined, for each $j \in N$, as follows:*

$$\tilde{\delta}_j = \begin{cases} a_{ij} & \text{if } j \notin N(\hat{S}) \\ 0 & \text{otherwise,} \end{cases} \qquad \tilde{\sigma}_j = \begin{cases} a_{ij} & \text{if } j \in N(\hat{S}) \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Proof Due to Lemma 1, $(\tilde{\delta}, \tilde{\sigma})$ is an extreme point of $\mathcal{P}_i^e$ obtained by setting $\gamma_j := 0$ for all $j \in N$. By the choice of multipliers in (6), we have $\sum_{j \in N} \tilde{\sigma}_j = \sum_{j \in N(\hat{S})} a_{ij}$. Thus,

$d_i + \sum_{j \in N} \tilde{\sigma}_j = \hat{q}_i(\hat{S})$ holds. Furthermore, the coefficient of y_ℓ (i.e., $\sum_{j \in N(\ell)} \tilde{\delta}_j$) is equal to $\sum_{j \in N(\ell) \backslash N(\hat{S})} a_{ij} = \varrho_\ell^{\hat{q}_i}(\hat{S})$ for all $\ell \in \hat{N}$. The claim follows. $\square$

As observed, Constraints (SC$_3^{\hat{q}}$) are dominated by (SC$_1^{\hat{q}}$). We show that the latter are subsumed by Constraints (BC).

Proposition 10 *For each $i \in M, \hat{S} \subseteq \hat{N}$, (SC$_1^{\hat{q}}$) is equal to the (BC) obtained by setting for each $j \in N$:*

$$\tilde{\delta}_j = \begin{cases} a_{ij} & if \ j \notin N(\hat{S}) \ or \ j \in N_1 \\ 0 & otherwise, \end{cases} \qquad \tilde{\sigma}_j = \begin{cases} a_{ij} & if \ j \in N(\hat{S}) \setminus N_1 \\ 0 & otherwise \end{cases} \quad (7)$$

(i.e., by setting $\gamma_j = 1$ for all $j \in N_1$, according to Lemma 1). Such a (BC) reads:

$$\hat{\mathcal{Q}}_i(y) \le \hat{q}_i(\hat{S}) + \sum_{\ell \notin \hat{S}} \left(\sum_{j \in N(\ell) \backslash N(\hat{S})} a_{ij} \right) y_\ell - \sum_{\ell \in \hat{S}} \left(\sum_{j \in N(\ell) \cap N_1} a_{ij} \right) (1 - y_\ell). \quad (\text{BC}_1)$$

The proof is reported in Appendix C.

Proposition 10 implies that, for each $i \in M$ and $\hat{S} \subseteq \hat{N}$, Constraints (OA+SC$_1$) coincides with the following (OA+BC) constraints:

$$w_i \le f(\hat{q}_i(\hat{S})) + \sum_{\ell \in \hat{N} \backslash \hat{S}} f'(\hat{q}_i(\hat{S})) \left(\sum_{j \in N(\ell) \backslash N(\hat{S})} a_{ij} \right) y_\ell$$

$$- \sum_{\ell \in \hat{S}} f'(\hat{q}_i(\hat{S})) \left(\sum_{j \in N(\ell) \cap N_1} a_{ij} \right) (1 - y_\ell), \qquad \hat{S} \subseteq \hat{N}, i \in M. \quad (\text{OA+BC}_1)$$

4.2.2 Comparison between (OA+SC$_2$) and (OA+BC)

We start by comparing (SC$_2^{\hat{q}}$) to (BC).

Proposition 11 *For each $i \in M, \hat{S} \subseteq \hat{N}$, (SC$_2^{\hat{q}}$) is dominated by the (BC) obtained by setting for each $j \in N$:*

$$\tilde{\delta}_j = \begin{cases} a_{ij} & if \ j \notin N(\hat{S}) \ or \ j \in N_1(\hat{S}) \\ 0 & otherwise, \end{cases} \qquad \tilde{\sigma}_j = \begin{cases} a_{ij} & if \ j \in N(\hat{S}) \setminus N_1(\hat{S}) \\ 0 & otherwise \end{cases} \quad (8)$$

(i.e., by setting $\gamma_j = 1$ for all $j \in N$ according to Lemma 1), and strictly dominated if $\exists \ell \in \hat{N} \setminus \hat{S} : \varrho_\ell^{\hat{q}_i}(\emptyset) > \varrho_\ell^{\hat{q}_i}(\hat{S}) + \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij}$. Such a (BC) reads:

$$\hat{Q}_i(y) \leq d_i + \sum_{j \in N(\hat{S})} a_{ij} + \sum_{\ell \in \hat{N} \setminus \hat{S}} \left(\sum_{j \in N(\ell) \setminus N(\hat{S})} a_{ij} + \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) y_\ell$$
$$- \sum_{\ell \in \hat{S}} \left(\sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) (1 - y_\ell). \tag{BC$_2$}$$

The proof is reported in Appendix D.

Proposition 11 implies that, for each $i \in M$ and $\hat{S} \subseteq \hat{N}$, Constraint (OA+SC$_2$) is dominated by the following (OA+BC) constraint:

$$w_i \leq f\big(\hat{q}_i(\hat{S})\big) + \sum_{\ell \in \hat{N} \setminus \hat{S}} f'\big(\hat{q}_i(\hat{S})\big) \left(\sum_{j \in N(\ell) \setminus N(\hat{S})} a_{ij} + \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) y_\ell$$
$$- \sum_{\ell \in \hat{S}} f'\big(\hat{q}_i(\hat{S})\big) \left(\sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) (1 - y_\ell), \qquad \hat{S} \subseteq \hat{N}, i \in M. \quad (\text{OA+BC}_2)$$

4.2.3 Comparison between (OA+BC$_1$) and (OA+BC$_2$)

One may wonder whether (OA+BC$_1$) and (OA+BC$_2$) dominate one another. We answer the question in the negative:

Remark 2 There is at least an instance of Problem (2) where (OA+BC$_1$) and (OA+BC$_2$) do not dominate one another.

To see this, consider the instance used in Remark 1 and let $\hat{S} = \{1, 2\}$. (OA+BC$_1$) and (OA+BC$_2$) read:

$$w \leq 0.105 - 0.000(1 - y_1) - 0.000(1 - y_2) + 0.000 y_3 + 0.099 y_4 + 0.045 y_5$$
$$w \leq 0.105 - 0.000(1 - y_1) - \mathbf{0.044}(1 - y_2) + \underline{0.045} y_3 + 0.099 y_4 + \underline{0.089} y_5.$$

As the latter is tighter at the coefficient of $1 - y_2$ (in boldface) and weaker at the coefficients of y_3 and y_5 (underlined), the two inequalities do not dominate one another.
□

4.3 Comparison of the inequalities in Ref(SC) and Ref(OA+BC)

Due to Proposition 10, Constraints (OA+SC$_1$) and (OA+BC$_1$) coincide. Thus:

Corollary 1 *For each $i \in M$ and $\hat{S} \subseteq \hat{N}$, ($SC_1^{\hat{h}}$) and (OA+BC$_1$) enjoy the same relationship as ($SC_1^{\hat{h}}$) and (OA+SC$_1$), as reported in Proposition 7.*

A similar result does not hold for ($SC_2^{\hat{h}}$) and (OA+BC$_2$).

Corollary 2 *There are instances where ($SC_2^{\hat{h}}$) and (OA+BC$_2$) do not dominate one another.*
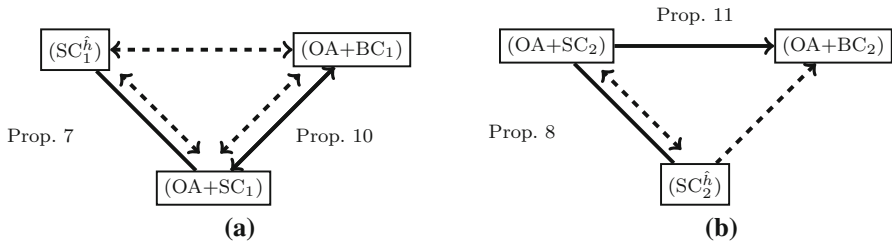
Fig. 3 A visual summary of the main theoretical results on the relative strength of the inequalities studied in this paper.

Proof Consider the instance used in Example 1 and let $\hat{S} = \{1, 2, 4\}$. ($SC_2^{\hat{h}}$) and (OA+BC$_2$) read:

$$w \leq 0.198 - 0(1 - y_1) - 0.041(1 - y_2) + 0.049 y_3 - 0.093(1 - y_4) + 0.148 y_5$$
$$w \leq 0.198 - 0(1 - y_1) - \underline{0.040}(1 - y_2) + \mathbf{0.041} y_3 - \underline{0.088}(1 - y_4) + \mathbf{0.081} y_5.$$

The latter inequality is tighter at the coefficients of y_3 and y_5 (in boldface) but weaker at those of $1 - y_2$ and $1 - y_4$ (underlined). Thus, the two inequalities do not dominate one another. □

4.4 Summary of the theoretical results on the relative strength of the inequalities

A visual summary of the relative strength of the families of inequalities studied in this paper is provided in Fig. 3. An arrow pointing from A to B indicates that A is strictly dominated by B (subject to the indicated conditions). An arrow in both directions indicates that the two inequalities are identical. A connection without directions indicates that domination between the two inequalities may or may not hold based on the conditions we derived. The proposition corresponding to each relationship is reported next to its arrow. Solid lines refer to the general case where no assumptions on f are made. Dashed lines refer to the case where f is an affine function. Such a case is analyzed in the following proposition.

Proposition 12 *If f is affine, (OA+BC$_1$) coincides with ($SC_1^{\hat{h}}$) and (OA+BC$_2$) dominates ($SC_2^{\hat{h}}$) for each $i \in M$, $\hat{S} \subseteq \hat{N}$.*

Proof Let $f(z) := \alpha z + \beta$ for some $\alpha \in \mathbb{R}^+$ and $\beta \in \mathbb{R}$. Then, $f'(\hat{q}_i(\hat{A})) = \alpha$ for all $\hat{A} \subseteq \hat{N}$. Thus, $\varrho_\ell^{\hat{h}_i}(\hat{B}) = \alpha \, \varrho_\ell^{\hat{q}_i}(\hat{B}) = f'(\hat{q}_i(\hat{A})) \, \varrho_\ell^{\hat{q}_i}(\hat{B})$ for all $\ell \in \hat{N}$ and $\hat{A}, \hat{B} \subseteq \hat{N}$. It follows that (OA+SC$_1$) coincides with ($SC_1^{\hat{h}}$) and that (OA+SC$_2$) coincides with ($SC_2^{\hat{h}}$). Thus, the claim follows since (OA+BC$_1$) coincides with (OA+SC$_1$) due to Proposition 10 and (OA+BC$_2$) dominates (OA+SC$_2$) due to Proposition 11. □

To summarize, the theoretical results we obtained allow us to conclude that, given a pair $i \in M$ and $\hat{S} \in \hat{N}$, there are instances where the constraints of Ref(OA+BC) strictly dominate those of Ref(OA+SC), whereas in general there is no domination

between the constraints of Ref(SC) and those of Ref(OA+SC) and Ref(OA+BC). For the special case in which the utility function f is affine, the OA part of the reformulation cancels out and we have Ref(OA+BC) $\subseteq$ Ref(SC) = Ref(OA+SC).

4.5 The case of Problem (1)

We conclude this section by comparing the relative strength of Ref(SC), Ref(OA+SC), and Ref(OA+BC) for the case where Problem (1) of Ahmed and Atamtürk [2], Yu and Ahmed [39] is solved as an instance of Problem (2).

As sketched in Sect. 1, we first reduce Problem (1) to (2) by setting $\hat{N} = N$ and defining the edge set E of the graph G as a matching. This induces a bijection between the sets N and $\hat{N}$ thanks to which each $\ell \in \hat{N}$ is mapped into a single $j = j(\ell) \in N$ and each $j \in N$ is mapped into a single $\ell = \ell(j) \in \hat{N}$. Thus, we have $x_j = y_{\ell(j)}$ and $x_{j(\ell)} = y_\ell$ for each $j \in N$ and $\ell \in \hat{N}$. This implies $\hat{Q}_i(y) = \sum_{j \in N} a_{ij} x_j + d_i$. Letting $S := \{j \in N : x_j = 1\}$, $\hat{Q}_i(y)$ coincides with the function $\hat{q}_i(S)$ of Problem (2) as well as with the function $q_i(S)$ of Problem (1). Thus, $\hat{Q}_i$, $\hat{q}_i$ and q_i are linear functions and the application of either the submodular cuts (SC$_1^{\hat{q}}$)–(SC$_2^{\hat{q}}$) to $\hat{q}_i$ or the Benders cut (BC) to $\hat{Q}(y)$ only results in restating the hypograph constraints (HYPO$_2$) as the inequalities $\eta_i \leq \sum_{j \in N} a_{ij} x_j + d_i, i \in M$. Therefore, Ref(OA+SC) and Ref(OA+BC) both coincide with the following outer-approximation reformulation, to which we refer as Ref(OA):

$$\left(\text{Ref(OA)}\right) \quad \max_{\substack{w \in \mathbb{R}^m \\ x \in X \cap \{0,1\}^n}} \quad \sum_{i \in M} \pi_i \, w_i \tag{9a}$$

$$\text{s.t.} \quad w_i \leq f(p) - f'(p) \, p + f'(p) \left(d_i + \sum_{j \in N} a_{ij} \, x_j \right),$$

$$i \in M, p \in \left[0, \sum_{j \in N} a_{ij} + d_i \right]. \tag{9b}$$

Given a (fractional) point $x^* \in X \cap [0, 1]^n$, the separation problem for Constraints (9b) for each $i \in M$ is solved in linear time by simply setting $p := \sum_{j \in N} a_{ij} x_j^* + d_i$ and computing $f(p)$ and $f'(p)$ (this takes linear time $O(n)$ assuming that f and its derivative can be computed in constant time).

In Yu and Ahmed [39], the authors exploit the submodularity of the objective function of Problem (1) and propose a version of the inequalities obtained in Ahmed and Atamtürk [2] that is stronger when the problem is subject to a single knapsack constraint. In particular, they propose a branch-and-cut algorithm based on the separation of the uplifted version of the submodular constraint (SC$_2^{\hat{h}}$) (see Inequality (9) on page 153 of Yu and Ahmed [39]), as well as on the separation of the standard submodular inequality (SC$_1^{\hat{h}}$).

Let us assume the same setting as Yu and Ahmed [39] where Problem (1) is subject to a knapsack constraint $\sum_{j \in N} \beta_j x_j \leq B$ and the utility function is defined as in (f).

We compare the constraints featured in Ref(OA) to those of Yu and Ahmed [39] with the following two numerical results.

Remark 3 ($SC_1^{\hat{h}}$) and (9b) do not dominate one another.

To see this, consider the following instance taken from Yu and Ahmed [39] where G is the matching graph with $n = \hat{n} = 6$, $m = 1$, $f = -\exp(-ax)$, the knapsack budget is $B = 1$ and the weights and item values are defined as:

$$\beta = (0.3023, 0.1892, 0.3884, 0.1047, 0.5938, 0.6699)$$
$$a = (0.3008, 0.3621, 0.4233, 0.6395, 0.1164, 0.0448).$$

For $S = \{1, 2\}$, the ($SC_1^{\hat{h}}$) constraint featured in the formulation of Yu and Ahmed [39] and Constraint (9b) featured in Ref(OA) read:

$$w \leq -0.515 - 0.053(1 - x_1) - 0.066(1 - x_2) + 0.178x_3 + 0.244x_4 + 0.057x_5 + 0.023x_6$$
$$w \leq -0.515 - \mathbf{0.155}(1 - x_1) - \mathbf{0.186}(1 - x_2) + \underline{0.218}x_3 + \underline{0.330}x_4 + \underline{0.060}x_5 + \underline{0.024}x_6.$$

Since the latter is tighter at the coefficients of $1 - x_1$ and $1 - x_2$ (in boldface) and weaker at the coefficients of x_3, x_4, x_5, x_6 (underlined), the two inequalities do not dominate one another. □

Remark 4 The lifted version of ($SC_2^{\hat{h}}$) proposed by Yu and Ahmed [39] and (9b) do not dominate one another.

To see this, consider the same example as in the proof of Proposition 3. For $S = \{1, 2, 3, 4\}$, the inequality obtained with the lifting procedure in Yu and Ahmed [39] (Inequality (9) in the paper) and Constraint (9b) in Ref(OA) read:

$$w \leq -0.178 - 0.062(1 - x_1) - 0.077(1 - x_2) - 0.093(1 - x_3) - 0.159(1 - x_4)$$
$$+ 0.037x_5 + 0.017x_6$$
$$w \leq -0.178 - \underline{0.053}(1 - x_1) - \underline{0.064}(1 - x_2) - \underline{0.075}(1 - x_3) - \underline{0.113}(1 - x_4)$$
$$+ \mathbf{0.021}x_5 + \mathbf{0.008}x_6.$$

Since the latter is tighter than the previous one at the coefficients of $1 - x_1, 1 - x_2, 1 - x_3, 1 - x_4$ (underlined) and weaker at the coefficients of x_5 and x_6 (in boldface), the two inequalities do not dominate one another. □

Computational experiments comparing the impact of these inequalities on solving instances of Problem (1) are reported in Sect. 6.

5 Generalization of Problem (2) to concave and differentiable utility functions

In this section, we discuss the extension of our methods to the case where the utility function is concave and differentiable, but not necessarily strictly increasing (as
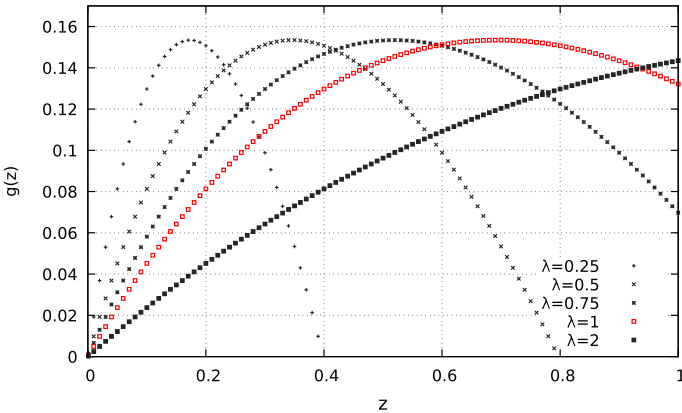
Fig. 4 An example of a concave, differentiable, and not increasing utility function $g(z) = 1 - \exp(-\frac{z}{\lambda}) - \alpha \frac{z}{\lambda}$, with $\alpha = 0.5$, graphed for different values of λ.

assumed up to this point). To distinguish this new utility function from the previous one, we will refer to it as g. Such a case arises in applications where, e.g., the utility function g is composed of two terms: a revenue term which increases with the amount of customers (having a positive decreasing derivative due to economies of scale) and a negative component which accounts for the total amount of money spent for running the service (assuming a unit cost per customer which decreases with the number of customers). In such cases, once the system becomes overloaded and the optimal running capacity is exceeded, g switches from increasing to decreasing. Applications of this type can be found in, e.g., competitive facility location problems where the objective function is the difference between the market capture and the investment cost (see Küçükaydın et al. [22], Lin and Tian [24] and further references therein).

As an example of a utility function of this type, consider the following one

$$g(z) = \left(1 - \exp\left(-\frac{z}{\lambda}\right)\right) - \alpha \frac{z}{\lambda}, \tag{g}$$

where λ is the risk-tolerance parameter and $\alpha > 0$ is an additional parameter establishing a trade-off between the increasing and the decreasing terms. An illustration of this function for different values of λ is provided in Fig. 4. As shown in the figure, larger values of λ lead to the function achieving its maximum for larger values of z. For the case illustrated in the figure (where $\alpha = 0.5$), such a maximum is reached for values of z larger than 1 whenever $\lambda \geq 2$.

When combining a not increasing function such as the one defined in (g) with the set function $\hat{q}_i$, the obtained function $\hat{h}_i := g \circ \hat{q}_i$ (which reads $\hat{h}_i(\hat{S}) = 1 - \exp(-\frac{\hat{q}_i(\hat{S})}{\lambda}) - \alpha \frac{\hat{q}_i(\hat{S})}{\lambda}$ for the case of (g)) is the difference of two submodular functions. As such a function is not submodular in the general case, Ref(SC) and the submodular constraints in Proposition 1 are not valid for Problem (2). This implies that the inequalities introduced in Ahmed and Atamtürk [2] and Yu and Ahmed [39] are not valid either.

To account for this case, we extend the double-hypograph decomposition we proposed in Sect. 3.1 as follows:

$$\max_{\substack{w,\eta\in\mathbb{R}^m \\ y\in Y\cap\{0,1\}^{\hat{n}}}} \sum_{i\in M} \pi_i\, w_i,$$

$$\text{s.t.} \qquad w_i \leq g(\eta_i), \qquad\qquad i \in M \qquad\qquad (\text{HYPO}_1')$$

$$\eta_i \leq \hat{\mathcal{Q}}_i(y), \qquad\qquad i \in M \qquad\qquad (\text{HYPO}_2)$$

$$\eta_i \geq \hat{\mathcal{Q}}_i(y), \qquad\qquad i \in M. \qquad\qquad (\text{EPI})$$

The new epigraph constraints (EPI) are added since $g(\eta_i)$ can be decreasing for a sufficiently large η_i and, thus, imposing $\eta_i \leq \hat{\mathcal{Q}}_i(y)$ with (HYPO$_2$) does not suffice to guarantee $\eta_i = \hat{\mathcal{Q}}_i(y)$ in an optimal solution.

Thanks to a result by Edmonds [10], Constraints (EPI) can be restated as the *polymatroid constraints*

$$\eta_i \geq d_i + \sum_{\ell\in\hat{N}} \vartheta_\ell\, y_\ell, \qquad\qquad \vartheta \in P^{\hat{q}_i}, i \in M,$$

where $P^{\hat{q}_i} := \left\{ \vartheta \in \mathbb{R}^{\hat{n}} : \sum_{\ell\in\hat{S}} \vartheta_\ell \leq \hat{q}_i(\hat{S}) - d_i, \ \hat{S} \subseteq \hat{N} \right\}$ is the *submodular polyhedron* (or *polymatroid*) of the submodular function $\hat{q}_i, i \in M$. These constraints can be separated by maximizing a linear objective function over $P^{\hat{q}_i}$ using Edmonds' (polynomial time) greedy algorithm. Given a pair (η_i^*, y^*) with $\eta_i^* \in \mathbb{R}_+$ and $y^* \in Y\cap[0,1]^{\hat{n}}$, the algorithm sorts the metaitems in $\hat{N}$ in nonincreasing order of the associated y^* values, i.e., such that $y_{\mu(1)}^* \geq y_{\mu(2)}^* \geq \cdots \geq y_{\mu(\hat{n})}^*$ holds, where μ is a suitable permutation of $\hat{N}$. Let $\hat{S}_{\mu(\ell)} := \{\mu(1), \mu(2), \ldots, \mu(\ell)\}$ be the set of the first ℓ metaitems according to new order. By setting

$$\vartheta_{\mu(\ell)} := \hat{q}_i(\hat{S}_{\mu(\ell)}) - \hat{q}_i(\hat{S}_{\mu(\ell-1)}), \qquad \ell = 1, 2, \ldots, \hat{n}, \qquad (10)$$

where $\hat{S}_{\mu(0)} := \emptyset$, the pair (η_i^*, y^*) violates one of the polymatroid constraints if and only if $\sum_{\ell\in\hat{N}} \vartheta_{\mu(\ell)}\, y_\ell^* + d_i > \eta_i^*$. Since ϑ only depends on the permutation μ, the polymatroid constraints can be restated as

$$\eta_i \geq d_i + \sum_{\ell\in\hat{N}} \vartheta_{\mu(\ell)}\, y_{\mu(\ell)}, \qquad \mu \text{ permutation of } \hat{N}, i \in M, \qquad (\text{POLY})$$

where ϑ is defined as in (10) and μ is any permutation of the elements of $\hat{N}$. Due to $\hat{q}_i$ being submodular, Constraints (POLY) are facet defining for the epigraph of $\hat{\mathcal{Q}}_i(y)$, see, e.g., Atamtürk and Narayanan [3].

As before, we apply an outer approximation to Constraints (HYPO$_1'$) and combine them with Constraints (HYPO$_2$) via Fourier-Motzkin elimination. Following

the same derivations used for the case with f strictly increasing, we obtain Constraints (OA+SC$_1$)–(OA+SC$_2$) and (OA+BC$_1$)–(OA+BC$_2$). Since g' can be either 0 or negative, though, such constraints are valid only for a set $\hat{S} \subseteq \hat{N}$ satisfying $g'(\hat{q}_i(\hat{S})) \geq 0$ and for a vector $\hat{y}$ satisfying $g'(\hat{\mathcal{Q}}_i(\hat{y})) \geq 0$, respectively (note that, at any point where g' takes value 0, these constraints just boil down to $w_i \leq UB$, where UB is the maximum value of g). For every vector $\hat{y} \in Y$ satisfying $g'(\hat{\mathcal{Q}}_i(\hat{y})) < 0$, the combination of the outer approximation of Constraints (HYPO$_1'$) with Constraints (EPI) via Fourier-Motzkin elimination leads to

$$w_i \leq g(\hat{\mathcal{Q}}_i(\hat{y})) - g'(\hat{\mathcal{Q}}_i(\hat{y})) + g'(\hat{\mathcal{Q}}_i(\hat{y}))\, d_i + g'(\hat{\mathcal{Q}}_i(\hat{y})) \sum_{\ell \in \hat{N}} \vartheta_{\mu(\ell)}\, y_{\mu(\ell)},$$

$$i \in M, \hat{y} \in Y : g'(\hat{\mathcal{Q}}_i(\hat{y})) < 0, \quad \mu \text{ permutation of } \hat{N}. \qquad \text{(OA+POLY)}$$

We call Ref(OA+SC+POLY) and Ref(OA+BC+POLY) the reformulations featuring (POLY) (imposed only for $\hat{S} \subset \hat{N} : g'(\hat{\mathcal{Q}}_i(\hat{y})) < 0$) and either (OA+SC$_1$)–(OA+SC$_2$) or (OA+BC$_1$)–(OA+BC$_2$) (the former imposed only for $\hat{S} \subset \hat{N} : g'(\hat{q}_i(\hat{S})) \geq 0$ and the latter only for $\hat{y} \in Y : g'(\hat{\mathcal{Q}}_i(\hat{y})) \geq 0$)). Experiments carried out with these two reformulations with the utility function (g) are reported in Sect. 6.

It is worth mentioning that for the case where the utility function is not strictly increasing and one has access to a decomposition of $\hat{h}_i$ (for each scenario $i \in M$) as a difference of two submodular functions, the recently-proposed method of Atamtürk and Narayanan [3] can be an alternative to the approach proposed in this section. Even though such a decomposition always exists, it is not always trivial to find it. Hence, one of the advantages of our method, compared to the one of Atamtürk and Narayanan [3], is its potentially wider applicability as it does not require finding such a decomposition.

6 Computational results

In this section, we assess the computational effectiveness of the MILP reformulations proposed in the paper. We carry out three sets of experiments, considering i) instances of Problem (2) featuring a strictly increasing function f (Sect. 6.1), ii) instances of Problem (2) featuring a not increasing function g (Sect. 6.2), and iii) instances of Problem (1), studied in Ahmed and Atamtürk [2] (Sect. 6.3).

Our experiments are performed on a single node of the IRIDIS 5 cluster at the University of Southampton, equipped with 40 dual 2.0 GHz Intel Skylake processors and 192 GB of DDR4 memory. Our code is written in C/C++ and compiled with gcc 9.2.1, using the -O3 compiler optimization flag. Our branch-and-cut algorithms are implemented in CPLEX 12.9.0 using the CALLABLE LIBRARY

framework, and run in single-threaded mode.[3] We set a time limit of 600 seconds for each run.[4]

6.1 Experiments for Problem (2) with a strictly increasing utility function

We experiment with four branch-and-cut (B&C) algorithms: `SC`, `OA+SC`, `OA+BC`, and `OA+BC+f`. `SC` is based on the standard reformulation Ref(SC) that relies on Constraints (SC$_1^{\hat{h}}$) and (SC$_2^{\hat{h}}$) (see Sect. 2). `OA+SC` is based on the reformulation Ref(OA+SC) that relies on Constraints (OA+SC$_1$) and (OA+SC$_2$) (see Sect. 3.3). `OA+BC` and `OA+BC+f` are based on the reformulation Ref(OA+BC) that relies on Constraints (OA+BC$_1$) and (OA+BC$_2$) (see Sect. 3.4); the separation is carried out only for an integer y^* in the former and for both an integer and a fractional y^* in the latter. The separation is performed with a violation tolerance of 10^{-6}.

6.1.1 Instances

In this set of experiments, we adopt the strictly increasing function (f) (see Sect. 1.1), also adopted in Ahmed and Atamtürk [2], Yu and Ahmed [38]. We consider five values of $\lambda \in \{0.25, 0.5, 0.75, 1, 2\}$ and two values of $m \in \{50, 100\}$ (number of scenarios).

We consider four different sets of constraints to define the feasible region Y of Problem (2): (i) a *k-cardinality constraint* $\sum_{\ell \in \hat{N}} y_\ell \leq k$, used in, e.g., Nemhauser et al. [29]; (ii) a *knapsack constraint* $\sum_{\ell \in \hat{N}} \beta_\ell \, y_\ell \leq B$ with weights $\beta_\ell \geq 0$, $\ell \in \hat{N}$, and budget $B \geq 0$, used in, e.g., Sviridenko [34], Ahmed and Atamtürk [2], Yu and Ahmed [38]; (iii) a set of *partition-matroid constraints* $\sum_{\ell \in \hat{N}_q} y_\ell \leq k_q$, with $1 \leq q \leq p$, according to which $\hat{N}$ is partitioned into p groups $\hat{N}_1, \ldots, \hat{N}_p$, each subject to a cardinality constraint with budget k_p; these constraints are used in, e.g., Friedrich et al. [14], Sviridenko [34]; (iv) a set of *conflict constraints* $y_\ell + y_{\ell'} \leq 1$, $\{\ell, \ell'\} \in \mathcal{E}$, used to model pairwise incompatibilities between metaitems; these constraints are based on a *conflict graph* $\mathcal{G} = (\hat{N}, \mathcal{E})$, where $\mathcal{E}$ is the set of edges representing the conflicts.

Motivated by the fact that Problem (2) subsumes the maximal-coverage location problem with uncertain customer demands and a concave utility function, we generate our instances by extending the procedure proposed by ReVelle et al. [30] and Cordeau et al. [6] for the deterministic version of such a problem. To model the covering relationship between the sets of items N and metaitems $\hat{N}$, N and $\hat{N}$ are embedded in $\mathbb{R}^2$ and the covering relationship is spatially induced by associating a covering radius $\hat{R}$ with each metaitem (see Fig. 2). The coordinates of each item and metaitem in $N \cup \hat{N}$ are chosen uniformly at random from [0, 30].

---

[3] To compute solutions with a high numerical accuracy, we set the parameters `CPX_PARAM_EPRHS` (feasibility tolerance) and `CPX_PARAM_EPGAP` (relative MIP gap tolerance) to 10^{-9}. We separate any constraint that is violated by either a fractional or an integer point via, respectively, the callback function `CPXsetusercutcallbackfunc` and `CPXsetlazyconstraintcallbackfunc`; as recommended by the `CPLEX` user manual, when these functions are used we set the parameters `Preprocessing_Linear` and `MIP_Strategy_CallbackReducedLP` to 0.

[4] The source code of our implementation can be downloaded from https://github.com/fabiofurini/Maximizing-submodular-utility-functions.

For each item $j \in N$, the set $\hat{N}(j)$ contains all the metaitems $\ell \in \hat{N}$ whose Euclidean distance from j is no larger than the covering radius $\hat{R} \in \{2, 4, 5, 6\}$ (smaller values of $\hat{R}$ result in sparser instances). The items are sampled in such a way that each of them is covered by at least a metaitem. For each scenario $i \in M$ and item $j \in N$, we set a_{ij} to zero with probability $r \in \{\frac{1}{4}, \frac{1}{2}\}$ and draw the value of u_{ij} uniformly at random in $\{1, 2, \ldots, 10\}$ with probability $1 - r$. We then introduce the coefficient $\Delta := \sum_{i \in M} \pi_i \sum_{j \in N} u_{ij}$, corresponding to the expected maximum return across all scenarios, and define the normalized value $a_{ij} := u_{ij}/\Delta$. For each choice of n and $\hat{n}$ and for each set of constraints, we generate 5 instances using a different seed for the random-number generator, 4 different values for the radius $\hat{R}$, and 2 different values for the scenario-item probability r.

We consider three main classes of instances: (i) the class k-card, containing 7200 instances subject to a k-cardinality constraint; they are obtained by choosing $n \in \{5000, 10{,}000, 20{,}000\}, \hat{n} \in \{40, 50, 60\}$, and $k \in \{10, 15\}$. (ii) the class kp, containing 3600 instances subject to a knapsack constraint; they are obtained by choosing $n \in \{5000, 10{,}000, 20{,}000\}$ and $\hat{n} \in \{40, 50, 60\}$; β_ℓ and B are defined following the procedure described by Martello et al. [26] for generating strongly correlated instances, which are among the most difficult instances for the knapsack problem.[5] (iii) the class p-matroid, containing 1200 instances subject to partition-matroid constraints; they are obtained by choosing $n \in \{5000, 10{,}000, 20{,}000\}$, $\hat{n} = 60$, $p = 3$, and $k_q = 5$, $q = 1, \ldots, 3$. In all these instances, we set $\pi_i = \frac{1}{m}$ and $d_i = 0$, $i \in M$. For each of these three classes, we generate another class that is also subject to conflict constraints induced by an Erdös-Rény graph $\mathcal{G} = (\hat{N}, \mathcal{E})$ with edge probability 0.01; in the following, we refer to these three classes as k-card & confl, kp & confl, and p-matroid & confl. Overall, the testbed contains six classes and a total of 24,000 instances.

6.1.2 Initialization of the B&C algorithms with a greedy solution

The *greedy algorithm* proposed in Nemhauser et al. [29] is a simple and effective way to obtain feasible solutions to the problem of maximizing a submodular function subject to a k-cardinality constraint. As shown in Nemhauser et al. [29], such solutions are $(1 - \frac{1}{e})$-approximate (where e is the base of the natural logarithm). For the basic variant of Problem (2) with a single k-cardinality constraint, the greedy algorithm starts with an empty set of metaitems ($\hat{S} = \emptyset$), and, at each iteration, adds to $\hat{S}$ the metaitem $\ell \in \arg\max_{\ell \notin \hat{S}} \sum_{i \in M} \pi_i \, \varrho_\ell^{\hat{h}_i}(\hat{S})$ (thereby selecting the metaitem with the largest marginal contribution) until $|\hat{S}| = k$.

The algorithm can be extended to work with all the constraints we introduced by selecting the item of maximum marginal contributions among those that can be added to $\hat{S}$ without violating any constraints. In the partition-matroid case, this algorithm achieves a $\frac{1}{2}$ approximation ratio, see Cornuejols et al. [7]. In the knapsack case, an extension that combines it with a partial enumeration scheme achieves a $1 - \frac{1}{e}$

---

[5] These instances are obtained by setting $\beta_\ell := p_\ell + R/10$, where $p_\ell := \sum_{i \in N} \pi_i \sum_{j \in \hat{N}(\ell)} u_{ij}$, $R := 1000$, and $B := \lfloor 0.5 \sum_{\ell \in \hat{N}} \beta_\ell \rfloor$.

Table 1 Computational performance of the four B&C algorithms within a time limit of 600 secs. The gaps are reported in percentage for the instances not solved to proven optimality.

Instance class	# total	SC		OA+SC		OA+BC		OA+BC+f		greedy
		# opt	gap	# opt	gap	# opt	gap	# opt	gap	gap
k-card	7200	837	7.0	3002	2.6	3821	1.6	**7058**	**0.2**	1.8
k-card & confl	7200	890	7.1	3097	2.8	3923	1.7	**6948**	**0.4**	1.9
kp	3600	2548	4.7	2991	0.3	3060	0.2	**3160**	**0.1**	0.7
kp & confl	3600	2639	5.0	3066	0.4	3165	0.3	**3246**	**0.2**	0.9
p-matroid	1200	98	6.5	398	2.9	398	1.8	**1056**	**0.3**	1.7
p-matroid & confl	1200	102	6.6	400	3.0	398	1.9	**979**	**0.4**	2.1
Total/average	24000	7114	6.1	12954	2.0	14765	1.2	**22447**	**0.3**	1.5

The best results are highlighted in bold

approximation, see Sviridenko [34]. As the problem of maximizing a linear function subject to a set of conflicts is inapproximable to within $\hat{n}^{1-\epsilon}$ for any $\epsilon > 0$ unless $\mathcal{ZPP} = \mathcal{NP}$ Håstad [17], no constant-factor approximation factor is obtained when conflict constraints are imposed.

In all of our experiments, we adopt the greedy algorithm to provide a first feasible solution to our branch-and-cut algorithms. In Appendix E, we illustrate how to construct a set of *worst-case* instances of Problem (2) subject to a k-cardinality constraint on which the greedy algorithm achieves the worst-case approximation factor of $1 - \frac{1}{e}$.

6.1.3 Comparison of the four B&C algorithms

In Table 1, we compare the performance of the four B&C algorithms (SC, OA+SC, OA+BC, and OA+BC+f) for solving Problem (2) with f strictly increasing. The table is vertically divided in six parts. The first part reports the instance class and the total number of instances it contains (# total). The next four parts report, for each of the four exact methods, the total number of instances solved to optimality within the time limit (denoted by # opt) and the average gap for the unsolved instances at the time limit. The gaps are calculated as $\frac{UB-LB}{UB} \cdot 100$, where UB and LB are, respectively, the best upper bound and the incumbent solution value achieved by the corresponding method within the time limit. The last part reports the gap obtained with the greedy algorithm for the instances that are not solved to optimality by OA+BC+f.

When comparing OA+SC to SC, the number of instances solved to optimality by OA+SC is about 80% larger and the average gap is about 3 times smaller. For all classes of instances except for the kp one, OA+SC solves three to four times more instances than SC.

Further performance improvements are obtained with OA+BC, which solves about 1700 more instances than OA+SC and reduces the average gap of OA+SC by about 40%. This is in line with our theoretical results, showing that Constraints (OA+BC$_2$) strictly dominate Constraints (OA+SC$_2$) (see Proposition 11) and that Constraints (OA+BC$_1$) are equivalent to Constraints (OA+SC$_1$) (see Proposition 10).

OA+BC+f achieves a drastic improvement when compared to OA+BC thanks to separating Constraints (OA+BC$_1$)–(OA+BC$_2$) also on fractional points. In particular, OA+BC+f is able to solve 94% of the instances across all classes, obtaining an average gap for the remaining 1553 unsolved instances of only 0.3%. Performance profiles that clearly demonstrate the superiority of OA+BC+f for each class of instances are reported in Appendix G.[6]

We note that our implementation of the greedy algorithm requires on average ≈ 1 sec of computing time. Let us consider the gap of this algorithm for the instances that cannot be solved to optimality by OA+BC+f, as reported in the last column of Table 1 (this gap is obtained by replacing LB by the value of the greedy solution). By comparing the gaps obtained with OA+BC+f to those obtained with the greedy algorithm, we observe that the average gap of the instances that are not solved by OA+BC+f decreases from 1.5% to 0.3%. Overall, Table 1 shows that the quality of the incumbent solutions found by the B&C algorithm OA+BC+f is superior to the quality of the solutions found by the greedy algorithm not only on those which OA+BC+f manages to solve to optimality, but also on those where the algorithm incurs in the time limit.

We now focus on the 7200 instances from the class k-card to identify the main instance features by which the computing times of OA+BC+f are affected. Figure 5 reports the computing times of OA+BC+f after aggregating the instances by the covering radius $\hat{R}$, the value of λ, the number of items n, and the number of metaitems $\hat{n}$.[7] The four box plots in the figure show that, by doubling the number of items, the average computing time doubles. However, by increasing the number of metaitems by 50%, the average computing time grows by a factor of 10. Increasing the length of the covering radius (and, thus, the density of G) also results in slightly larger computing times. In spite of this, very sparse instances obtained for $\hat{R} = 2^8$ are extremely easy to solve for our B&C methods. For larger values of λ, the instances tend to become easier. This can be due to the fact that, the larger λ, the more the function f is close to being linear. Further analysis (not illustrated in the tables for reasons of space) indicates that, by increasing the number of scenarios from 50 to 100, the average computing time doubles; differently, changing the value of k from 10 to 15 does not have a significant impact on the computing time of OA+BC+f.

Table 2 illustrates the most relevant features concerning the execution of OA+BC+f (the best-performing algorithm) on the 22,447 instances it solves to optimality: LP gap (computed as $\frac{Z^{\mathrm{LP}}-Z}{Z^{\mathrm{LP}}} \cdot 100$, where Z^{LP} is the optimal solution value of the LP relaxation at the root node and Z is the optimal solution value of the problem), number of generated inequalities, number of branching nodes, and average computing time.

---

[6] We also experimented with solving the problem directly using the formulation reported in (2) with BARON 20.4 (Sahinidis [31], Tawarmalani and Sahinidis [35]), one of the state-of-the-art solvers for global optimization. Extensive preliminary tests showed that BARON is computationally outperformed by all of the B&C algorithms. It fails to solve many instances with $n = 5,000$ items, and it is unable to solve any instance with $n = 10,000$.

[7] The computing times are reported in logarithmic scale through their quantiles. The lines extending vertically from the boxes indicate the variability outside the upper and lower quantiles. The outliers are plotted as individual points.

[8] These are the only instances with a curvature strictly smaller than one—see Appendix F.
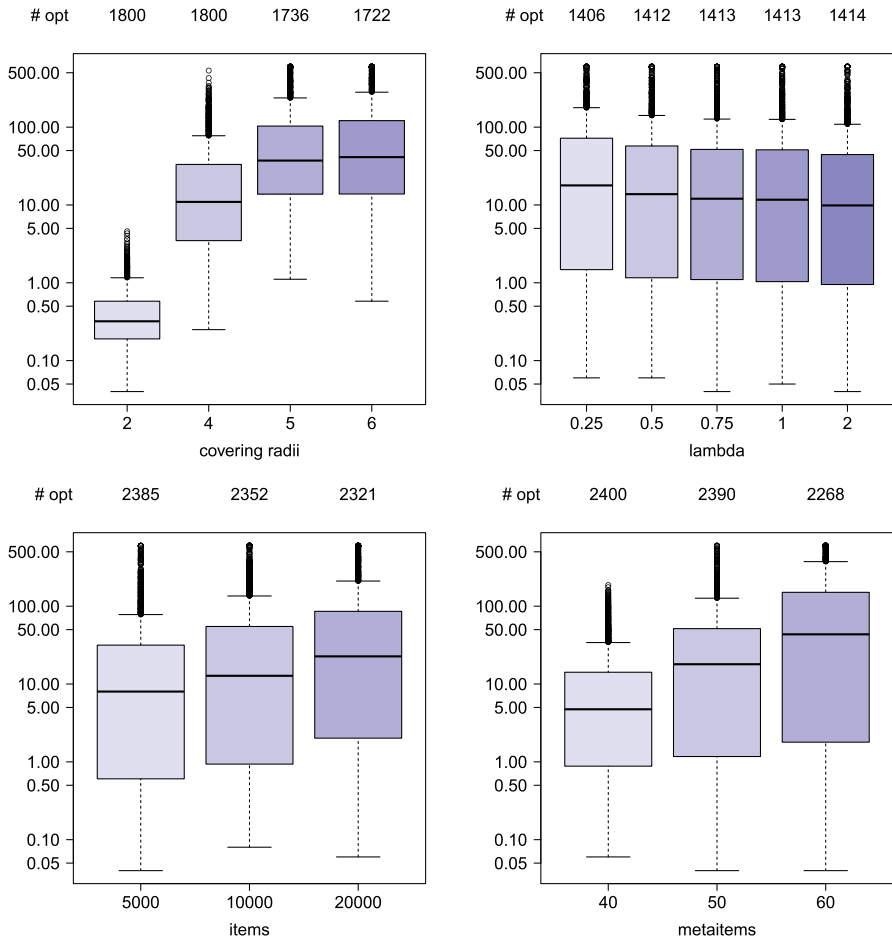
Fig. 5 Computing times (in logarithmic scale) of `OA+BC+f` for the `k-card` instances aggregated by: (*i*) the covering radius $\hat{R}$, (*ii*) the risk-tolerance λ, (*iii*) the number of items n, and (*iv*) the number of metaitems $\hat{n}$. The number of instances solved within the time limit of 600 secs is reported above each boxplot.

Table 2 Detailed results obtained by using the `OA+BC+f` algorithm on the instances solved to proven optimality within a time limit of 600 seconds.

Instance class	LP gap [%]		LP time [s]	avg # of ineq.		# of nodes		time [s]
	avg	max	avg	frac. sol.	int. sol.	avg	max	avg
`k-card`	0.048	0.798	12.4	8038.6	2750.8	5.9	156	46.1
`k-card & confl`	0.066	1.244	12.4	8505.6	2459.1	6.6	119	52.7
`kp`	0.036	0.699	2.4	6703.7	2484.0	250.3	29430	31.5
`kp & confl`	0.047	0.797	2.3	6155.0	2288.3	317.4	16960	38.3
`p-matroid`	0.066	0.650	22.6	14455.3	3384.7	11.9	136	116.1
`p-matroid & confl`	0.081	1.191	21.1	13529.2	2846.7	11.1	78	129.3

The column "LP time [s]" reports the average computing time necessary to solve the LP relaxation at the root node (including the time spent to generate the inequalities). The last column of the table, denoted by "time [s]", reports the total average computing time of OA+BC+f. For what concerns the average and the maximum LP gaps, the table shows that very small LP gaps are obtained across all classes, suggesting that the tight LP gap is one of the main drivers of the excellent performance of OA+BC+f. In particular, the gaps are never larger than $\approx 1.25\%$ and their average values are smaller than $\approx 0.1\%$. The time taken to solve the LP relaxation is quite small and it varies as a function of the class of the instances. For what concerns the average number of Constraints (OA+BC$_1$) and (OA+BC$_2$) generated when separating fractional and integer points, the table show that, as one may expect, many more cuts are generated when fractional solutions are separated. Such an increased number of inequalities is well compensated by the smaller LP gaps. The average number of nodes is small in most of the instances, further evidencing the superior performance of OA+BC+f. The only exceptions are the instances from the kp class, for which several thousands of nodes are explored, albeit the time taken to solve each of them is extremely small—smaller, on average, than for the other instances. The average computing time is in line with the percentage of instances solved for the different classes, showing that the hardest instances are the ones subject to partition-matroid constraints. Interestingly, while the introduction of conflict constraints leads to larger LP gaps, it does not affect too much the computing times, leading to an increase by only 14%, 22%, and 12%, for, respectively, the k-card, kp, and p-matroid classes.

6.2 Experiments for the generalization of Problem (2) with a not increasing utility function

We report the results obtained on instances of Problem (2) where the utility function is not increasing and $\hat{h}_i, i \in M$, is not submodular. We consider the utility function (g) introduced in Sect. 5, and set $\lambda \in \{0.25, 0.5, 0.75, 1, 2, 3, 4\}$ and $\alpha = 0.5$, considering two values of $m \in \{50, 100\}$.

We consider three B&C algorithms: OA+SC+POLY, OA+BC+POLY and OA+BC+f+POLY. In each of them, Constraints (OA+POLY) are separated at every integer point and also at every fractional point but (based on preliminary experiments) only at the root node. The other constraints are separated as we did before, but only if g is positive at the point being separated: in OA+SC+POLY, we separate Constraints (OA+SC$_1$)–(OA+SC$_2$) at integer points; in OA+BC+POLY, we separate Constraints (OA+BC$_1$)–(OA+BC$_2$) at integer points; in OA+BC+f+POLY, we separate Constraints (OA+BC$_1$)–(OA+BC$_2$) at integer and fractional points.

As testbed, we adopt a subset of 1120 instances of the subclass k-card with $\hat{n} = 40$, $n \in \{5000, 10,000\}$, and $k = 10$. In line with the previous experiments, each B&C algorithm is initialized with the greedy algorithm (see Sect. 6.1.2), halting it as soon as the marginal contribution becomes negative (note that, since $\hat{h}_i, i \in M$, is not submodular, the greedy algorithm does not provide approximate solutions).

The computational results are summarized in Table 3. They are aggregated by the risk-tolerance parameter λ (for a total (# total) of 160 instances per value of λ). For each

Table 3 Comparison of three B&C algorithms for the utility function (g), with $\alpha = 0.5$, aggregated w.r.t. the value of λ.

	# total	OA+SC+POLY		OA+BC+POLY		OA+BC+f+POLY	
		# opt	gap	# opt	gap	# opt	gap
$\lambda = 0.25$	160	7	0.5	6	0.6	**8**	0.5
$\lambda = 0.5$	160	**8**	0.7	**8**	0.8	**8**	0.9
$\lambda = 0.75$	160	**40**	0.2	**40**	0.3	**40**	0.3
$\lambda = 1$	160	56	0.1	77	0.1	**80**	0.1
$\lambda = 2$	160	64	1.4	96	0.5	**160**	0.0
$\lambda = 3$	160	68	1.4	102	0.4	**160**	0.0
$\lambda = 4$	160	68	1.5	103	0.4	**160**	0.0
Total/avg	1120	311	0.8	432	0.5	616	0.2

The best results are highlighted in bold

of the three B&C algorithms, we report the number of instances solved to optimality within the time limit of 600 seconds (# opt) and the average gaps for the remaining instances (gap).

The results are in line with those obtained for the strictly increasing case, and they indicate that the outer approximation method combined with Benders cuts (OA+BC+POLY) clearly outperforms the outer approximation one combined with submodular cuts (OA+SC+POLY). This phenomenon is even more pronounced if Benders cuts are also separated at fractional points (OA+BC+f+POLY). Indeed, OA+BC+f+POLY solves the largest number of instances (616 out of 1120) and, on average, provides the smallest gaps for the unsolved ones (0.2%).

We note that the problem is harder for small values of λ where the decreasing term of g becomes dominant. For $\lambda = 0.25, 0.5$, only 16 instances are solved to optimality by OA+BC+f+POLY and the largest average gaps (between 0.5 and 0.9) are obtained. On such instances, the performance of the three algorithms is almost comparable. For larger values of λ, though, OA+BC+f+POLY becomes much better than the other two methods. In particular, all the instances with $\lambda = 2, 3, 4$ are solved to optimality by it. On average around 20 thousand (POLY) constraints are generated at integer points (with a maximum of 50 thousand) and less than one thousand constraints are generated at fractional points at the root node of the branching tree (with a maximum of three thousand). It is worth noticing that the number of generated (POLY) constraints tends to decrease by increasing the value of λ. This is not surprising since in this case the increasing part of the objective function becomes dominant.

The tests show that our decomposition method based on a double-hypograph reformulation (and, in particular, algorithm OA+BC+f+POLY) can be effectively applied to tackle Problem (2) also when the utility function is not increasing and, thus, the objective function is not submodular and the Ref(SC) reformulation that relies on the submodular cuts of Nemhauser et al. [29] cannot be adopted.

6.3 Experiments when solving the problem of Ahmed and Atamtürk [2]

We now present the results obtained with our methods applied for the solution of Problem (1) with a single knapsack constraint that is studied in Ahmed and Atamtürk [2], Yu and Ahmed [39] and which Problem (2) generalizes. We adopt the reformulation Ref(OA) derived in Sect. 4.5. In the experiments, we separate its constraints at integer points only (the computational benefits of separating fractional points are negligible for these instances). We refer to the resulting B&C method as OA.

We compare OA to the state-of-the-art B&C method of Yu and Ahmed [39] (which outperforms the previous one proposed in Ahmed and Atamtürk [2]), to which we refer as BnC-LI (LI stands for lifted inequalities).

To test the performance of OA against BnC-LI, we consider the same set of instances tested in Yu and Ahmed [39] with $n \in \{100, 150, 200\}$ (number of items), $m \in \{50, 100\}$ (number of scenarios) and $\lambda \in \{0.8, 1, 2\}$ (parameter of the utility function $f(z) = 1 - \exp(-\frac{z}{\lambda})$). For the way a and β are constructed, we refer the reader to Ahmed and Atamtürk [2]. For each value of these parameters, we consider 20 random instances as in Yu and Ahmed [39], for a total of 360 instances.[9]

Since the source code of BnC-LI is not available, its results reported in Table 4 are directly taken from the table reported on page 161 in Yu and Ahmed [39], for which the authors used a Python code based on Gurobi 5.6.3, running on a 2.3 GHz x86 Linux workstation with 7 GB of RAM. While these results were obtained with Gurobi with an optimality gap of 10^{-5}, in ours we run CPLEX with a tighter optimality gap of 10^{-9}, thus obtaining solutions that are numerically more accurate.

In Table 4, we compare the performance of OA, BnC-LI, as well as of BARON (the latter is used for solving the MILP formulation of Problem (1) given in Sect. 1). The results are aggregated by the value of n, λ and m. For the two B&C methods OA and BnC-LI, we report the average number of cuts, the average number of B&C nodes, and the average computing time. For BARON, we report the average computing time.

While all three methods are able to solve every instance to optimality, they exhibit a substantial difference in terms of performance. As far as the computing time is concerned, the best approach is OA, which solves all the instances in an average time always smaller than 0.2 seconds, whereas the average computing time of BnC-LI can be more than one hundred seconds. OA scales much better than BnC-LI for an increasing number of items n, while BnC-LI tends to struggle for small values of λ. On average, the number of B&C nodes of OA is smaller (albeit only slightly) than that of BnC-LI (despite the fact that OA adopts a more stringent optimality tolerance). The difference is more substantial for the instances with a larger number of items n and a smaller value of λ. As for BARON, its computing times, while smaller than that of BnC-LI, are up to three orders of magnitude larger than those of OA.

When inspecting the maximum values (not reported in Table 4 for reasons of space) rather than average ones, we observed that the maximum computing time of OA never exceeds 0.32 seconds, whereas for BnC-LI it is as high as 800 seconds and for BARON as high as 15 seconds. The maximum number of cuts for OA never exceeds

---

[9] The instances can be downloaded at https://github.com/fabiofurini/Maximizing-submodular-utility-functions/blob/main/exputil_data_set.tar.xz.

Table 4 Average computational performance of OA, BnC-LI of Yu and Ahmed [39] and BARON when solving Problem (1) with a single knapsack constraint and employing function (f), aggregated w.r.t. n, λ, or m.

	# total	BnC-LI			OA			BARON
		Cuts	Nodes	Time [s]	Cuts	Nodes	Time [s]	Time [s]
$n = 100$	120	706.3	214.2	27.0	1308.5	285.1	**0.1**	2.7
$n = 150$	120	1012.5	410.2	79.0	1207.0	345.1	**0.2**	3.8
$n = 200$	120	1689.3	882.0	135.7	1322.4	319.3	**0.2**	4.3
$\lambda = 0.8$	120	1702.3	784.2	115.8	1368.5	305.9	**0.2**	3.7
$\lambda = 1$	120	1188.5	516.8	81.8	1350.3	303.3	**0.2**	3.6
$\lambda = 2$	120	517.3	205.3	44.0	1119.1	340.4	**0.2**	3.5
$m = 50$	180	757.3	524.8	58.8	810.2	329.2	**0.1**	2.6
$m = 100$	180	1514.8	479.4	102.3	1748.4	303.8	**0.2**	4.6
All instances	360	1136.1	502.1	80.6	1279.3	316.5	**0.2**	3.6

The best results are highlighted in bold

1600, whereas it can be as high as 12,864 for BnC-LI. The maximum number of nodes for OA never exceeds 1400, whereas it can be as high as 7031 for BnC-LI.

Overall, our analysis shows that, in spite of the experiments being run on different architectures and with different solvers, the difference in the computing times between OA and BnC-LI is of various orders of magnitude (up to 3) in favor of OA. We believe that such a difference is mainly due to the different complexity of the corresponding separation algorithms. Indeed, while the one in OA only takes linear time $O(n)$, the one in BnC-LI requires a larger time, at least quadratic as shown in Proposition 12 and Corollary 14 of Yu and Ahmed [39]. Moreover, the separation problem in BnC-LI can be solved in polynomial time only for integer solutions and, for fractional solutions, a heuristic separation approach is adopted, differently from our implementation of OA, where fractional points are not separated.

7 Conclusions

In this article, we have studied the generalization of the problem of maximizing a concave, strictly increasing, and differentiable utility function over a discrete set studied by Ahmed and Atamtürk [2] that is obtained by combining the utility function with a set-union operator. We have introduced a double-hypograph reformulation for the problem that has allowed us to decompose it into its two main components, one related to the utility function and the other one related to the set-union operator. We have then linearized the two components: we have replaced the utility function by its first order outer-approximation (OA) and the set-union operator by either submodular cuts (SCs) or via Benders cuts (BCs). We have obtained two MILP reformulations: Ref(OA+SC) and Ref(OA+BC). We have studied the relationship between these two reformulations and the standard formulation Ref(SC) of Nemhauser et al. [29], and derived theoretical results to compare the strength of the associated constraints. In particular, we have

shown that the constraints in Ref(OA+BC) are at least as tight as those in Ref(OA+SC) and that there are instances in which there is no dominance between these constraints and those of Ref(SC). For the special case in which the utility function is affine, we have shown that the constraints of Ref(SC) and Ref(OA+SC) are dominated by those of Ref(OA+BC). Considering Problem (1), we have shown that there is no dominance between the constraints proposed in Ahmed and Atamtürk [2], Yu and Ahmed [39] and those of our OA-based reformulation.

We have then extended our methods to the case where the utility function is not necessarily increasing and, thus, the objective function of the problem is not submodular, combining our double-hypograph reformulation with an epigraph reformulation of a submodular function obtained via polymatroid inequalities.

In our computational study, we have shown that a branch-and-cut approach based on Ref(OA+BC) drastically outperforms the alternative MILP reformulations as well as the state-of-the-art solver for mixed-integer nonlinear optimization BARON, also for the case where the utility function is not increasing. This is particularly true when tackling problems in which the number of items is many orders of magnitude larger than the number of metaitems. We remark that our approach can be straightforwardly generalized to the case in which a different utility function is considered in each scenario. The same is true if the utility function is a piecewise linear function. In particular, since, in such a case, only as many supporting functions as the number of pieces are needed, the overall approach is likely to be more efficient.

Many open questions still remain to be addressed for this rich optimization problem. One way to linearize the problem is to work in the (x, y) space and proceed along the lines of Ahmed and Atamtürk [2] to "project f out onto the x variables", which is a way of deriving outer-approximation cuts for a hypograph reformulation of the problem. Once the linearization is performed in the (x, y) space, the x variables could be projected out in a Benders-like fashion. Finally, one could obtain an alternative MILP formulation in the y space by generating Generalized Benders Cuts (Geoffrion [15], Fischetti et al. [12]) derived from the associated convex continuous subproblem. The connection between these different types of decomposition and linearizations and the strength of the respective MILP reformulations (both in the (x, y) and in the y space) remain an open topic for future research.

Another line of research addresses the investigation of problems where the function $\hat{q}_i$ is not induced by the set-union operator. In particular, our decomposition approach and the inequalities we derived based on the outer approximation combined with submodular cuts can be applied to any problem involving a function $h_i = f \circ \hat{q}_i$ where $\hat{q}_i$ is an increasing submodular set function. A notable example is the one where $\hat{q}_i$ is the rank (or weighted rank) function of a matroid, e.g., the value function of a graphic matroid whose ground set is determined by the set variable $\hat{S}$. In such a case, one could transform $\hat{q}_i$ into the value function $\hat{Q}_i$ of a suitable optimization problem and construct Benders cuts analogue to those we derived by leveraging the full description of the convex hull of the problem underlying $\hat{Q}_i$.

A Proof of Proposition 7 and related numerical examples

We remark that, since f is concave, strictly increasing, and differentiable, $f(x) - f(x_0) \leq f'(x_0)(x - x_0)$ and $f(x) - f(x_0) \geq f'(x)(x - x_0)$ hold for every $x, x_0 \in \mathbb{R}$ with $x \geq x_0$. As $f' > 0$, both inequalities are strict whenever $x > x_0$.

The proof relies on the following two lemmata.

Lemma 2 *For each $i \in M$, $\hat{S} \subseteq \hat{N}$, and $\ell \in \hat{N} \setminus \hat{S}$:*

1. *the coefficient of y_ℓ in ($SC_1^{\hat{h}}$) is at least as tight as the coefficient of y_ℓ in ($OA+SC_1$), and strictly tighter if $\varrho_\ell^{\hat{q}}(\hat{S}) > 0$.*
2. *if $\varrho_\ell^{q_i}(\hat{S}) = 0$, the coefficients of y_ℓ in ($OA+SC_1$) and ($SC_1^{\hat{h}}$) are equivalently tight (and equal to 0).*

Proof 1. As f is concave, strictly increasing, and differentiable and $\hat{q}_i$ is increasing:

$$\varrho_\ell^{\hat{h}_i}(\hat{S}) = f\big(\hat{q}_i(\hat{S} \cup \{\ell\})\big) - f\big(\hat{q}_i(\hat{S})\big) \leq f'\big(\hat{q}_i(\hat{S})\big) \big(\hat{q}_i(\hat{S} \cup \{\ell\}) - \hat{q}_i(\hat{S})\big)$$
$$= f'\big(\hat{q}_i(\hat{S})\big)\, \varrho_\ell^{\hat{q}_i}(\hat{S}).$$

Thus, the coefficient of y_ℓ in ($SC_1^{\hat{h}}$) is at least as tight as the coefficient of y_ℓ in ($OA+SC_1$) and strictly tighter if $\varrho_\ell^{\hat{q}_i}(\hat{S}) > 0$ due to the inequality being strict.

2. Assume $\varrho_\ell^{\hat{q}_i}(\hat{S}) = 0$. By definition of $\hat{h}_i$, we deduce $\varrho_\ell^{\hat{h}_i}(\hat{S}) = 0$; thus, the coefficients of the two constraints are equal.

$\square$

Lemma 3 *For each $i \in M$, $\hat{S} \subseteq \hat{N}$, and $\ell \in \hat{S}$:*

1. *if $\hat{q}_i(\hat{N} \setminus \{\ell\}) \geq \hat{q}_i(\hat{S})$, the coefficient of $1 - y_\ell$ in ($OA+SC_1$) is at least as tight as the coefficient of $1 - y_\ell$ in ($SC_1^{\hat{h}}$), and strictly tighter if in addition $\varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\}) > 0$ holds.*
2. *if $\hat{q}_i(\hat{N}) = \hat{q}_i(\hat{S})$, the coefficient of $1 - y_\ell$ in ($SC_1^{\hat{h}}$) is at least as tight as the coefficient of $1 - y_\ell$ in ($OA+SC_1$).*

Proof 1. As f is concave, strictly increasing, and differentiable and $\hat{q}_i$ is increasing:

$$\varrho_\ell^{\hat{h}_i}(\hat{N} \setminus \{\ell\}) = f\big(\hat{q}_i(\hat{N})\big) - f(\hat{q}_i(\hat{N} \setminus \{\ell\})) \leq f'\big(\hat{q}_i(\hat{N} \setminus \{\ell\})\big) \big(\hat{q}_i(\hat{N}) - \hat{q}_i(\hat{N} \setminus \{\ell\})\big)$$

$$= f'(\hat{q}_i(\hat{N} \setminus \{\ell\})) \, \varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\}).$$

As f is increasing, $\hat{q}_i(\hat{N} \setminus \{\ell\}) \geq \hat{q}_i(\hat{S})$ implies $f'(\hat{q}_i(\hat{N} \setminus \{\ell\})) \leq f'(\hat{q}_i(\hat{S}))$. Thus, $\varrho_\ell^{\hat{h}_i}(\hat{N} \setminus \{\ell\}) \leq f'(\hat{q}_i(\hat{S})) \varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\})$, i.e., the coefficient of $1 - y_\ell$ in (OA+SC$_1$) is at least as tight as the one in (SC$_1^{\hat{h}}$), and strictly tighter if in addition $\varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\}) > 0$ due to f being strictly increasing.

2. As f is concave, strictly increasing, and differentiable and $\hat{q}_i$ is increasing:

$$\varrho_\ell^{\hat{h}_i}(\hat{N} \setminus \{\ell\}) = f\big(\hat{q}_i(\hat{N})\big) - f\big(\hat{q}_i(\hat{N} \setminus \{\ell\})\big) \geq f'\big(\hat{q}_i(\hat{N})\big) \big(\hat{q}_i(\hat{N}) - \hat{q}_i(\hat{N} \setminus \{\ell\})\big)$$
$$= f'(\hat{q}_i(\hat{N})) \, \varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\}).$$

If $\hat{q}_i(\hat{N}) = \hat{q}_i(\hat{S})$, the right-hand side coincides with the coefficient of $1 - y_\ell$ in (OA+SC$_1$). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad$ $\square$

Proof of Proposition 7 Since $\hat{h}_i(\hat{S}) = f(\hat{q}_i(\hat{S}))$, the constant terms of both inequalities are identical.

1. If $\varrho_\ell^{\hat{q}}(\hat{S}) > 0$ for some $\ell \in \hat{N} \setminus \hat{S}$, part 1 of Lemma 2 applies and (SC$_1^{\hat{h}}$) is strictly tighter than (OA+SC$_1$) w.r.t. at least a coefficient. If $\hat{q}_i(\hat{N} \setminus \{\ell'\}) \geq \hat{q}_i(\hat{S})$ and $\varrho_{\ell'}^{\hat{q}_i}(\hat{N} \setminus \{\ell'\}) > 0$ for some $\ell' \in \hat{S}$, part 1 of Lemma 3 applies and (OA+SC$_1$) is strictly tighter than (SC$_1^{\hat{h}}$) w.r.t. at least a coefficient. Thus, (SC$_1^{\hat{h}}$) and (OA+SC$_1$) do not dominate one another.
2. As $\hat{q}_i(\hat{S}) = \hat{q}_i(\hat{N})$ implies $\varrho_\ell^{\hat{q}}(\hat{S}) = 0$ for all $\ell \in \hat{N} \setminus \hat{S}$, (SC$_1^{\hat{h}}$) is at least as tight as (OA+SC$_1$) w.r.t. all the coefficients of y_ℓ due to part 2 of Lemma 2 (the coefficients coincide). If $\hat{q}_i(\hat{S}) = \hat{q}_i(\hat{N})$, (SC$_1^{\hat{h}}$) is at least as tight as (OA+SC$_1$) w.r.t. the coefficients of $1 - y_\ell$ due to part 2 of Lemma 3.

$\qquad$ $\square$

The following example illustrates the no domination result between (SC$_1^{\hat{h}}$) and (OA+SC$_1$) of Proposition 7:

Example 1 Consider the instance used in Remark 1 and let $\hat{S} = \{4\}$. (SC$_1^{\hat{h}}$) and (OA+SC$_1$) read:

$$w \leq 0.105 + 0.053 y_1 + 0.094 y_2 + 0.044 y_3 - 0.088(1 - y_4) + 0.133 y_5$$
$$w \leq 0.105 + \underline{0.054} y_1 + \underline{0.099} y_2 + \underline{0.045} y_3 - \mathbf{0.098}(1 - y_4) + \underline{0.144} y_5.$$

Since the latter is tighter at the coefficient of $1 - y_4$ (in boldface) but weaker at those of y_1, y_2, y_3, and y_5 (underlined), the two inequalities do not dominate one another. $\square$

If f is almost linear in the interval $[\hat{q}_i(\hat{S}), \hat{q}_i(\hat{S} \cup \{\ell\})]$, then we have

$$\varrho_\ell^{\hat{h}_i}(\hat{S}) = f(\hat{q}_i(\hat{S} \cup \{\ell\})) - f(\hat{q}_i(\hat{S})) \simeq f'(\hat{q}_i(\hat{S})) \left(\hat{q}_i(\hat{S} \cup \{\ell\}) - \hat{q}_i(\hat{S})\right)$$

$$= f'(\hat{q}_i(\hat{S}))\varrho_\ell^{\hat{q}_i}(\hat{S}).$$

Thus, the difference in tightness of the coefficients of y_ℓ for each $\ell \in \hat{N} \setminus \hat{S}$ (in which $(SC_1^{\hat{h}})$ is tighter) almost vanishes. This is illustrated in the following example:

Example 2 Increasing the value of $\lambda = 100$ to make the function f closer to linear, for the same instance we used before and the same set $\hat{S}$ we have the following $(SC_1^{\hat{h}})$ and $(OA+SC_1)$ inequalities (note that we increased the numerical precision to 5 decimal places):

$$w \le 0.01095 + 0.00592y_1 + 0.01083y_2 + 0.00494y_3 - 0.01076(1 - y_4) + 0.01570y_5$$
$$w \le 0.01095 + \underline{0.00594}y_1 + \underline{0.01089}y_2 + \underline{0.00495}y_3 - \mathbf{0.01089}(1 - y_4) + \underline{0.01583}y_5.$$

While the latter is tighter than the former at the coefficient of $1 - y_4$ (in boldface) and weaker at those of y_1, y_2, y_3, and y_5 (underlined), the difference in the coefficients of y_1, y_2, y_3, and y_5 is less pronounced. □

B Proof of Proposition 8 and related numerical example

The proof relies on the following two lemmata.

Lemma 4 *For each $i \in M$, $\hat{S} \subseteq \hat{N}$, and $\ell \in \hat{N} \setminus \hat{S}$:*

1. *if $\hat{q}_i(\hat{S}) \ge \hat{q}_i(\{\ell\})$, the coefficient of y_ℓ in $(OA+SC_2)$ is at least as tight as the coefficient of y_ℓ in $(SC_2^{\hat{h}})$, and strictly tighter if $\varrho_\ell^{\hat{q}_i}(\emptyset) > 0$ in addition holds.*
2. *if $\hat{q}_i(\hat{S}) = \hat{q}_i(\emptyset)$, the coefficient of y_ℓ in $(SC_2^{\hat{h}})$ is at least as tight as the coefficient of y_ℓ in $(OA+SC_2)$.*

Proof 1. As f is concave, strictly increasing, and differentiable and $\hat{q}_i$ is increasing:

$$\varrho_\ell^{\hat{h}_i}(\emptyset) = f(\hat{q}_i(\{\ell\})) - f(\hat{q}_i(\emptyset)) \ge f'(\hat{q}_i(\{\ell\}))\,(\hat{q}_i(\{\ell\}) - \hat{q}_i(\emptyset)) = f'(\hat{q}_i(\{\ell\}))\,\varrho_\ell^{\hat{q}_i}(\emptyset).$$

As f is increasing, $\hat{q}_i(\hat{S}) \ge \hat{q}_i(\{\ell\})$ implies $f'(\hat{q}_i(\hat{S})) \le f'(\hat{q}_i(\{\ell\}))$. Thus, $\varrho_\ell^{\hat{h}_i}(\emptyset) \ge f'(\hat{q}_i(\hat{S}))\,\varrho_\ell^{\hat{q}_i}(\emptyset)$, i.e., the coefficient of y_ℓ in $(OA+SC_2)$ is at least as tight as the one in $(SC_2^{\hat{h}})$, and strictly tighter if $\varrho_\ell^{\hat{q}_i}(\emptyset) > 0$ in addition holds, due to f being strictly increasing.

2. As f is concave, strictly increasing, and differentiable and $\hat{q}_i$ is increasing:

$$\varrho_\ell^{\hat{h}_i}(\emptyset) = f(\hat{q}_i(\{\ell\})) - f(\hat{q}_i(\emptyset)) \le f'(\hat{q}_i(\emptyset))\,(\hat{q}_i(\{\ell\}) - \hat{q}_i(\emptyset))$$
$$= f'(\hat{q}_i(\emptyset))\,\varrho_\ell^{\hat{q}_i}(\emptyset).$$

As f is increasing, $f'(\hat{q}_i(\emptyset)) \leq f'(\hat{q}_i(\hat{S}))$ if and only if $\hat{q}_i(\emptyset) \geq \hat{q}_i(\hat{S})$, which holds if and only if $\hat{q}_i(\hat{S}) = \hat{q}_i(\emptyset)$. In such a case, the coefficient of y_ℓ in (SC$_2^{\hat{h}}$) is at least as tight as the coefficient of y_ℓ in (OA+SC$_2$).

$\square$

Lemma 5 *For each $i \in M$, $\hat{S} \subseteq \hat{N}$, and $\ell \in \hat{S}$:*

1. *the coefficient of $1 - y_\ell$ in (SC$_2^{\hat{h}}$) is at least as tight as the coefficient of $1 - y_\ell$ in (OA+SC$_2$), and strictly tighter if $\varrho_\ell^{\hat{q}}(\hat{S} \setminus \{\ell\}) > 0$.*
2. *if $\varrho_\ell^{\hat{q}_i}(\hat{S} \setminus \{\ell\}) = 0$, the coefficients of $1 - y_\ell$ in (OA+SC$_2$) and (SC$_2^{\hat{h}}$) are equivalently tight (and equal to 0).*

Proof 1. Since f is concave, strictly increasing, and differentiable and $\hat{q}_i$ is increasing, we have:

$$\varrho_\ell^{\hat{h}_i}(\hat{S} \setminus \{\ell\}) = f(\hat{q}_i(\hat{S})) - f(\hat{q}_i(\hat{S} \setminus \{\ell\})) \geq f'(\hat{q}_i(\hat{S})) \left(\hat{q}_i(\hat{S}) - \hat{q}_i(\hat{S} \setminus \{\ell\}) \right)$$
$$= f'(\hat{q}_i(\hat{S})) \, \varrho_\ell^{\hat{q}_i}(\hat{S} \setminus \{\ell\}).$$

Thus, the coefficient of $1 - y_\ell$ in (SC$_2^{\hat{h}}$) is at least as tight as the coefficient of $1 - y_\ell$ in (OA+SC$_2$), and strictly tighter if $\varrho_\ell^{\hat{q}}(\hat{S} \setminus \{\ell\}) > 0$ due to the inequality being strict.

2. Assuming $\varrho_\ell^{\hat{q}_i}(\hat{S} \setminus \{\ell\}) = 0$, by definition of $\hat{h}_i$ we deduce $\varrho_\ell^{\hat{h}_i}(\hat{S} \setminus \{\ell\}) = 0$; thus, the coefficients of the two constraints are equal.

$\square$

Proof of Proposition 8 As $\hat{h}_i(\hat{S}) = f(\hat{q}_i(\hat{S}))$, the constant terms of the two inequalities are identical.

1. If $\hat{q}_i(\hat{S}) \geq \hat{q}_i(\{\ell\})$ and $\varrho_\ell^{\hat{q}_i}(\emptyset) > 0$ for some $\ell \in \hat{N} \setminus \hat{S}$, part 1 of Lemma 4 applies and (OA+SC$_2$) is strictly tighter than (SC$_2^{\hat{h}}$) w.r.t. at least a coefficient. If $\varrho_{\ell'}^{\hat{q}}(\hat{S} \setminus \{\ell'\}) > 0$ for some $\ell' \in \hat{S}$, part 1 of Lemma 5 applies and (SC$_2^{\hat{h}}$) is strictly tighter than (OA+SC$_2$) w.r.t. at least a coefficient. Thus, (SC$_2^{\hat{h}}$) and (OA+SC$_2$) do not dominate one another.
2. If $\hat{q}_i(\hat{S}) \geq \hat{q}_i(\{\ell\})$ holds for all $\ell \in \hat{N} \setminus \hat{S}$, part 1 of Lemma 4 applies and the coefficients of all y_ℓ of (OA+SC$_2$) are at least as tight as the ones of (SC$_2^{\hat{h}}$). If $\varrho_{\ell'}^{\hat{q}}(\hat{S} \setminus \{\ell'\}) = 0$ for all $\ell' \in \hat{S}$, part 2 of Lemma 5 applies and the coefficients of $1 - y_\ell$ of (SC$_2^{\hat{h}}$) and (OA+SC$_2$) coincide. Thus, (OA+SC$_2$) dominates (SC$_2^{\hat{h}}$).
3. Assume $\hat{q}_i(\hat{S}) = \hat{q}_i(\emptyset)$ holds. Due to part 2 of Lemma 4, (SC$_2^{\hat{h}}$) is at least as tight as (OA+SC$_2$) w.r.t. all the coefficients of y_ℓ. Since $\hat{q}_i(\hat{S}) = \hat{q}_i(\emptyset)$ implies $\varrho_\ell^{\hat{q}}(\hat{S} \setminus \{\ell\}) = 0$ for all $\ell \in \hat{S}$, part 2 of Lemma 5 applies and the coefficients of $1 - y_\ell$ of (SC$_2^{\hat{h}}$) and (OA+SC$_2$) coincide. Thus, (SC$_2^{\hat{h}}$) dominates (OA+SC$_2$).

$\square$

The following example illustrates the no domination result between ($\text{SC}_2^{\hat{h}}$) and (OA+SC_2) of Proposition 8:

Example 3 Consider the instance of Remark 1 and let $\hat{S} = \{5\}$. ($\text{SC}_2^{\hat{h}}$) and (OA+SC_2) read:

$$w \le 0.148 + 0.059y_1 + 0.105y_2 + 0.049y_3 + 0.105y_4 - 0.148(1 - y_5)$$
$$w \le 0.148 + \mathbf{0.052}y_1 + \mathbf{0.094}y_2 + \mathbf{0.043}y_3 + \mathbf{0.094}y_4 - \underline{0.136}(1 - y_5).$$

The latter is tighter at the coefficients of y_1, y_2, y_3, and y_4 (highlighted in boldface), but weaker at the coefficient of $1 - y_5$ (underlined). Thus, the two inequalities do not dominate one another. $\square$

C Proof of Proposition 10

Proof As a consequence of Lemma 1, the pair $(\tilde{\delta}, \tilde{\sigma})$ in Equation (7) corresponds to the extreme solution to the dual of the Benders subproblem defined in (BS) that is obtained by setting $\gamma_j := 1$ if $j \in N_1$ and $\gamma_j := 0$ otherwise. Due to (7), we have:

$$d_i + \sum_{j \in N} \tilde{\sigma}_j = d_i + \sum_{j \in N(\hat{S}) \setminus N_1} a_{ij} = \hat{q}_i(\hat{S}) - \sum_{j \in N(\hat{S}) \cap N_1} a_{ij} \tag{11}$$

$$\sum_{j \in N(\ell)} \tilde{\delta}_j = \begin{cases} \sum_{j \in N(\ell) \setminus N(\hat{S})} a_{ij} & \ell \in \hat{N} \setminus \hat{S} \\ \sum_{j \in N(\ell) \cap N_1} a_{ij} & \ell \in \hat{S}. \end{cases} \tag{12}$$

Using (11) and (12), (BC) can be rewritten as:

$$\hat{\mathcal{Q}}_i(y) \le \hat{q}_i(\hat{S}) - \sum_{j \in N(\hat{S}) \cap N_1} a_{ij} + \sum_{\ell \notin \hat{S}} \left(\sum_{j \in N(\ell) \setminus N(\hat{S})} a_{ij} \right) y_\ell + \sum_{\ell \in \hat{S}} \left(\sum_{j \in N(\ell) \cap N_1} a_{ij} \right) y_\ell. \tag{13}$$

The coefficient of y_ℓ in the third term is equal to $\varrho_\ell^{\hat{q}_i}(\hat{S})$. As the summation in the second term only considers items $j \in N(\hat{S})$ which belong to N_1, we have:

$$\sum_{j \in N(\hat{S}) \cap N_1} a_{ij} = \sum_{\ell \in \hat{S}} \sum_{j \in N(\ell) \cap N_1} a_{ij}.$$

Thus, we can reorder the terms of the last two summations in (13), obtaining (BC_1). The proof is concluded by noting that

$$\varrho_\ell^{\hat{q}_i}(\hat{N} \setminus \{\ell\}) = \sum_{j \in N(\ell) \cap N_1} a_{ij}. \qquad \square$$

D Proof of Proposition 11

Proof With the choice of dual multipliers in (8), we rewrite (BC) as:

$$\hat{Q}_i(y) \leq d_i + \sum_{j \in N(\hat{S}) \setminus N_1(\hat{S})} a_{ij} + \sum_{\ell \in \hat{N} \setminus \hat{S}} \left(\sum_{j \in N(\ell) \setminus N(\hat{S})} a_{ij} + \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) y_\ell$$
$$+ \sum_{\ell \in \hat{S}} \left(\sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) y_\ell. \tag{14}$$

Since $\varrho_\ell^{\hat{q}_i}(\emptyset) = \sum_{j \in N(\ell)} a_{ij}$ and $\varrho_\ell^{\hat{q}_i}(\hat{S} \setminus \{\ell\}) = \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij}$, the right-hand side of ($\text{SC}_2^{\hat{q}}$) can be rewritten as:

$$d_i + \sum_{j \in N(\hat{S})} a_{ij} + \sum_{\ell \in \hat{N} \setminus \hat{S}} \left(\sum_{j \in N(\ell)} a_{ij} \right) y_\ell - \sum_{\ell \in \hat{S}} \left(\sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) (1 - y_\ell). \tag{15}$$

Since $\sum_{j \in N(\hat{S})} a_{ij} - \sum_{\ell \in \hat{S}} \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} = \sum_{j \in N(\hat{S}) \setminus N_1(\hat{S})} a_{ij}$, (15) can be restated as:

$$d_i + \sum_{j \in N(\hat{S}) \setminus N_1(\hat{S})} a_{ij} + \sum_{\ell \in \hat{N} \setminus \hat{S}} \left(\sum_{j \in N(\ell)} a_{ij} \right) y_\ell + \sum_{\ell \in \hat{S}} \left(\sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij} \right) y_\ell.$$

Such a quantity is larger than the right-hand side of (14). By rewriting the term $\sum_{j \in N(\hat{S}) \setminus N_1(\hat{S})} a_{ij}$ as $\sum_{j \in N(\hat{S})} a_{ij} - \sum_{\ell \in \hat{S}} \sum_{j \in N(\ell) \cap N_1(\hat{S})} a_{ij}$ and collecting y_ℓ for all $\ell \in \hat{S}$, (14) is shown to coincide with (BC_2). $\qquad\square$

E Worst-case for instances for the greedy algorithm

As mentioned in Sect. 6.1.2, the greedy algorithm proposed in Nemhauser et al. [29] allows to compute good-quality feasible solutions to the problem of maximizing a submodular function subject to a k-cardinality constraint. In this section, we show how to construct *worst-case instances* for Problem (2) subject to a k-cardinality constraint, i.e., instances on which the greedy algorithm achieves the worst-case approximation factor of $1 - 1/e$, asymptotically.

Letting $m = 1$ and $f(z) := z$, Problem (2) with a k-cardinality constraint corresponds to the Maximum Covering Problem (MCP), see, e.g., Cordeau et al. [6]. A procedure for creating worst-case instances for the MCP can be found in Hochbaum and Pathria [18]. Here, we show how to adapt such a procedure to create worst-case instances for Problem (2) subject to a k-cardinality constraint and with the function $f(z)$ defined as in (f).

Let us introduce the matrix $U = [u_{st}]$ with $s = 0, 1, \ldots, k$ and $t = 1, 2, \ldots, k$, for some $k \in \mathbb{N}$. This matrix has $k + 1$ rows and k columns with entries

$$u_{0t} := \begin{cases} k - 1, & t = 1, \\ k - 2, & \text{otherwise} \end{cases} \qquad t = 1, \ldots, k,$$

$$u_{1t} := \begin{cases} 0, & t = 1, \\ 1, & \text{otherwise} \end{cases} \qquad t = 1, \ldots, k$$

$$u_{st} := \left(\frac{k}{k-1} \right)^{s-2} \qquad s = 2, \ldots, k, \ t = 2, \ldots, k.$$

Each entry (s, t) of the matrix U is associated with an item j (there are $k \ (k + 1)$ items in N in total) whose value in the single scenario $i = 1$ is

$$a_{1j} := \frac{u_{st}}{\Delta} \quad \text{where} \quad \Delta := k \ (k - 1) \left(\frac{k}{k-1} \right)^{k-1}.$$

The constant Δ is equal to the sum of all the entries of the matrix U. In addition, we introduce $2k$ metaitems defined as follows. The first k metaitems, which we denote by $\mathcal{R} = \{r_1, r_2, \ldots, r_k\}$, correspond to the rows $1, 2, \ldots, k$ of the matrix U; each metaitem r_s, $s = 1, \ldots, k$, covers all the items of row s. The second k metaitems, which we denote by $\mathcal{C} = \{c_1, c_2, \ldots, c_k\}$, correspond to the columns $1, 2, \ldots, k$ of the matrix U; each metaitem c_t, $t = 1, \ldots, k$, covers all the items of column t. Overall, the set of items is $N = \{(s, t) : s = 0, \ldots, k, t = 1, \ldots, k\}$ and the set of metaitems is $\hat{N} = \{\mathcal{R}\} \cup \{\mathcal{C}\}$.

With f defined as $f(z) := 1 - \exp(-\frac{z}{\lambda})$, the optimal solution value is $Z = 1 - \exp(-\frac{1}{\lambda})$. While an optimal solution of this value is obtained by selecting all k metaitems in $\mathcal{C}$, the greedy algorithm constructs one by picking (in this order) the metaitems $r_k, r_{k-1}, \ldots, r_1$, which results in a greedy solution of value $Z^G = 1 - \exp(-\frac{\Delta - (k-1)^2}{\Delta \lambda})$. Hence, we have

$$\lim_{\lambda \to \infty} \frac{Z^G}{Z} = \lim_{\lambda \to \infty} \frac{1 - \exp(-\frac{\Delta - (k-1)^2}{\Delta \lambda})}{1 - \exp(-\frac{1}{\lambda})}$$

$$= \lim_{\lambda \to \infty} \frac{\frac{\Delta - (k-1)^2}{\Delta \lambda}}{\frac{1}{\lambda}} = \frac{\Delta - (k-1)^2}{\Delta} = 1 - \left(1 - \frac{1}{k} \right)^k$$

and

$$\lim_{k \to \infty} \lim_{\lambda \to \infty} \frac{Z^G}{Z} = \lim_{k \to \infty} 1 - \left(1 - \frac{1}{k} \right)^k = 1 - \frac{1}{e}.$$

Figure 6 illustrates the value of the approximation factor as a function of λ and k, also considering the case where f is the identity function. In the latter case, the optimal solution value is equal to 1 and the ratio Z^G/Z is $1 - \left(1 - \frac{1}{k} \right)^k$ (see Fig. 6).
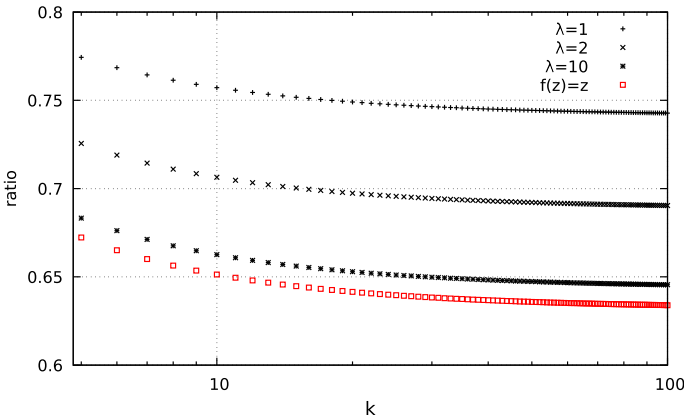
Fig. 6 Ratio between the greedy heuristic solution value Z^G and the optimal solution value Z for the worst-case instances with utility functions $f(z) = 1 - \exp(-\frac{z}{\lambda})$ for different values of $\lambda \in \{1, 2, 10\}$ and $f(z) = z$. The horizontal axis reports in logarithmic scale the value of $k \in \{5, 6, \ldots, 100\}$.

We observe that, with $f(z) = 1 - \exp(-\frac{z}{\lambda})$, the ratio Z^G/Z already falls below 0.65 for $\lambda = 10$ and $k > 30$.

We conclude by briefly commenting a set of computational experiments carried out with OA+BC+f on the worst-case instances we introduced above with values of k up to 1000 (these results are not shown in a table for reasons of space). The experiments reveal that OA+BC+f achieves an excellent computational performance on these instances. Indeed, it is capable to solve to optimality all the instances at the root node of the branching tree within a fraction of a second, including the largest instances with $k = 1000$ that feature $n = 1,001,000$ items and $\hat{n} = 2000$ metaitems.

F On the relation with the curvature of submodular functions

The set N_1, corresponding to the set of items in N which can be covered by a single metaitem $\ell \in \hat{N}$, is related to the notion of *curvature* of a submodular function introduced by Conforti and Cornuéjols [5]. For a monotone increasing set function $f : 2^N \to \mathbb{R}_+$, we call *curvature w.r.t. a set* $S \subseteq N$ the quantity

$$c^f(S) := 1 - \min_{\ell \in S, \varrho_\ell(\emptyset) > 0} \frac{\varrho_\ell(S \setminus \{\ell\})}{\varrho_\ell(\emptyset)},$$

and *total curvature* the quantity

$$c^f := 1 - \min_{S \subseteq N} c^f(S) = 1 - \min_{\ell \in N : \varrho_\ell(\emptyset) > 0} \frac{\varrho_\ell(N \setminus \{\ell\})}{\varrho_\ell(\emptyset)},$$

(where the equality holds due to the monotonicity of f).

The notion of total curvature plays a central role in establishing the approximation factor of the greedy algorithm used for solving the problem of maximizing a submod-

ular function subject to a cardinality constraint. In particular, Conforti and Cornuéjols [5] show that the greedy algorithm for this problem achieves a $(1 - \frac{1}{e^{c^f}})\frac{1}{c^f}$ approximation factor, which ranges from 1 when $c^f = 0$ (i.e., when f is linear and the greedy algorithm is exact) to $1 - \frac{1}{e}$ when $c^f = 1$ (the tight worst-case factor already established by Nemhauser et al. [29]).

Let, for each $i \in M$, $c^{\hat{q}_i}$ be the total curvature of $\hat{q}_i$, and let $c^{\hat{h}_i}$ be the total curvature of $\hat{h}_i$. The following holds:

Observation 1 *For each $i \in M$, we have $c^{\hat{q}_i} < 1$ (respectively $c^{\hat{h}_i} < 1$) if and only if $N_1 \neq \emptyset$ and, for all $\ell \in \hat{N}$, there is some $j \in N(\ell)$ with $a_{ij} > 0$ which is only covered by ℓ (i.e., with $\hat{N}(j) = \{\ell\}$).*

Indeed, by definition of $c^{\hat{q}_i}$ (respectively, $c^{\hat{h}_i}$), we have $c^{\hat{q}_i} < 1$ (respectively, $c^{\hat{h}_i} < 1$) if and only if $\varrho_\ell^{\hat{q}_i}(N \setminus \{\ell\}) > 0$ (respectively, $\varrho_\ell^{\hat{q}_i}(N \setminus \{\ell\}) > 0$) holds for all $\ell \in N$, which is the case if and only if, for each $\ell \in \hat{N}$, there is some $j \in N(\ell)$ with $a_{ij} > 0$ which is only covered by ℓ (i.e., with $\hat{N}(j) = \{\ell\}$). As shown in the following, some of the results we derived for the comparison of the different inequalities we considered in this paper can be related to the total curvature $c^{\hat{h}_i}$ and $c^{\hat{q}_i}$ of $\hat{h}_i$ and $\hat{q}_i$, respectively.

Observation 2 *If $c^{\hat{h}_i} < 1$, then Constraints ($SC_3^{\hat{h}}$) are strictly dominated by Constraints ($SC_1^{\hat{h}}$). Similarly, if $c^{\hat{q}_i} < 1$, then Constraints ($SC_3^{\hat{q}}$) are strictly dominated by Constraints ($SC_1^{\hat{q}}$).*

Indeed, in Sect. 2, we notice that Constraints ($SC_3^{\hat{h}}$) are dominated by ($SC_1^{\hat{h}}$) if there exists an item $j \in N$ which is covered by a single metaitem $\ell \in \hat{N}$, i.e., such that $|\hat{N}(j)| = 1$, and $a_{ij} > 0$. If $c^{\hat{h}_i} < 1$, this is satisfied for all $\ell \in \hat{N}$.

When comparing the strength of Constraints ($SC_1^{\hat{h}}$) and ($OA+SC_1$), we showed in Proposition 7 that a sufficient condition for Constraint ($SC_1^{\hat{h}}$) to dominate ($OA+SC_1$) is given by $\hat{q}_i(\hat{S}) = \hat{q}_i(\hat{N})$. The latter condition implies $c^{\hat{q}_i} = 1$.

Concerning the strength of Constraints ($SC_2^{\hat{h}}$) and ($OA+SC_2$), we showed in Proposition 8 that a sufficient condition for Constraint ($OA+SC_2$) to dominate ($SC_2^{\hat{h}}$) is given by $\hat{q}_i(\hat{S}) > \hat{q}_i(\{\ell\})$ for all $\ell \in \hat{N} \setminus \hat{S}$ and $\hat{q}_i(\hat{S}) = \hat{q}_i(\hat{S} \setminus \{\ell\})$. Condition $\hat{q}_i(\hat{S}) = \hat{q}_i(\hat{S} \setminus \{\ell\})$ for all $\ell \in \hat{S}$ implies a curvature $c^{\hat{q}_i}(\hat{S}) = 1$.

G Performance profiles comparing SC, OA+SC, OA+BC, and OA+BC+f for the case of a strictly increasing utility function

We provide a graphical representation of the relative computational performance of the four branch-and-cut algorithms SC, OA+SC, OA+BC, and OA+BC+f by the performance profiles (see Dolan and Moré [8]) reported in Fig. 7 for classes k-card and k-card & confl, in Fig. 8 for classes kp and kp & confl, and in Fig. 9 for classes p-matroid and p-matroid & confl.

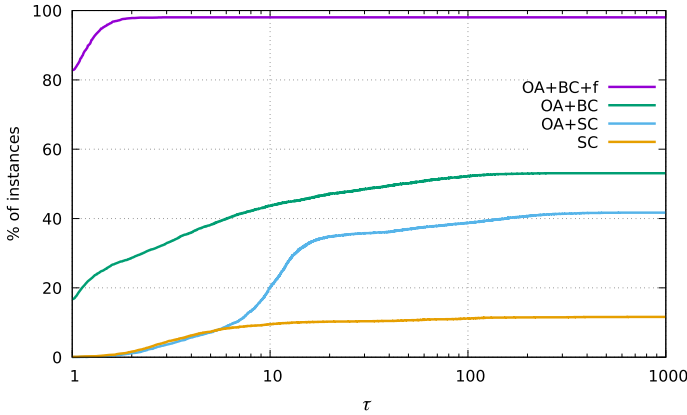**Fig. 7** Performance profile comparing the four branch-and-cut algorithms for the `k-card` and `k-card & confl` classes.
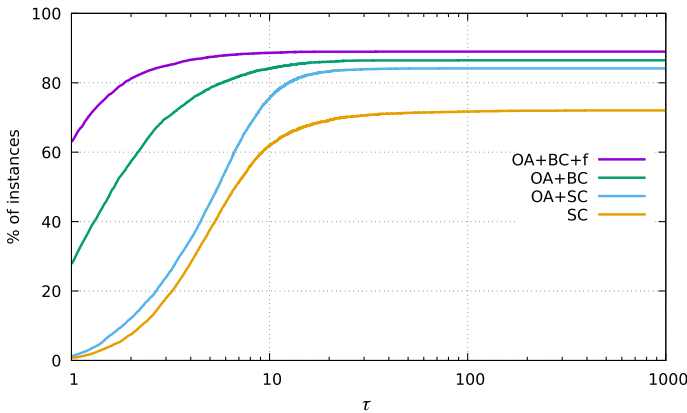

Fig. 8 Performance profile comparing the four branch-and-cut algorithms for the `kp` and `kp & confl` classes.

For each instance and algorithm, let τ be the normalized time defined as the ratio of the computing time required to solve the instance taken by the algorithm and the minimum computing time taken by all algorithms. For each algorithm and for each value of τ in the horizontal axis, the vertical axis reports the percentage of the instances for which the corresponding algorithm spent at most τ times the computing time of the fastest algorithm. Each curve starts from the percentage of instances in which the corresponding algorithm is the fastest. At the right end of the chart, we can read the percentage of instances solved by a specific algorithm. The best performance is achieved by the algorithm whose curve occupies the upper part of the figure. The values on the horizontal axis are reported in logarithmic scale.

The figure clearly shows that OA+BC+f is the best algorithm for this sets of instances, as it is able to solve more instances and it is the fastest algorithm in most of the cases. The second best algorithm is OA+BC. These figure show that separat-
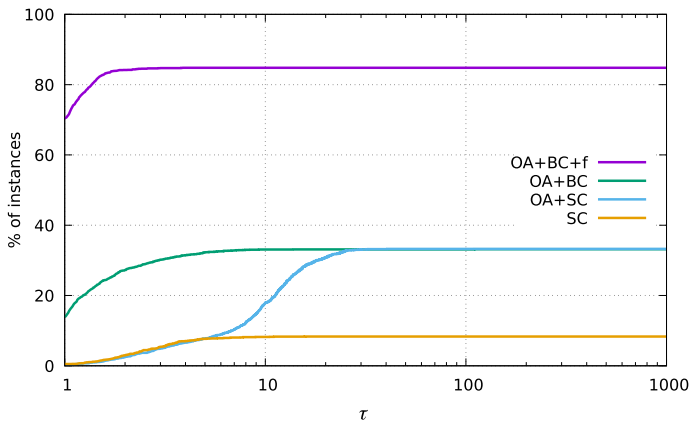
Fig. 9 Performance profile comparing the four branch-and-cut algorithms for the `p-matroid` and `p-matroid & confl` classes.

ing Benders cuts not just for integer points but also for for fractional ones is crucial for the efficiency of the method. The two branch-and-cut algorithms based on submodular cuts, i.e., `OA+SC` and `SC`, achieve the worst performance and they are both outperformed by `OA+BC+f` (and by `OA+BC`).

References

1. Aboolian, R., Berman, O., Krass, D.: Competitive facility location model with concave demand. Eur. J. Oper. Res. **181**(2), 598–619 (2007)
2. Ahmed, S., Atamtürk, A.: Maximizing a class of submodular utility functions. Math. Program. **128**(1–2), 149–169 (2011)
3. Atamtürk, A., Narayanan, V.: Submodular function minimization and polarity. Mathematical Programming, pp. 1–11 (2021)
4. Berman, O., Krass, D.: Flow intercepting spatial interaction model: a new approach to optimal location of competitive facilities. Locat. Sci. **6**(1–4), 41–65 (1998)
5. Conforti, M., Cornuéjols, G.: Submodular set functions, matroids and the greedy algorithm: tight worst-case bounds and some generalizations of the rado-edmonds theorem. Discret. Appl. Math. **7**(3), 251–274 (1984)
6. Cordeau, J.-F., Furini, F., Ljubić, I.: Benders decomposition for very large scale partial set covering and maximal covering location problems. Eur. J. Oper. Res. **275**(3), 882–896 (2019)
7. Cornuejols, G., Fisher, M.L., Nemhauser, G.L.: Exceptional paper-location of bank accounts to optimize float: An analytic study of exact and approximate algorithms. Manage. Sci. **23**(8), 789–810 (1977)
8. Dolan, E.D., Moré, J.J.: Benchmarking optimization software with performance profiles. Math. Program. **91**(2), 201–213 (2002)
9. Duran, M.A., Grossmann, I.E.: An outer-approximation algorithm for a class of mixed-integer nonlinear programs. Math. Program. **36**(3), 307–339 (1986)
10. Edmonds, J.: Submodular functions, matroids, and certain polyhedra. In: Guy, R. (ed.) Combinatorial structures and their applications, pp. 69–87. Gordon and Breach, New York, NY (1971)
11. Feige, U., Mirrokni, V.S., Vondrak, J.: Maximizing non-monotone submodular functions. SIAM J. Comput. **40**(4), 1133–1153 (2011)
12. Fischetti, M., Ljubić, I., Sinnl, M.: Benders decomposition without separability: A computational study for capacitated facility location problems. Eur. J. Oper. Res. **253**(3), 557–569 (2016)

13. Fletcher, R., Leyffer, S.: Solving mixed integer nonlinear programs by outer approximation. Math. Program. **66**, 327–349 (1994)
14. Friedrich, T., Göbel, A., Neumann, F., Quinzan, F., Rothenberger, R.: Greedy maximization of functions with bounded curvature under partition matroid constraints. In: The Thirty-Third AAAI Conference on Artificial Intelligence, Honolulu, Hawaii, USA, January 27 - February 1, 2019, pages 2272–2279. AAAI Press (2019)
15. Geoffrion, A.: Generalized Benders Decomposition. J. Optim. Theory Appl. **10**, 237–260 (1972)
16. Güney, E., Leitner, M., Ruthmair, M., Sinnl, M.: Large-scale influence maximization via maximal covering location. Eur. J. Oper. Res. **289**(1), 144–164 (2021)
17. Håstad, J.: Clique is hard to approximate withinn 1- ε. Acta Math. **182**(1), 105–142 (1999)
18. Hochbaum, D.S., Pathria, A.: Analysis of the greedy approach in problems of maximum k-coverage. Nav. Res. Logist. **45**(6), 615–627 (1998)
19. Karimi, M.R., Lucic, M., Hassani, H., Krause, A.: Stochastic submodular maximization: The case of coverage functions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems, NIPS'17, page 6856-6866, Red Hook, NY, USA (2017). Curran Associates Inc
20. Kempe, D., Kleinberg, J.M., Tardos, É.: Maximizing the spread of influence through a social network. In: Getoor, L., Senator, T.E., Domingos, P.M., Faloutsos, C. (eds.) Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Washington, DC, USA, August 24 - 27, 2003, pages 137–146. ACM (2003)
21. Klastorin, T.: On a discrete nonlinear and nonseparable knapsack problem. Oper. Res. Lett. **9**(4), 233–237 (1990)
22. Küçükaydın, H., Aras, N., Altınel, İK.: A discrete competitive facility location model with variable attractiveness. Journal of the Operational Research Society **62**(9), 1726–1741 (2011)
23. Lehmann, B., Lehmann, D., Nisan, N.: Combinatorial auctions with decreasing marginal utilities. Games Econom. Behav. **55**(2), 270–296 (2006)
24. Lin, Y.H., Tian, Q.: Generalized Benders decomposition for competitive facility location with concave demand and zone-specialized variable attractiveness. Comput. Oper. Res. **130**, 105236 (2021)
25. 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)
26. Martello, S., Pisinger, D., Toth, P.: Dynamic programming and strong bounds for the 0–1 knapsack problem. Manage. Sci. **45**(3), 414–424 (1999)
27. Mehrez, A., Sinuany-Stern, Z.: Resource allocation to interrelated risky projects using a multiattribute utility function. Manage. Sci. **29**(4), 430–439 (1983)
28. Nemhauser, G.L., Wolsey, L.A.: Integer and Combinatorial Optimization. Wiley, New York (1988)
29. 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)
30. ReVelle, C., Scholssberg, M., Williams, J.: Solving the maximal covering location problem with heuristic concentration Comput. Oper. Res. **35**(2), 427–435 (2008)
31. Sahinidis, N.V.: BARON 17.8.9: Global Optimization of Mixed-Integer Nonlinear Programs. User's Manual, (2017)
32. Schrijver, A.: Combinatorial optimization: polyhedra and efficiency, vol. 24. Springer Science & Business Media (2003)
33. Shi, X., Prokopyev, O., Zeng, B.: Sequence independent lifting for the set of submodular maximization problem. In: 25th international Conference on Integer programming and Combinatorial Optimization (IPCO), pages 1–12. Springer (2020)
34. Sviridenko, M.: A note on maximizing a submodular set function subject to a knapsack constraint. Oper. Res. Lett. **32**(1), 41–43 (2004)
35. Tawarmalani, M., Sahinidis, N.V.: A polyhedral branch-and-cut approach to global optimization. Math. Program. **103**, 225–249 (2005)
36. Weingartner, H.M.: Capital budgeting of interrelated projects: survey and synthesis. Manage. Sci. **12**(7), 485–516 (1966)
37. Wu, H., Küçükyavuz, S.: A two-stage stochastic programming approach for influence maximization in social networks. Comput. Optim. Appl. **69**(3), 563–595 (2018)
38. Yu, J., Ahmed, S.: Maximizing expected utility over a knapsack constraint. Oper. Res. Lett. **44**(2), 180–185 (2016)

39. Yu, J., Ahmed, S.: Maximizing a class of submodular utility functions with constraints. Math. Program.
    162(1–2), 145–164 (2017)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps
and institutional affiliations.