
Volume 198, Issue 1, 1 October 2009
ISSN 0377-2217
ELSEVIER
EUROPEAN JOURNAL OF
OPERATIONAL RESEARCH
Available online at
ScienceDirect
www.sciencedirect.com
www.elsevier.com/locate/ejor

European Journal of Operational Research 198 (2009) 102–112

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Discrete Optimization

Maximization of submodular functions: Theory and enumeration algorithms ☆

Boris Goldengorin *

Department of Operations, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 14 March 2008
Accepted 29 August 2008
Available online 9 September 2008

Keywords:
Maximization
Submodular functions
Enumeration algorithms

ABSTRACT

Submodular functions are powerful tools to model and solve either to optimality or approximately many operational research problems including problems defined on graphs. After reviewing some long-standing theoretical results about the structure of local and global maxima of submodular functions, Cherenin's selection rules and his Dichotomy Algorithm, we revise the above mentioned theory and show that our revision is useful for creating new non-binary branching algorithms and finding either approximation solutions with guaranteed accuracy or exact ones.

1. Introduction

In this paper, we give some theoretical results fundamental to the problem of finding a global maximum of a general submodular (or, equivalently, global minimum of a general supermodular) set function which we call the Problem of Maximization of Submodular Functions (PMSF) [34]. By a set function we mean a mapping from 2^N to the real numbers, where $N = \{1, 2, \ldots, n\}$. Another well-known term for an arbitrary set function is a *pseudo-Boolean function* [29] which is a mapping from $\{0,1\}^n$ to the real numbers.

PMSF is known to be NP-hard [20] but the problem of minimization of submodular functions (or, equivalently, the problem of finding a global maximum of a supermodular function) is known to be polynomially solvable [42]. In this paper, we are not dealing with these polynomially solvable problems.

Enormous interest in studying PMSF arises from the fact that several classes of important combinatorial optimization problems belong to PMSF, including the Simple or "uncapacitated" Plant (facility) Location Problem (SPLP) [4,9,16] and its competitive version [8], the Quadratic Cost Partition Problem (QCP) with non-negative edge weights [27], and its special case – the Max-Cut Problem, the generalized transportation problem [37,39]. Genkin et al. [21] have reduced many different problems of data mining and knowledge discovery in biomedical and bioinformatics research (e.g., diagnostic hypothesis generation, logical methods of data analysis, conceptual clustering, and proteins functional annotations) to the PMSF. Many models in mathematics [35], including the rank function of elementary linear algebra, which is a special case of matroid rank functions [15,18] require the solution of a PMSF. Ballester et al. [5] studied the properties and the applicability of policies affecting the structure of the network. They show that this problem is computationally hard and that a simple greedy algorithm used for maximizing submodular set functions finds an acceptable approximation. Submodular functions are used as utility functions in the so called combinatorial auctions (see e.g. [17] and references within) and supermodular functions are applied to model different preferences of choice [11].

Although the general problem of the maximization of a submodular function is known to be NP-hard, there has been a sustained research effort aimed at developing practical procedures for solving medium and large-scale instances of the PMSF. Often the approach taken has been problem specific, and submodularity of the underlying objective function has been only implicit to the analysis. For example, Barahona et al. [6] have addressed the Max-Cut Problem from the point of view of polyhedral combinatorics and developed a branch and cut algorithm, suitable for applications in statistical physics and circuit layout design. Beasley [7] applies Lagrangean heuristics to several classes of location problems including SPLPs and reports results of extensive experiments on a Cray supercomputer. Lee et al. [34] have made a study of the quadratic cost partition problem (QCP) of which max-cut with nonnegative edge weights is a special case, again from the standpoint of polyhedral combinatorics.

There have been fewer published attempts to develop algorithms for minimization of a general supermodular function. We believe that the earliest attempt to exploit supermodularity is the work of Petrov and Cherenin [40], who identified a supermodular structure in their study of railway timetabling. Their procedure was subsequently published by Cherenin [14] as the "method of

☆ This paper is based on the mini-course "Theory and Algorithms for Maximization of Submodular Functions" presented by the author on the 2nd Open Summer School-Seminar "Achievements and Applications of Contemporary Informatics, Mathematics and Physics" (AACIMP 2007), August 8–20, National Technical University of Ukraine "KPI", Kyiv, Ukraine.

* Tel.: +31 50 3633794.
E-mail address: b.goldengorin@rug.nl

0377-2217/$ - see front matter
doi:10.1016/j.ejor.2008.08.022

successive calculations". Their algorithm however is not widely known in the West [3] where, as far we are aware of, the only general procedures that have been studied in depth are the greedy approximation algorithm from Nemhauser et al. [37], and the algorithm for maximization of submodular functions subject to linear constraints from [39]. In a comment to a note by Frieze [19], Babayev [3] demonstrated that Frieze's Property P and Cherenin's theorem Ch (Theorem 3 in this paper) are equivalent. Moreover, Frieze [19] has defined two sets of conditions, namely OP1 and OP2, the application of which in a search procedure are equivalent to the so called Cherenin's "selection rules" for PMSF [14] (see Section 3 in this paper). Note that Alcouffe and Muratet's [2] algorithm is based on a special case of Cherenin's [14] "method of successive calculations".

Indeed the only practical algorithmic implementation of solution procedure known in the West appears to be the "accelerated greedy" (AG) algorithm of Minoux [36], which has been applied to optimal planning and design of telecommunication networks. We note that the AG algorithm has also been applied to the problem of D-optimal experimental design [41]; see also Ko et al. [32] and Lee [33] for further examples of "hard" D-optimal design problems in environmental monitoring. In Genkin and Muchnik [22] an optimal algorithm is constructed with exponential time complexity for the well-known Shannon max–min problem. This algorithm is applied to the maximization of submodular functions subject to a convex set of feasible solutions, and to the problem of – what is known as – decoding monotonic Boolean functions.

Recently based on an elegant pipage rounding technique, introduced by Ageev and Sviridenko [1], many authors (see e.g. Vondrák [43] and references within) are trying to create an approximation algorithm for maximization of submodular functions with additional constraints similarly to the seminal work of Nemhauser and Wolsey [38]. We are not going to address such approximations issues in this paper because they deserve a separately written manuscript.

In this paper, we present a fundamental theorem of Cherenin and our revision of this theorem, both of which provide the basis of selection (preservation) rules, and in particular, for the justification of the Preliminary Preservation (Dichotomy) algorithm. We also revise Cherenin's selection rules dealing with local maxima in the form of our "preservations rules" dealing with global maxima. These rules have been used in Benati [8], Goldengorin [24], Goldengorin et al. [25], and Goldengorin and Ghosh [27]. Moreover, our preservations rules can be used for implicit enumeration of subproblems in a Branch-and-Bound (BnB) approach for solving PMSF.

The paper is organized as follows: In Section 2, we motivate a theoretical development of selection rules based on the essential structural results for local and global maxima of submodular functions due to Cherenin [14] and Khachaturov [30,31]. In this section, a fundamental theorem of Cherenin is stated, which provides the basis of "the method of successive calculations". Section 2 contains an important characterization of local maxima as disjoint components of "strict" and "saddle" vertices which greatly assists the understanding of the difference between the properties of Cherenin's "selection rules" and our "preservation rules" discussed in Section 3. In Section 4, we present our main Theorem 11 from which generalized bounds for implicit enumeration can be derived, and allow the rules of Section 3 to be extended to other cases (ε-*optimality*). We present the two different representations (a) and (b) of the partition of the current set of feasible solutions (vertices) defined by a strictly inner vertex with respect to this set. By using our main Theorem 11 and representations (a) and (b) we prove the correctness of Cherenin's selection rules in the form of our preservation rules. These rules are the basis of Cherenin's Preliminary Preservation Algorithm (PPA) [40]. In Section 5, we outline the main steps of the PPA and illustrate how our new preservation rules (see Corollary 18) can be applied to a small example of the SPLP. We show that if the PPA terminates with a global maximum then the given submodular function has exactly one strict component of local maxima. We introduce the so called non-binary branching rules, based on Theorem 11 in Section 6. Non-binary branching rules are illustrated by means of a SPLP instance. Section 7 gives a number of concluding remarks.

2. The structure of local and global maxima of submodular set functions

In this section, we present results of Cherenin–Khachaturov [14,30], which are hardly known in the Western literature [3].

Let z be a real-valued function defined on the power set 2^N of $N = \{1, 2, \dots, n\}; n \geqslant 1$. For each $S, T \in 2^N$ with $S \subseteq T$, define

$$[S, T] = \{I \in 2^N | S \subseteq I \subseteq T\}.$$

Note that $[\emptyset, N] = 2^N$. Any *interval* $[S, T]$ is, in fact, a *subinterval* of $[\emptyset, N]$ if $\emptyset \subseteq S \subseteq T \subseteq N$; notation $[S, T] \subseteq [\emptyset, N]$. In this paper, we mean by an interval always a subinterval of $[\emptyset, N]$. Throughout this paper we consider a set of PMSFs defined on any interval $[S, T] \subseteq [\emptyset, N]$ as follows:

$$\max\{z(I) | I \in [S, T]\} = z^*[S, T], \text{ for all } [S, T] \subseteq [\emptyset, N].$$

The function z is called *submodular* on $[S, T]$ if for each $I, J \in [S, T]$ it holds that

$$z(I) + z(J) \geqslant z(I \cup J) + z(I \cap J).$$

Expressions of the form $S \setminus \{k\}$ and $S \cup \{k\}$ will be abbreviated to $S - k$ and $S + k$.

The following theorem given in Nemhauser et al. [37] gives a number of equivalent formulations for submodular functions which is useful for a clearer understanding of the concept of submodularity. Since sometime we use the incremental or decremental value of $z(S)$, we define $d_j^+(S) = z(S + j) - z(S)$ and $d_j^-(S) = z(S - j) - z(S)$.

Theorem 1. *All the following statements are equivalent and define a submodular function.*

(i) $z(A) + z(B) \geqslant z(A \cup B) + z(A \cap B), \ \forall A, B \subseteq N.$
(ii) $d_j^+(S) \geqslant d_j^+(T), \ \forall S \subseteq T \subseteq N$ and $j \in N \setminus T$.
(iii) $d_j^+(S) \geqslant d_j^+(S + k), \ \forall S \subseteq N$ and $j \in N \setminus (S + k)$ and $k \in N \setminus S$.
(iv) $z(T) \leqslant z(S) + \sum_{j \in T \setminus S} d_j^+(S), \ \forall S \subseteq T \subseteq N.$
(v) $z(S) \leqslant z(T) + \sum_{j \in T \setminus S} d_j^-(T), \ \forall S \subseteq T \subseteq N.$

As an example consider the Quadratic Cost Partition Problem (QCP), for which it is well known that the objective function $z(Q)$ is a submodular function [34]. For given real numbers p_i and nonnegative real numbers q_{ij} with $i, j \in N$, the QCP is the problem of finding a subset Q of N such that the weight $z(Q) = \sum_{i \in Q} p_i - \frac{1}{2} \sum_{i,j \in Q} q_{ij}$ is as large as possible. Let N be the vertex set, $E \subseteq N \times N$ the edge set of an edge-weighted graph $G = (N, E)$, and $w_{ij} \geqslant 0$ are edge weights. For each $Q \subseteq N$, the cut $\delta(Q)$ is defined as the edge set for which each edge has one end in Q and the other one in $N \setminus Q$. It is easy to see that the Max-Cut Problem with nonnegative edge weights is a QCP where $p_i = \sum_{j \in N} w_{ij}$ and $q_{ij} = 2w_{ij}$, for $i, j \in N$.

Lemma 2. *The objective* $z(S)$ *of the Quadratic Cost Partition problem is submodular.*

Proof. According to Theorem 1(iii) a function is submodular if

$$d_l^+(S) \geqslant d_l^+(S + k), \ \forall S \subseteq N \text{ and } l \in N \setminus (S + k) \text{ and } k \in N \setminus S.$$

Substituting $d_l^+(S) = z(S + l) - z(S)$ we get

$z(S+l) - z(S) \geqslant z(S+k+l) - z(S+k).$

Substituting $z(S) = \sum_{i \in S} p_i - \frac{1}{2}\sum_{i,j \in S} q_{ij}$ gives

$$\sum_{i \in S+l} p_i - \frac{1}{2}\sum_{i,j \in S+l} q_{ij} - \left(\sum_{i \in S} p_i - \frac{1}{2}\sum_{i,j \in S} q_{ij}\right)$$
$$\geqslant \sum_{i \in S+k+l} p_i - \frac{1}{2}\sum_{i,j \in S+k+l} q_{ij} - \left(\sum_{i \in S+k} p_i - \frac{1}{2}\sum_{i,j \in S+k} q_{ij}\right).$$

Canceling out terms involving p_i we obtain

$$-\sum_{i,j \in S+l} q_{ij} + \sum_{i,j \in S} q_{ij} \geqslant -\sum_{i,j \in S+k+l} q_{ij} + \sum_{i,j \in S+k} q_{ij}.$$

This result, after some bookkeeping, implies

$q_{kl} + q_{lk} \geqslant 0.$

Since q_{ij} is nonnegative for all $i, j \in N$, the proof is completed. □

Hence, the QCP problem is a special case of the PMSF.

A subset $L \in [\emptyset, N]$ is called a *local maximum* of z if for each $i \in N$

$z(L) \geqslant \max\{z(L-i), z(L+i)\}.$

A subset $S \in [\emptyset, N]$ is called a *global maximum* of z if $z(S) \geqslant z(I)$ for each $I \in [\emptyset, N]$. We will use the Hasse diagram (see e.g. Fig. 1 [28]) as the ground graph $G = (V, E)$ in which $V = [\emptyset, N]$ and a pair (I, J) is an edge iff either $I \subset J$ or $J \subset I$, and $|I \setminus J| + |J \setminus I| = 1$.

The graph $G = (V, E)$ is called *z-weighted* if the weight of each vertex $I \in V$ is equal to $z(I)$; notation $G = (V, E, z)$. In terms of $G = (V, E, z)$ the PMSF means finding a vertex $S \in V$ of the weight $z(S)$ which is as large as possible. An example of the weighted G with $N = \{1, 2, 3, 4\}$ is depicted in Fig. 2, where the weight $z(I)$ is indicated inside the corresponding vertex I.

Here among others the vertices {1,2,3} and {4} are local maxima, and {4} is a global maximum (see Fig. 2).

A sequence $\Gamma = (I^0, I^1, \ldots, I^n)$ of subsets $I^t \in 2^N$, $t = 0, 1, \ldots, n$ such that $|I^t| = t$ and

$\emptyset = I^0 \subset I^1 \subset I^2 \subset \cdots \subset I^t \subset \cdots \subset I^{n-1} \subset I^n = N$

is called a *chain* in $[\emptyset, N]$. An example of the chain $\emptyset \subset \{2\} \subset \{2,4\} \subset \{1,2,4\} \subset \{1,2,3,4\}$ in Fig. 3 is shown.

Similarly, a chain of any interval $[S, T]$ can be defined. A submodular function z is *nondecreasing (nonincreasing)* on the chain Γ if $z(I^l) \leqslant z(I^m)$ $(z(I^l) \geqslant z(I^m))$ for all l, m such that $0 \leqslant l \leqslant m \leqslant n$; concepts of *increasing, decreasing and constant* (signs, respectively, $<, >, =$) are defined in an obvious manner (see e.g. Fig. 4).

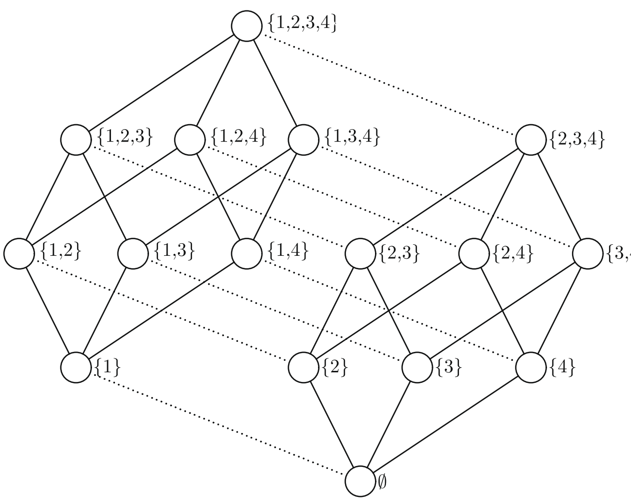

Fig. 1. The Hasse diagram of {1,2,3,4}.

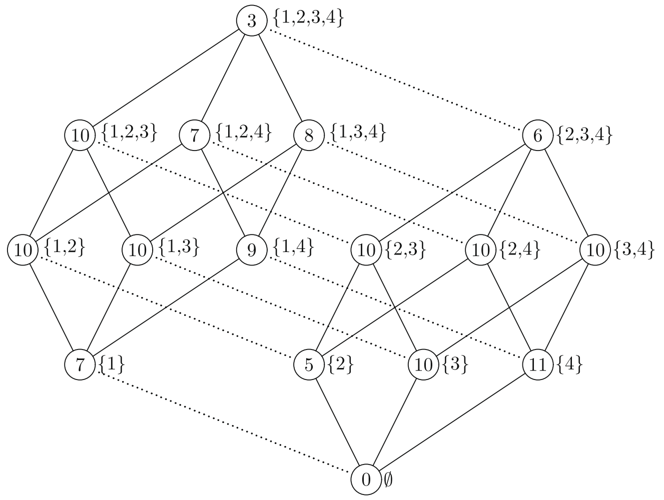

Fig. 2. Example of local maxima {1,2}, {1,2,3}, {1,3}, {2,3}, {3}, and the global maximum {4} on the Hasse diagram.

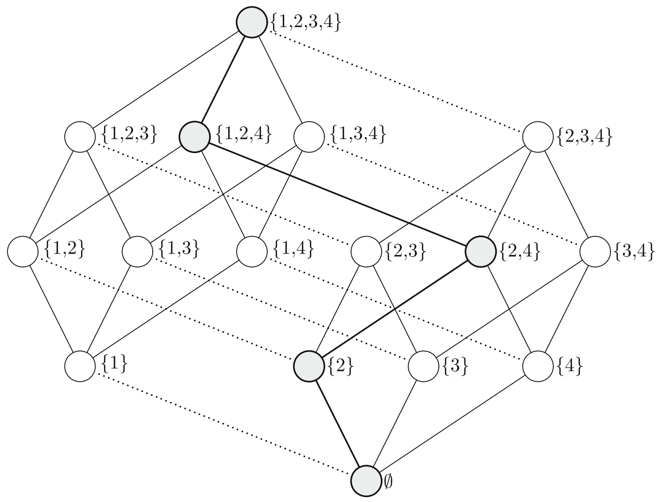

Fig. 3. Example of the chain $\emptyset \subset \{2\} \subset \{2,4\} \subset \{1,2,4\} \subset \{1,2,3,4\}$ in the Hasse diagram of {1,2,3,4}.

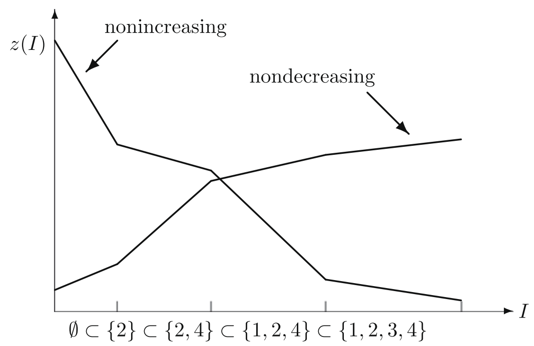

Fig. 4. Example of a nondecreasing (nonincreasing) function on the chain in the Hasse diagram of {1,2,3,4}.

The following theorem [14] shows the quasiconcavity of a submodular function on any chain that includes a local maximum (see Fig. 5).

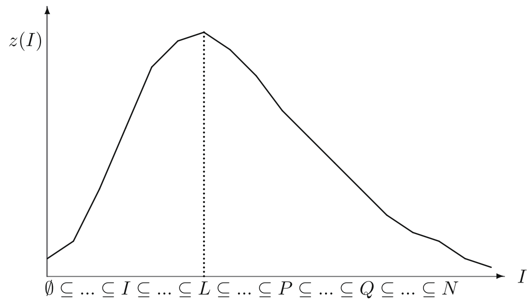

Fig. 5. A quasi-concave behaviour of a submodular function on the chain with a local maximum L (Cherenin's theorem).

Theorem 3. *Let z be a submodular function on 2^N and let L be a local maximum. Then z is nondecreasing on any chain in $[\emptyset, L]$, and nonincreasing on any chain in $[L, N]$.*

Proof. We show that z is nondecreasing on any chain in $[\emptyset, L]$. If either $L = \emptyset$ (we obtain the nonincreasing case) or $|L| = 1$, the assertion is true, since L is a local maximum of z. So, let $|L| > 1$ and $I, J \in [\emptyset, L]$ such that $J = I + k$, $k \in L \setminus I$.

Note that $\emptyset \subseteq \cdots \subseteq I \subset J \subseteq \cdots \subset L$. The submodularity of z implies $z(J) + z(L - k) \geqslant z(I) + z(L)$, or $z(J) - z(I) \geqslant z(L) - z(L - k)$. Since L is a local maximum, $z(L) - z(L - k) \geqslant 0$. Hence $z(J) \geqslant z(I)$, and we have finished the proof of nondecreasing case. The proof for $[L, N]$ is similar. $\square$

Corollary 4. *Let z be a submodular function on 2^N and let L_1 and L_2 be local maxima with $L_1 \subseteq L_2$. Then z is a constant on $[L_1, L_2]$, and every $L \in [L_1, L_2]$ is a local maximum of z.*

Proof. First we show that z is a constant function on $[L_1, L_2]$. Let us apply Theorem 3 to a chain including $\emptyset \subseteq \cdots \subseteq L_1 \subseteq L_2 \subseteq \cdots \subseteq N$, first to the single (isolated) local maximum L_2 and second to the single local maximum L_1. For the first case we obtain $z(\emptyset) \leqslant \cdots \leqslant z(L_1) \leqslant \cdots \leqslant z(I) \leqslant z(L_2)$. For any subchain of the interval $[L_1, L_2]$ we have $z(L_1) \leqslant \cdots \leqslant z(L_2)$. By the same reasons for the second case we have $z(L_1) \geqslant \cdots \geqslant z(L_2)$. Combining both sequences of inequalities we have proved the constancy of z.

Now we show that every $L \in [L_1, L_2]$ is a local maximum of z. Assume to the contrary that there exists $L \in [L_1, L_2]$ that is not a local maximum of z. Then either there is a $L - i \notin [L_1, L_2]$ with $z(L) < z(L - i)$ or there is a $L + i \notin [L_1, L_2]$ with $z(L) < z(L + i)$. For the first case we get according the definition of submodularity $z(L) + z(L_2 - i) \geqslant z(L - i) + z(L_2)$ or $z(L) - z(L - i) \geqslant z(L_2) - z(L_2 - i) \geqslant 0$. This contradicts $z(L) < z(L - i)$. For the second case a similar argument holds by using L_1 instead of L_2. $\square$

In Corollary 4, we have indicated two important structural properties of a submodular function considered on intervals whose end points are local maxima. Namely, on such an interval a submodular function preserves a constant value and every point of this interval is a local maximum. It will be natural to consider the widest intervals with above mentioned properties.

A local maximum $\underline{L} \in 2^N (\overline{L} \in 2^N)$ is called a *lower* (respectively, *upper*) *maximum* if there is no another local maximum L such that $L \subset \underline{L}$ (respectively, $\overline{L} \subset L$). For example, in Fig. 6 the vertex {1,2,3} is an upper local maximum and the vertices {1,2}, {3} are lower local maxima. If an interval $[\underline{L}, \overline{L}]$ with $\underline{L} \subseteq \overline{L}$ has as its end points lower and upper maxima then it is the widest interval on which the submodular function is a constant and each point is a local maximum. We call a pair of intervals $[\underline{L}_i, \overline{L}_i]$ with $\underline{L}_i \subseteq \overline{L}_i$, $i = 1, 2$ *connected* iff $[\underline{L}_1, \overline{L}_1] \cap [\underline{L}_2, \overline{L}_2] \neq \emptyset$. The intervals of local maxima form a set of *components of local maxima*. Two intervals belong to the same component if they are connected. Hence, two local maxima L_1 and L_2 are in the same component if there is a path in $G = (V, E, z)$ with end vertices L_1 and L_2, and all intermediate vertices of this path are local maxima.

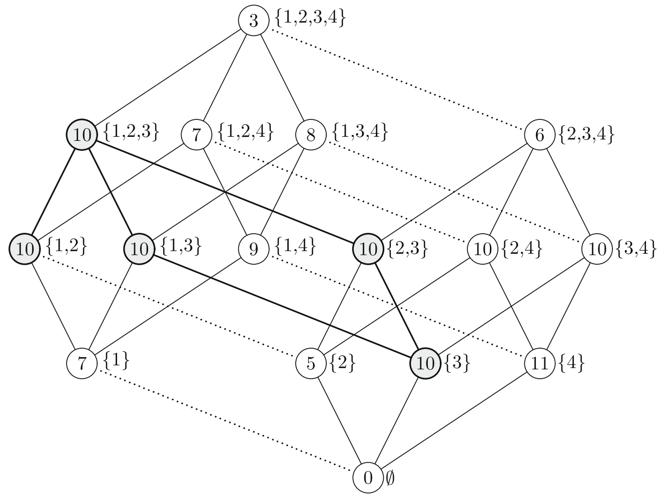

Fig. 6. Lower local maxima: {1,2}, {3}; upper local maximum: {1,2,3}; SDC (shadowed); global maximum: {4}.

By the following definitions Khachaturov [30] (see also Ref. [24]) introduced two kinds of components of subgraphs of local maxima.

Let V_0 be the subset of V corresponding to all local maxima of z and let $H_0 = (V_0, E_0, z)$ be the subgraph of G induced by V_0. This subgraph consists of at least one component. We denote the components by $H_0^j = (V_0^j, E_0^j, z)$, with $j \in J_0 = \{1, \ldots, r\}$. Note that if L_1 and L_2 are vertices in the same component then $z(L_1) = z(L_2)$.

A component H_0^j is called a *strict local maximum component* (STC) if for each $I \notin V_0^j$, for which there is an edge (I, L) with $L \in V_0^j$, we have $z(I) < z(L)$. A component H_0^j is called a *saddle local maximum component* (SDC) if for some $I \notin V_0^j$, there exists an edge (I, L) with $L \in V_0^j$ such that $z(I) = z(L)$. An example of the SDC defined by two intervals [{1,2},{1,2,3}] and [{3},{1,2,3}] in Fig. 6 is shown. The values of a submodular function in Fig. 6 are printed inside the vertices. Here a trivial STC by the vertex {4} is defined. Note that {3,4} is not a local maximum because its neighbor {4} is the global maximum with value $z(\{4\}) = 11$.

All vertices in a component H_0^j are local maxima of the same kind. Therefore, the index set J_0 of these components can be split into two subsets: J_1 being the index set of the STCs, and J_2 being the index set of the SDCs.

The following theorem of Khachaturov [30] is an application of Theorem 3 to the case of a nontrivial STC (see Fig. 7).

Theorem 5. *Let z be a submodular function on 2^N and let $\underline{L}$ and $\overline{L}$ be lower and upper maxima with $\underline{L} \subseteq \overline{L}$, both located in an STC. Then z is strictly increasing on each subchain $\emptyset \subseteq \cdots \subseteq \underline{L}$ of $[\emptyset, \underline{L}]$, constant on $[\underline{L}, \overline{L}]$, and strictly decreasing on each subchain $\overline{L} \subseteq \cdots \subseteq N$ of $[\overline{L}, N]$.*

Proof. We first show that z is strictly increasing on $[\emptyset, \underline{L}]$. The proof of the strictly decreasing case is similar. If either $\underline{L} = \emptyset$ (we obtain the decreasing case) or $|\underline{L}| = 1$, the assertion is true, since $\underline{L}$ is a local maximum of z. So, let $|\underline{L}| > 1$ and $I, J \in [\emptyset, \underline{L}]$ such that $J = I + k$, $k \in L \setminus I$. Note that $\emptyset \subseteq I \subset J \subseteq \cdots \subseteq \underline{L}$. The submodularity of z implies $z(J) + z(\underline{L} - k) \geqslant z(I) + z(\underline{L})$, or $z(J) - z(I) \geqslant z(\underline{L}) - z(\underline{L} - k)$. Since $\underline{L} \in V_0^j$ for some $j \in J_1$, $z(\underline{L}) - z(\underline{L} - k) > 0$. Hence $z(J) > z(I)$, and we have finished the proof of the strictly increasing case.

The property that z is constant on $[\underline{L}, \overline{L}]$ follows from Corollary 4. $\square$

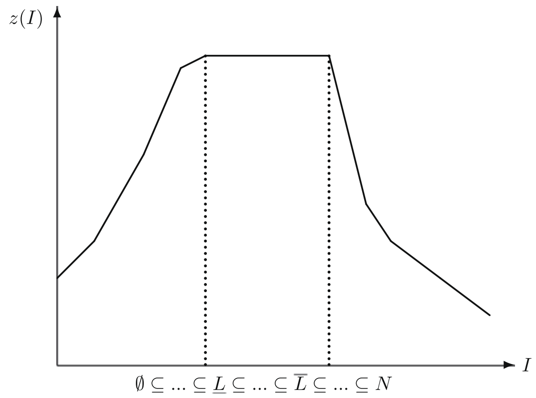

Fig. 7. The behaviour of a submodular function on a chain with lower and upper local maxima (Khachaturov's theorem).

Note that $\underline{L}$ and $\overline{L}$ need not be lower and upper maxima in Theorem 5. It is clear from the proof of Theorem 5 that any pair of embedded local maxima L_1 and L_2 located on a chain $\emptyset \subseteq \cdots \subseteq L_1 - i \subset L_1 \subseteq \cdots \subseteq L_2 \subset L_2 + k \subseteq \cdots \subseteq N$ such that $z(L_1 - i) < z(L_1)$ and $z(L_2 + k) < z(L_2)$ will imply that z is strictly increasing on each subchain $\emptyset \subseteq \cdots \subseteq L_1 - i \subset L_1$ and strictly decreasing on each subchain $L_2 \subset L_2 + k \subseteq \cdots \subseteq N$. We call such a local maxima *boundary local maxima*. In other words, a boundary local maximum is connected with vertices outside the component.

Lemma 6. *Let $L \in V_0^j$ for some $j \in J_1$, and let I satisfy $z(I) = z(L)$ with $(I, L) \in E$. Then $I \in V_0^j$ for the same $j \in J_1$.*

Proof. Let $L \in V_0^j$ for some $j \in J_1$. If $I \notin V_0^j$, then $z(I) < z(L)$, since $(I, L) \in E$ and L is a local maximum of the STC. □

Khachaturov [30] has observed that any global maximum belongs to a STC.

Theorem 7. *Let S be a global maximum of the submodular function z defined on 2^N. Then $S \in V_0^j$ for some $j \in J_1$.*

Proof. Suppose, to the contrary, that $S \in V_0^i$ with $i \in J_2$. Then there exists an $I \in V \setminus V_0$, adjacent to some $J \in V_0^i$ with $z(I) = z(J)$. This I is not a local maximum and hence, I has an adjacent vertex M with $z(M) > z(I)$. Thus, $z(S) = z(J) = z(I) < z(M)$, contradicting the assumption that S is a global maximum of z. □

Theorem 7 implies that we may restrict the search for a global maximum of a submodular function z to STC's. Based on Corollary 4, and definitions of strict and saddle components we can represent each component of local maxima as a maximal connected set of intervals whose end points are lower and upper local maxima.

3. Selection rules: An old proof

There are two "selection rules" [40,19,2], that can be used to eliminate certain subsets from $[\emptyset, N]$ when determining a global maximum of a submodular function.

Theorem 8. *Let z be a submodular function on $[\emptyset, N]$ and V_0^j with $j \in J_0$ be the components of local maxima. Then the following assertions hold.*

(a) *First Strict Selection Rule (FSSR).If for some T_1 and T_2 with $\emptyset \subseteq T_1 \subset T_2 \subseteq N$ we have $z(T_1) > z(T_2)$, then $V_0^j \cap [T_2, T] = \emptyset$ for all $j \in J_0$.*
(b) *Second Strict Selection Rule (SSSR).If for some S_1 and S_2 with $\emptyset \subseteq S_1 \subset S_2 \subseteq N$ we have $z(S_1) < z(S_2)$, then $V_0^j \cap [S, S_1] = \emptyset$ for all $j \in J_0$.*

Proof. We prove case (a) because a proof of case (b) is similar. Let us consider a chain $\emptyset \subseteq \cdots \subseteq T_1 \subset T_2 \subseteq L \subseteq T \subset \cdots \subset N$ with $L \in V_0^j \cap [T_2, T] \neq \emptyset$ for some $j \in J_0$. Applying Theorem 3 to the subchain $\emptyset \subseteq \cdots \subseteq T_1 \subset T_2 \subseteq L$ we have $z(\emptyset) \leqslant \cdots \leqslant z(T_1) \leqslant z(T_2) \leqslant z(L)$ which contradicts $z(T_1) > z(T_2)$. □

This theorem states that by applying the strict rules we do not exclude any local maximum. In other words, we preserve all local maxima. In the following theorem of Khachaturov [30], we will see that applying selection rules with nonstrict inequalities (nonstrict rules) will preserve at least one local maximum of each STC. We will call such a maximum a *representative* of the STC.

Theorem 9. *Let z be a submodular function on $[S, T] \subseteq [\emptyset, N]$ and for every $j \in J_1, V_0^j \cap [S, T] \neq \emptyset$. Then the following assertions hold.*

(a) *First Selection Rule (FSR). If for some T_1 and T_2 with $S \subseteq T_1 \subset T_2 \subseteq T$ holds that $z(T_1) \geqslant z(T_2)$, then $V_0^j \cap ([S, T] \setminus [T_2, T]) \neq \emptyset$ for all $j \in J_1$.*
(b) *Second Selection Rule (SSR). If for some S_1 and S_2 with $S \subseteq S_1 \subset S_2 \subseteq T$ holds that $z(S_1) \leqslant z(S_2)$, then $V_0^j \cap ([S, T] \setminus [S, S_1]) \neq \emptyset$ for all $j \in J_1$.*

Proof. We prove case (a) because the proof of case (b) is similar. Let us consider two cases:

Case 1: $z(T_1) > z(T_2)$. Theorem 8 implies that $V_0^j \cap [T_2, T] = \emptyset$ for all $j \in J_0 = J_1 \cup J_2$. Since for every $j \in J_1, V_0^j \cap [S, T] \neq \emptyset$ and $[T_2, T] \subset [S, T]$ we have $([S, T] \setminus [T_2, T]) \cap V_0^j \neq \emptyset$ for all $j \in J_1$.
Case 2: $z(T_1) = z(T_2)$. If we can construct a chain through two boundary local maxima L_1 and L_2 that also contains T_1 and T_2, then we have just two possibilities:
(1) $L_1 \subseteq T_1 \subset T_2 \subseteq L_2$;
(2) all others.

Each case of the possibility (2) contradicts Theorem 5. Therefore, $L_1 \subseteq T_1 \subset T_2 \subseteq L_2$, and $L_1 \subseteq T_1 \in ([S, T] \setminus [T_2, T]) \cap V_0^j \neq \emptyset$ for all $j \in J_1$. □

In Section 6, we will give an example of the SPLP in which by application of a nonstrict selection rule we discard the local minimum $\{2, 4\}$ of the corresponding supermodular function. This local minimum is an analogue of the trivial SDC for the corresponding supermodular function.

By applying Theorem 9a (respectively, 9b) we can discard $2^{|T \setminus T_2|}$ (respectively, $2^{|S_1 \setminus S|}$) subsets of interval $[T_2, T]$ (respectively, $[S, S_1]$) because this interval does not include a local maximum of any STC from $[S, T]$. If $T_1 = S$ and $T_2 = S + i$ then in case of Theorem 9a the interval $[S + i, T]$ can be discarded. If $S_1 = T - i$ and $S_2 = T$ then in case of Theorem 9b the interval $[S, T - i]$ can be discarded. These two special cases are important because we may exclude a *half subinterval* of the current interval while we preserve at least one representative from each STC. Based on the last special cases of selection rules, we present Cherenin's Preliminary Preservation (Dichotomy) Algorithm for the maximization of submodular functions in Section 5. Before we present the Dichotomy Algorithm we give in Theorem 10 an alternative proof of the correctness of these special cases of selection rules which is based only on Lemma 6, the definitions of a STC and a submodular function z. This proof shows that in case of submodular functions the definition of a STC is an insightful notion for understanding the correctness of Cherenin's Dichotomy Algorithm. Therefore, it is not necessary to use all the statements of the previous section in order to justify both prime rules. In the next section, we present a revision and a simple justification of the same rules.

Theorem 10. *Let z be a submodular function on 2^N. Suppose that for $\emptyset \subseteq S \subset T \subseteq N$ and for every $j \in J_1$, $V_0^j \cap [S,T] \neq \emptyset$. Then the following assertions hold.*

(a) *First Prime Selection Rule (FPSR). If for some $i \in T \setminus S$ it holds that $z(S+i) \leqslant z(S)$, then $[S, T-i] \cap V_0^j \neq \emptyset$ for all $j \in J_1$.*
(b) *Second Prime Selection Rule (SPSR). If for some $i \in T \setminus S$ it holds that $z(T-i) \leqslant z(T)$, then $[S+i, T] \cap V_0^j \neq \emptyset$ for all $j \in J_1$.*

Proof. We prove only part (a). The proof of part (b) is similar.

(a) Let $z(S+i) \leqslant z(S)$ for some $i \in T \setminus S$ and let $G \in V_0^j \cap [S,T]$ for any $j \in J_1$. Then $S \subset G$.

Case 1: $i \in G$. From the definition of submodularity applied to the sets $G-i$ and $S+i$

$$z(G-i) + z(S+i) \geqslant z(G \cup S + i) + z(S) \Rightarrow$$
$$z(G-i) - z(G \cup S + i) \geqslant z(S) - z(S+i) \geqslant 0 \Rightarrow$$
$$z(G-i) \geqslant z(G \cup S + i) = z(G) \Rightarrow (G \text{ is a local maximum})$$
$$z(G-i) = z(G). G \in V_0^j \Rightarrow (\text{by Lemma 6})$$
$$G - i \in V_0^j \Rightarrow G - i \in V_0^j \cap [S, T-i] \Rightarrow V_0^j \cap [S, T-i] \neq \emptyset.$$

Case 2: $i \notin G$.

$$i \notin G \Rightarrow G \in V_0^j \cap [S, T-i] \Rightarrow V_0^j \cap [S, T-i] \neq \emptyset. \qquad \square$$

Theorem 10a states that if $z(S+i) - z(S) \leqslant 0$ for some $i \in T \setminus S$, then by preserving the interval $[S, T-i]$ we preserve at least one strict local maximum from each STC, and hence we preserve at least one global maximum from each STC containing a global maximum. Therefore, in this case it is possible to exclude exactly the whole interval $[S+i, T]$ of $[S,T]$ from consideration when searching for a global maximum of the submodular function z on $[S,T] \subseteq [\emptyset, N]$. For example, see Fig. 8, if $z(\emptyset) - z(\emptyset + 1) \geqslant 0$, then the interval $[\{1\},\{1,2,3,4\}]$ can be excluded, i.e., the interval $[\emptyset, \{2,3,4\}]$ will be preserved (FPSR). If $z(\{1,2,3,4\}) - z(\{1,2,3,4\} - 1) \geqslant 0$, then the interval $[\emptyset, \{2,3,4\}]$ can be excluded, i.e., the interval $[\{1\},\{1,2,3,4\}]$ will be preserved (SPSR).

If the prime rules are not applicable it will be useful to discard less than a half subinterval of the current interval $[S,T] \subseteq [\emptyset, N]$. In the following section, we further relax most of the theoretical results presented in the previous sections of this paper with the purpose to show the correctness of all selection rules and their revisions (preservation rules) based only on the definitions of submodularity and the maximum value $z^*[S,T]$ of the function z on the interval $[S,T] \subseteq [\emptyset, N]$.

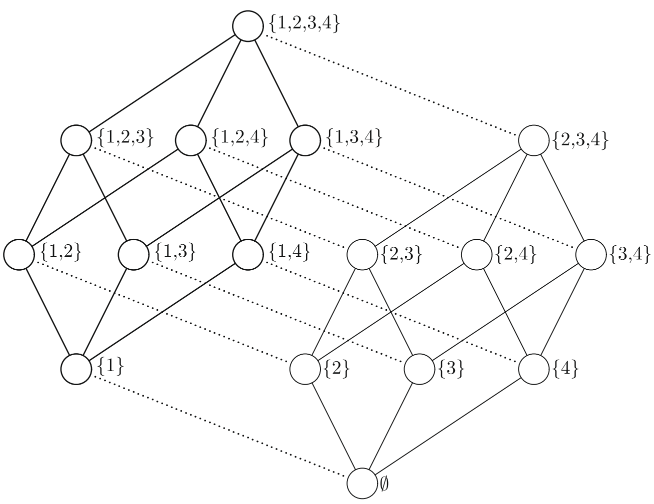

Fig. 8. Example of prime selection rules.

4. Preservation rules: Revision and a simple justification

In the following theorem, we give an important interpretation of the submodularity property which is based on two pairs of submodular function values [26]. For this purpose we introduce an *upper (respectively, lower) partition* of the current interval $[S,T]$ by an inner vertex $Q : S \subset Q \subset T$ into two parts $[Q,T]$ and $[S,T] \setminus [Q,T]$ (respectively, $[S,Q]$ and $[S,T] \setminus [S,Q]$). In terms of the maximum values of the function z defined on each of two parts of the above mentioned partitions a special case of submodularity can be read as either $z^*([S,T] \setminus [Q,T]) + z(Q) \geqslant z(S) + z^*[Q,T]$ or $z^*([S,T] \setminus [S,Q]) + z(Q) \geqslant z^*[S,Q] + z(T)$.

Here, the both maximal values of a submodular function and their arguments (vertices) involved in each of the above indicated inequalities are *unknown*. In other words, Theorem 11 establishes a relationship of the difference between the unknown optimal values of z on the two parts of the partition, for example, $([S,T] \setminus [Q,T])$ and $[Q,T]$ of $[S,T]$ and the corresponding difference $z(S) - z(Q)$ (see the FSR in Theorem 9); a symmetrical result is obtained for the SSR.

Theorem 11. *Let z be a submodular function on the interval $[S,T] \subseteq [\emptyset, N]$. Then for any Q such that $S \subset Q \subset T$ the following assertions hold.*

(a) $z^*([S,T] \setminus [Q,T]) - z^*[Q,T] \geqslant z(S) - z(Q)$.
(b) $z^*([S,T] \setminus [S,Q]) - z^*[S,Q] \geqslant z(T) - z(Q)$.

Proof. We prove only case (a) because the proof of case (b) is similar. Let $z^*[Q,T] = z(Q \cup J)$ with $J \subseteq T \setminus Q$. Define $I = S \cup J$. Then $I \in [S,T] \setminus [Q,T]$ since $Q \setminus S \not\subseteq I$. We have that $z^*([S,T] \setminus [Q,T]) - z(S) \geqslant z(I) - z(S) = z(S \cup J) - z(S)$. From the submodularity of z we have $z(S \cup J) - z(S) \geqslant z(Q \cup J) - z(Q)$. Therefore, $z^*([S,T] \setminus [Q,T]) - z(S) \geqslant z^*[Q,T] - z(Q)$. $\square$

Theorem 11 is a reformulation of Cherenin's Theorem 3 in terms of global maxima compared to local maxima analyzed in Cherenin–Khachaturov's Theorems 3, 5, 7–9, Corollary 4, and Lemma 6. Theorem 11 stating that the difference of values of a submodular function on any pair of embedded subsets is a lower bound for the difference between the optimal values of z on the two parts of the partition defined by this pair of embedded subsets. The theorem can be used to decide either in which part of the partition $([S,T] \setminus [Q,T])$ and $[Q,T]$ of $[S,T]$ a global maximum of z is located, or to partition further the set of $([S,T] \setminus [Q,T])$ such that each its subset will contain at least one global maximum (Corollary 18).

We may represent the partition of interval $[S,T]$ from Theorem 11 by means of its proper subintervals as follows:

(a) upper partition $$[S,T] \setminus [Q,T] = \bigcup_{i \in Q \setminus S} [S, T-i]$$

and

(b) lower partition $$[S,T] \setminus [S,Q] = \bigcup_{i \in T \setminus Q} [S+i, T].$$

Examples of upper and lower partitions in Figs. 9 and 10 are shown.

A disadvantage of representations (a) and (b) is a non-empty overlapping of each pairwise distinct intervals involved in these representations. As easy to see in Figs. 11 and 12 we can avoid such an overlapping by representing the remaining parts $([S,T] \setminus [Q,T])$

and $([S,T] \setminus [S,Q])$ with a sequence of "parallel" non-overlapping intervals. For example, the difference $[\emptyset, \{1,2,3,4\}] \setminus [\{1,2,3\}, \{1,2,3,4\}] = [\{1,2\},\{1,2,4\}] \cup [\{1\},\{1,3,4\}] \cup [\emptyset,\{2,3,4\}]$ (see Figs. 9 and 12), and the difference $[\emptyset,\{1,2,3,4\}] \setminus [\emptyset,\{1,2\}] = [\{3\},\{1,2,3\}] \cup [\{4\},\{1,2,3,4\}]$ (see Figs. 10 and 11).

The sequence of non-overlapping intervals can be created by the following iterative procedure. We will use the value $d = \dim([U,W])$ of the *dimension* of an interval $[U,W]$ interpreted as the corresponding subspace of the Boolean space $\{0,1\}^n$ which is another representation of the interval $[\emptyset, N]$.

If we have discarded the k-dimensional subinterval $[Q,T]$ in the upper partition of the interval $[S,T]$, then the first non-overlapping interval $[U_1,W_1]$ is the k-dimensional subinterval of the $(k+1)$-dimensional interval $[U_1,T] = [Q,T] \cup [U_1,W_1]$. In other words, the first non-overlapping interval $[U_1,W_1]$ is the k-dimensional complement to the $(k+1)$-dimensional interval $[U_1,T]$ such that $[U_1,W_1] = [U_1,T] \setminus [Q,T]$. The second non-overlapping interval $[U_2,W_2]$ is the $(k+1)$-dimensional subinterval of the $(k+2)$-dimensional interval $[U_2,T] = [U_1,T] \cup [U_2,W_2]$, and $[U_2,W_2] = [U_2,T] \setminus [U_1,T]$, etc. Finally, $[U_q,W_q] = [U_q,T] \setminus [U_{(q-1)},T]$. The number q of the non-overlapping intervals in the upper partition is equal to $n-k$, where $k = \dim[Q,T]$. The representation of a lower

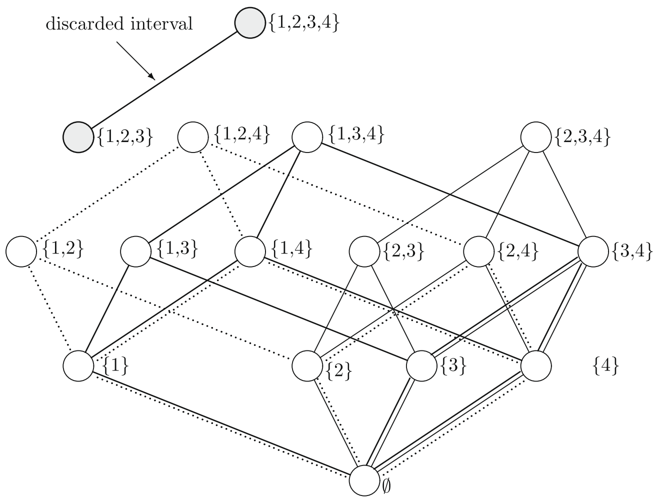

Fig. 9. A representation of the upper partition of the interval $[S,T] = [\emptyset,\{1,2,3,4\}]$ with $Q \setminus S = \{1,2,3\}$.

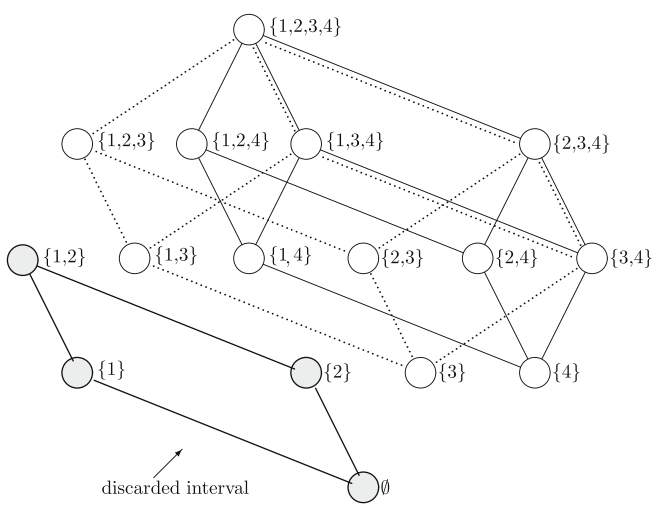

Fig. 10. A representation of the lower partition by $Q = \{1,2\}$ for the interval $[S,T] = [\emptyset,\{1,2,3,4\}]$ with $T \setminus Q = \{3,4\}$.

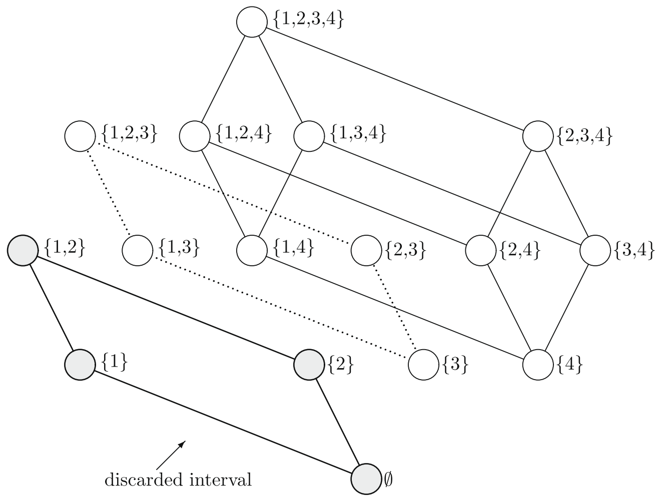

Fig. 11. The non-overlapping representation of the lower partition by parallel intervals $[\{3\},\{1,2,3\}]$ and $[\{4\},\{1,2,3,4\}]$.

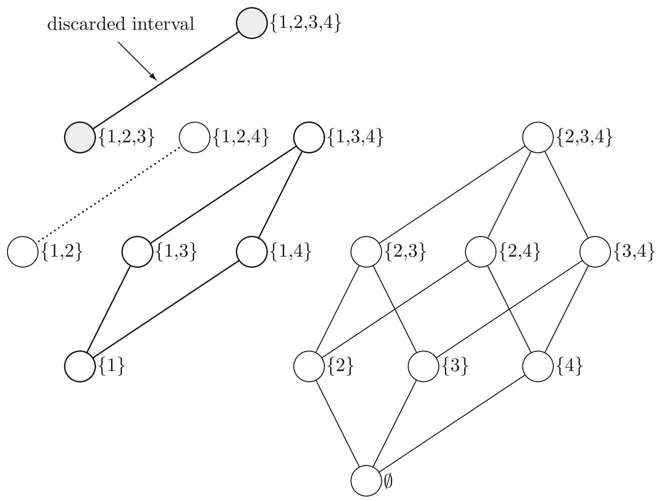

Fig. 12. The nonoverlapping representation of the upper partition by the parallel intervals $[\{1,2\},\{1,2,4\}]$, $[\{1\},\{1,3,4\}]$, and $[\emptyset,\{2,3,4\}]$.

partition by the sequence of non-overlapping intervals can be described in similar lines. Note that the above indicated representation of lower (upper) partition by a sequence of non-overlapping intervals has the *minimum number of mutually disjoint intervals.*

For example (see Fig. 12), the complement interval to $[\{1,2,3\},\{1,2,3,4\}]$ is $[\{1,2\},\{1,2,4\}]$ since $[\{1,2\},\{1,2,4\}] \cup [\{1,2,3\},\{1,2,3,4\}] = [\{1,2\},\{1,2,3,4\}]$, and the complement to $[\{1,2\},\{1,2,3,4\}]$ is $[\{1\},\{1,3,4\}]$. Finally, the complement to $[\{1\},\{1,2,3,4\}]$ is $[\emptyset,\{2,3,4\}]$.

If, in Theorem 11, we replace Q by $S+k$ in part (a), and Q by $T-k$ in part (b), we obtain the following reformulation of the prime rules stated in Theorem 10.

Corollary 12. *Let z be a submodular function on the interval $[S,T] \subseteq [\emptyset,N]$ and let $k \in T \setminus S$. Then the following assertions hold.*

(a) $z^*[S,T-k] - z^*[S+k,T] \geqslant z(S) - z(S+k)$.
(b) $z^*[S+k,T] - z^*[S,T-k] \geqslant z(T) - z(T-k)$.

By adding the condition $z(S) - z(S+k) \geqslant 0$ to part (a) and the condition $z(T) - z(T-k) \geqslant 0$ to part (b) of Corollary 12 we obtain another form (see Corollary 13) of two prime rules from Theorem

10 for preserving subintervals containing at least one global maximum of z on $[S,T]$.

Corollary 13. *Let z be a submodular function on the interval $[S,T] \subseteq [\emptyset,N]$ and $k \in T \setminus S$. Then the following assertions hold.*

(a) *First Preservation (FP) Rule. If $z(S) \geqslant z(S+k)$, then $z^*[S,T] = z^*[S,T-k] \geqslant z^*[S+k,T]$.*
(b) *Second Preservation (SP) Rule. If $z(T) \geqslant z(T-k)$, then $z^*[S,T] = z^*[S+k,T] \geqslant z^*[S,T-k]$.*

Proof. (a) From Corollary 12a we have $z^*[S,T-k] - z^*[S+k,T] \geqslant z(S) - z(S+k)$. By assumption $z(S) - z(S+k) \geqslant 0$. Hence, $z^*[S,T] = z^*[S,T-k] \geqslant z^*[S+k,T]$. (b) The proof is similar. □

From the calculation point of view these rules are the same as in Theorem 9 but Theorem 10 is more powerful than Corollary 13. In Theorem 10, we preserve at least one strict local maximum from each STC, and hence one global maximum from each STC that contains global maxima. Corollary 13 only states that we preserve at least one global maximum. However, we can use Corollary 13 for constructing some extension of the preservation rules.

For $\varepsilon \geqslant 0$, we may consider the problem of finding an approximate solution $J \in [S,T]$ such that $z^*[S,T] \leqslant z(J) + \varepsilon$; J is called an *ε–maximum* of z on $[S,T]$. The following corollary presents an extension of the rules from Corollary 13 which is appropriate to the problem of ε-maximization.

Corollary 14. *Let z be a submodular function on the interval $[S,T] \subseteq [\emptyset,N]$, and $k \in T \setminus S$. Then the following assertions hold.*

(a) *First θ-Preservation(θ-FP) Rule. If $z(S) - z(S+k) = \theta < 0$, then $z^*[S,T] - z^*[S,T-k] \leqslant -\theta$, which means that $[S,T-k]$ contains a $|\theta|$-maximum of $[S,T]$.*
(b) *Secondη-Preservation (η-SP) Rule. If $z(T) - z(T-k) = \eta < 0$, then $z^*[S,T] - z^*[S+k,T] \leqslant -\eta$, which means that $[S+k,T]$ contains a $|\eta|$-maximum of $[S,T]$.*

Proof. The proof of part (a) is as follows:

Case 1. If $z^*[S,T] = z^*[S,T-k]$ then $z^*[S,T-k] - z^*[S,T-k] \leqslant -\theta$ or $z^*[S,T] - z^*[S,T-k] \leqslant -\theta$.
Case 2. If $z^*[S,T] = z^*[S+k,T]$, then from Corollary 12a follows that $z^*[S,T-k] - z^*[S+k,T] \geqslant \theta$ or $z^*[S,T-k] - z^*[S,T] \geqslant \theta$. Hence $z^*[S,T] - z^*[S,T-k] \leqslant -\theta$. The proof of (b) is similar. □

5. The preliminary preservation algorithm (PPA)

By means of Corollary 13 it is often possible to exclude a large part of $[\emptyset,N]$ from consideration when determining a global maximum of z on $[\emptyset,N]$. The so called *Preliminary Preservation Algorithm (PPA)* [25] determines the smallest subinterval $[S,T]$ of $[\emptyset,N]$ containing a global maximum of z, by using the preservation rules of Corollary 13.

We call the PPA the *dichotomy algorithm* because in every successful step it halves the current domain of a submodular function.

Let $[S,T]$ be an interval. For each $i \in T \setminus S$, define $\delta^+(S,T,i) = z(T) - z(T-i)$ and $\delta^-(S,T,i) = z(S) - z(S+i)$; moreover, define $\delta^+_{\max}(S,T) = \max\{\delta^+(S,T,i) | i \in T \setminus S\}$, $r^+(S,T) = \min\{r | \delta^+(S,T,r) = \delta^+_{\max}(S,T)\}$. Similarly, for $\delta^-(S,T,i)$ define $\delta^-_{\max}(S,T) = \max\{\delta^-(S,T,i)) | i \in T \setminus S\}$, $r^-(S,T) = \min\{r | \delta^-(S,T,r) = \delta^-_{\max}(S,T)\}$. If no confusion is likely, we briefly write r^-, r^+, δ^-, δ^+ instead of $r^-(S,T)$, $r^+(S,T)$, $\delta^-_{\max}(S,T)$, and $\delta^+_{\max}(S,T)$, respectively. See Figs. 13.

Each time that either S or T are updated during the execution of the PPA, the conditions of Corollary 13 remain satisfied, and therefore $z^*[S,T] = z^*[U,W]$ remains invariant at each step of the PPA. At the end of the algorithm we have that $\max\{\delta^+, \delta^-\} < 0$, and therefore $z(S) < z(S+i)$ and $z(T) < z(T-i)$ for each $i \in T \setminus S$. Hence Corollary 13 cannot be applied to further reduce the interval $[S,T]$ without violating $z^*[S,T] = z^*[U,W]$. Note that this remark shows the correctness of the procedure PP(.).

If we replace in the PPA the rules of Corollary 13 by those of Corollary 14 we obtain an ε-maximization variant of the PPA. In this case the output of the ε-PPA will be a subinterval $[S,T]$ of $[U,W]$ such that $z^*[U,W] - z^*[S,T] \leqslant \varepsilon$ with postconditions $z(S) + \varepsilon < z(S+i)$ and $z(T) + \varepsilon < z(T-i)$ for each $i \in T \setminus S$.

The following theorem can also be found in [25]. It provides an upper bound for the worst case complexity of the PPA; the complexity function is dependent only on the number of comparisons of pairs of values for z.

Theorem 15. *The time complexity of the PP algorithm procedure is at most* $O(n^2)$.

Note that if the PPA terminates with $S = T$, then S is a global maximum of z. Any submodular function z on $[U,W]$ for which the PP algorithm returns a global maximum for z is called a *PP-function*.

An example of a set of PP-functions $\mathscr{P}$ is shown in Fig. 14. Here, for all vertices without prespecified values of $z(I)$ can be assigned an arbitrary value of $z(I)$ such that each corresponding set function $z(I) \in \mathscr{P}$ defined on the whole weighted graph G will be submodular. For example, if for all vertices without prespecified values of $z(I)$ in Fig. 14 we set $z(I) = a$, then for each constant $a : 2 \leqslant a \leqslant 3$ the corresponding function z is a submodular PP-function. It means that by applying the Dichotomy algorithm we have found an optimal solution to the PSMF for all PP-functions defined by a constant a.

Corollary 16 describes in terms of STCs some properties of the variables S and T during the iterations of the PPA. A representative $L^j_1 \in V^j_0$ with $j \in J_1$ which will be preserved through all iterations during the execution of the PPA by FPSR ($L^j_1 \in V^j_0 \cap [S,T-i] \neq \emptyset$ with $j \in J_1$) or SPSR ($L^j_1 \in V^j_0 \cap [S+i,T] \neq \emptyset$ with $j \in J_1$) is called a *PP-representative* of STC H^j_0 with $j \in J_1$ (see Theorem 8 and its discussion).

Corollary 16. *If z is a submodular PP-function on $[U,W] \subseteq [\emptyset,N]$, then at each iteration of the PPA $S \subseteq \cap_{j \in J_1} L^j_1$ and $T \supseteq \cup_{j \in J_1} L^j_1$.*

Procedure PP(U,W,S,T)
Input: A submodular function z on the subinterval $[U,W]$ of $[\emptyset,N]$
Output: A subinterval $[S,T]$ of $[U,W]$ such that $z^*[S,T] = z^*[U,W]$, $z(S) < z(S+i)$ and $z(T) < z(T-i)$ for each $i \in T \setminus S$.
begin
$S \leftarrow U$; $T \leftarrow W$;
Step 1: if $S = T$
then goto Step 4;
Step 2: Calculate δ^+ and r^+;
if $\delta^+ \geq 0$ (Corollary 13b)
then begin call PP$(S+r^+, T; S, T)$
goto Step 4
end;
Step 3: Calculate δ^- and r^-;
if $\delta^- \geq 0$ (Corollary 13a)
then begin call PP$(S, T-r^-; S, T)$
goto Step 4
end;
Step 4:
end;

Fig. 13. The Dichotomy (preliminary preservation) algorithm.

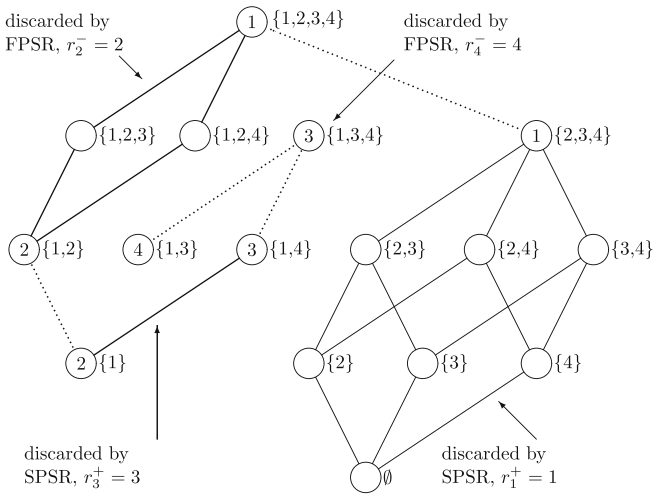

Fig. 14. The idea of the Dichotomy algorithm: $z(\{1,3\}) = 4$ is the global maximum for all submodular functions from the subclass of $\mathscr{P}$.

Proof. Theorem 10a says that if $z(S+i) - z(S) \leqslant 0$ for some $i \in T \setminus S$, then by preserving the interval $[S, T-i]$ we preserve at least one PP-representative L_1^j from each STC H_0^j, and hence $i \notin L_1^j$. In case of Theorem 10b we preserve PP-representatives L_1^j such that $i \in L_1^j$ for all STCs in $[S, T]$. Therefore, $i \in S \subseteq \cap_{j\in J_1} L_1^j$ and $T \supseteq \cup_{j\in J_1} L_1^j$. □

The following theorem gives a property of PP-functions in terms of STCs.

Theorem 17. *If z is a submodular PP-function on* $[U, W] \subseteq [\emptyset, N]$, *then* $[U, W]$ *contains exactly one STC.*

Proof. From $\cap_{j\in J_1} L_1^j \supseteq S = T \supseteq \cup_{j\in J_1} L_1^j$ we obtain $\cap_{j\in J_1} L_1^j = \cup_{j\in J_1} L_1^j$ or $L_1^j = L$ for all $j \in J_1$. □

Note that not each submodular function with exactly one STC on $[\emptyset, N]$ is a PP-function. For example, let $N = \{1,2,3\}$ and consider the submodular function z defined by $z(I) = 2$ for any $I \in [\emptyset, \{1,2,3\}] \setminus (\{\emptyset\} \cup \{1,2,3\})$ and $z(I) = 1$ for $I \in (\{\emptyset\} \cup \{1,2,3\})$. The vertex set of the unique STC defined by this function can be represented by $[\{1\},\{1,2\}] \cup [\{1\},\{1,3\}] \cup [\{2\},\{1,2\}] \cup [\{2\},\{2,3\}] \cup [\{3\},\{1,3\}] \cup [\{3\},\{2,3\}]$. The PPA terminates with $[S, T] = [\emptyset, \{1,2,3\}]$ and so, z is not a PP-function.

6. Non-binary branching rules

Usually in BnB type algorithms we use a *binary* branching rule by which the original set $[S, T]$ of feasible solutions will be split by an element k into two subsets $[S+k, T]$ and $[S, T-k]$. Let us consider an interval $[S, T]$ for which the postconditions of the PPA are satisfied, i.e., $z(S) < z(S+i)$ and $z(T) < z(T-i)$ for each $i \in T \setminus S$. Thus, the PPA cannot make the interval $[S, T]$ smaller. By using Corollary 18 we can sometimes find two subintervals $[S, T-k_1]$ and $[S, T-k_2]$ such that the postconditions of the PPA algorithm for each of these intervals are violated.

Corollary 18. *Let z be a submodular function on the interval* $[S, T] \subseteq [\emptyset, N]$ *and let* $k_1, k_2 \in T \setminus S$ *with* $k_1 \neq k_2$. *Then the following assertions hold.*

(a) $\max\{z^*[S, T-k_1], z^*[S, T-k_2]\} - z^*[S+k_1+k_2, T] \geqslant z(S) - z(S+k_1+k_2)$.
(b) $\max\{z^*[S+k_1, T], z^*[S+k_2, T]\} - z^*[S, T\setminus\{k_1,k_2\}] \geqslant z(T) - z(T\setminus\{k_1,k_2\})$.

Proof. We prove only part (a) because the proof of part (b) is similar. Replace Q by $S+k_1+k_2$ in Theorem 11a. Then, $z^*([S,T] \setminus [Q,T]) - z^*[Q,T] = z^*\left(\bigcup_{i\in Q\setminus S}[S, T-i]\right) - z^*[Q,T] = z^*([S,T-k_1] \cup [S,T-k_2]) - z^*[S+k_1+k_2, T] = \max\{z^*[S,T-k_1], z^*[S,T-k_2]\} - z^*[S+k_1+k_2, T] \geqslant z(S) - z(Q) = z(S) - z(S+k_1+k_2)$. □

In the case that $z(S) - z(S+k_1+k_2) \geqslant 0$ we can discard the interval $[S+k_1+k_2, T]$ and continue the search for an optimal solution by applying the PPA separately to each remaining interval $[S, T-k_1]$ and $[S, T-k_2]$, which are obtained by subtracting an element k_i from T. The symmetrical case will be obtained if $z(T) - z(T\setminus\{k_1,k_2\}) \geqslant 0$. Corollary 18 can easily be generalized to the case of m-ary branching by elements $k_1, k_2, \ldots, k_m$ with $m \leqslant |T \setminus S|$.

We conclude this section with a simple plant location example borrowed from Boffey [10] which the data are presented in Table 1.

For solving the SPLP it suffices to solve the problem $\min\{z(I) | I \in [\emptyset, N]\} = z^*[\emptyset, N] = z(G)$ with $N = \{1,2,3,4\}, m = 5$ and

$$z(I) = \sum_{i\in I} f_i + \sum_{j=1}^{m} \min_{i\in I} c_{ij}.$$

As usual for the SPLP, f_i is the fixed cost of opening a plant at location i, c_{ij} is the cost of satisfying the demand of customer j by plant i, and $z(I)$ is a supermodular function. Note that if in the definition of a submodular function we change the sign "$\geqslant$" to the opposite sign "$\leqslant$" then we obtain the definition of a supermodular function. For sake of completeness, let us show that $z(I)$ of the SPLP is supermodular.

Lemma 19. *The objective* $z(I)$ *of SPLP is supermodular.*

Proof. According to Theorem 1(i) a function is supermodular if

$$z(A) + z(B) \leqslant z(A\cup B) + z(A\cap B), \quad \forall A, B \subseteq N.$$

We use the following representation of this definition

$$z(A) + z(B) - z(A\cup B) - z(A\cap B) \leqslant 0, \quad \forall A, B \subseteq N.$$

Substituting

$$z(I) = \sum_{i\in I} f_i + \sum_{j=1}^{m} \min_{i\in I} c_{ij}$$

gives

$$\sum_{i\in A} f_i + \sum_{j=1}^{m} \min_{i\in A} c_{ij} + \sum_{i\in B} f_i + \sum_{j=1}^{m} \min_{i\in B} c_{ij} - \sum_{i\in A\cup B} f_i - \sum_{j=1}^{m} \min_{i\in A\cup B} c_{ij} - \sum_{i\in A\cap B} f_i - \sum_{j=1}^{m} \min_{i\in A\cap B} c_{ij}$$
$$= \left[\sum_{i\in A} f_i + \sum_{i\in B} f_i - \sum_{i\in A\cup B} f_i - \sum_{i\in A\cap B} f_i\right] + \sum_{j=1}^{m}[(\min_{i\in A} c_{ij} - \min_{i\in A\cup B} c_{ij}) + (\min_{i\in B} c_{ij} - \min_{i\in A\cap B} c_{ij})].$$

Table 1
The data of the SPLP

Location		Delivery cost to site				
i	f_i	$j = 1$	$j = 2$	$j = 3$	$j = 4$	$j = 5$
1	7	7	15	10	7	10
2	3	10	17	4	11	22
3	3	16	7	6	18	14
4	6	11	7	6	12	8

Note that

$$\left[\sum_{i\in A} f_i + \sum_{i\in B} f_i - \sum_{i\in A\cup B} f_i - \sum_{i\in A\cap B} f_i\right] = 0,$$

hence it is enough to show that for each $j = 1, \dots, m$

$$[(\min_{i\in A} c_{ij} - \min_{i\in A\cup B} c_{ij}) + (\min_{i\in B} c_{ij} - \min_{i\in A\cap B} c_{ij})] \leqslant 0.$$

Let us consider two cases. Case 1: $\min_{i\in A\cup B} c_{ij} = c_{aj}$ for some $a \in A$. Then $\min_{i\in A} c_{ij} = \min_{i\in A\cup B} c_{ij}$ and $\min_{i\in B} c_{ij} \leqslant \min_{i\in A\cap B} c_{ij}$.

Case 2: $\min_{i\in A\cup B} c_{ij} = c_{bj}$ for some $b \in B$. Then $\min_{i\in B} c_{ij} = \min_{i\in A\cup B} c_{ij}$ and $\min_{i\in A} c_{ij} \leqslant \min_{i\in A\cap B} c_{ij}$. □

We use this example for illustrating that the supermodular function defined by data from Table 1 is not a PP-function. Of course, here we mean the corresponding definition of a PP-function obtained by replacing the definitions of local, global maxima of a submodular function by the local, global minima of a supermodular function. It is easy to check that this supermodular function has two trivial analogues of STCs: {1,4}, {1,3} and one trivial analogue of SDC: {2,4} (see Fig. 15).

After the first execution of *Step 3* of the PPA, we have that $[S,T] = [\{1\},\{1,2,3,4\}]$, because $\delta^+ = z(\{1,2,3,4\}) - z(\{2,3,4\}) = 0$ and $r^+ = 1$. Together with interval $[\{\emptyset\},\{2,3,4\}]$ the PPA has discarded the trivial SDC {2,4}. After the second execution of *Steps 2* and *3* the PPA terminates with interval $[S,T] = [\{1\},\{1,2,3,4\}]$, because all postconditions of the PPA are satisfied. Hence, this function is not a PP-function. A global minimum of this SPLP can be found by application the following analogue of the inequality from Corollary 18b:

$$\min\{z^*[S+k_1,T], z^*[S+k_2,T]\} - z^*[S, T\setminus\{k_1,k_2\}] \leqslant z(T) - z(T\setminus\{k_1,k_2\}).$$

Let us substitute all possible pairs $\{k_1,k_2\}$ into the right-hand side of this inequality with $S = \{1\}$ and $T = \{1,2,3,4\}$. Then, we have that only $z(\{1,2,3,4\}) - z(\{1,2,3,4\} - \{3,4\}) = 52 - 53 < 0$. Hence, we can discard the interval $[\{1\},\{1,2,3,4\} - \{3,4\}]$ and we may continue to find $z^*[\{1\},\{1,2,3,4\}]$ by solving two remaining subproblems $z^*[\{1,3\},\{1,2,3\}]$ and $z^*[\{1,4\},\{1,2,3,4\}]$ defined on "parallel" intervals $[\{1,3\},\{1,2,3\}]$ and $[\{1,4\},\{1,2,3,4\}]$ (with disjoint set of feasible solutions) instead of two corresponding subproblems $z^*[S+k_1,T] = z^*[\{1,3\},\{1,2,3,4\}]$ and $z^*[S+k_2,T] = z^*[\{1,4\},\{1,2,3,4\}]$ which have the non-empty intersection on $[\{1,3,4\},\{1,2,3,4\}]$. Each of these subproblems can be solved by the corresponding analogue of the PPA.

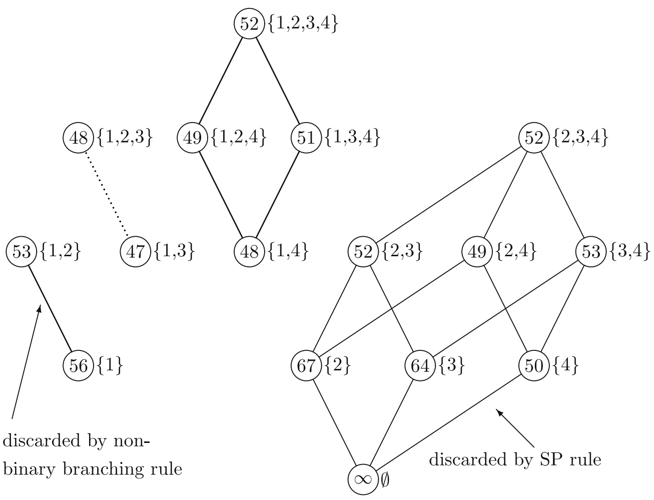

Fig. 15. The SPLP example: illustration of non-binary branching rule.

7. Concluding remarks

We have considered a submodular function z defined on the Boolean hypercube to which we can apply a classic theorem of Cherenin saying that z is quasi-concave on any chain that intersects a local maxima component. This result enables a clearer understanding of the structure of a submodular function in terms of components of the graph of local maxima. Specifically we may state that each component of the graph of local maxima is a maximal connected set of intervals whose end points are lower and upper local maxima. Cherenin's theorem provides a justification of "the method of successive calculations". This method was successfully applied to solve problems arising in railway logistics planning [12,13,40], and for constructing BnB type algorithms [30,31,19,2,24,25] for solving a number of NP-hard problems.

We have shown that if the Dichotomy algorithm (PPA) terminates with $S = T$ then the given submodular function has exactly one strict component of local maxima (STC). Hence the number of subproblems created in a branch without bounds type algorithm, which is based on the Dichotomy algorithm, can be used as an upper bound for the number of the STCs. In a similar way, an upper bound for the number of all components (STCs and SDCs) by using strict selection rules can be calculated. This information can be used for complexity analysis in terms of the number of local optima for a specific class of problems arose in practice (computational experiments).

We next proposed a reformulation of Cherenin's fundamental Theorem 3 dealing with local maxima in the form of our Theorem 11 analyzing only global maxima. Theorem 11 provides implicit enumeration bounds for a recursive implementation of any BnB procedure incorporating the Dichotomy algorithm. This reformulation is useful in three respects. Firstly we have shown that the original Cherenin's selection rules essentially are preservation rules which preserve at least one local maximum from each strict component of local maxima. Theorem 11 keeps the form of Cherenin's selection rules and explains that these rules are also preservation rules of at least one global maximum from some strict component of global maxima. Secondly it is suitable for use in ε-*optimal* procedures which obtain an approximate global maximum within specified bounds. Thirdly the theorem allows the derivation of alternatives to the prime selection rules by which we are able to discard subintervals of smaller cardinality than half original subinterval. We show that the remaining part of the current interval can be represented by a set of subintervals, some of which may include just one strict component. In other words, we try to prepare the necessary conditions for the Dichotomy algorithm to terminate on each subinterval. Moreover, Theorem 11 is based only on the definition of the maximum value of PMSF for an interval of $[\emptyset, N]$, and relaxed Cherenin–Khachaturov's theory presented in Sections 2 and 3 (which is based on notions of monotonicities on a chain, local and global maxima, strict and saddle components in the Hasse diagram).

Corollary 12 can be considered as the basis of our Data Correcting (DC) algorithm presented in Goldengorin [23] and Goldengorin et al. [25]. It states that if an interval $[S,T]$ is split into $[S, T-k]$ and $[S+k, T]$, then the difference between the submodular function values $z(S)$ and $z(S+k)$, or between the values of $z(T)$ and $z(T-k)$ is an upper bound for the difference of the (unknown!) optimal values on the two subintervals. This difference is used for 'correcting' the current data (values of a submodular function z) in the DC algorithm. Our computational experiments with the Quadratic Cost Partition Problem presented in [25,27] show that we can substantially reduce the calculation time for data correcting algorithms by recursive application of Theorem 11.

An interesting subject for future research is the investigation of the computational efficiency of m-ary branching rules (see Corollary 18) for specific problems which can be reduced to the maximization of submodular functions.

Acknowledgements

This article is dedicated to my former colleague Dr. Gert A. Tijssen who sadly passed away 18th of June 2008. He has contributed to the results presented here by lively and inspiring discussions. Thanks are due to the Editor, Professor Jean-Charles Billaut for many efforts in the reviewing process, my colleagues Professor Alex Belenky, Dr. Hans Nieuwenhuis for helpful discussions and suggestions, and the referees for their valuable comments.

References

[1] A. Ageev, M. Sviridenko, Pipage rounding: A new method of constructing algorithms with proven performance guarantee, Journal of Combinatorial Optimization 8 (2004) 307–328.
[2] A. Alcouffe, G. Muratet, Optimal location of plants, Management Science 23 (3) (1976) 267–274.
[3] Dj.A. Babayev, Comments on the note of frieze, Mathematical Programming 7 (1974) 249–252.
[4] E. Balas, M.W. Padberg, On the set covering problem, Operations Research 20 (1972) 1152–1161.
[5] C. Ballester, A. Calv-Armengol, Y. Zenou, Optimal Targets in Peer Networks. Working Paper, University of Alicante, Spain, 2005.
[6] F. Barahona, M. Grötschel, M. Jünger, G. Reinelt, An application of combinatorial optimization to statistical physics and circuit layout design, Operations Research 36 (3) (1988) 493–512.
[7] J.E. Beasley, Lagrangean heuristics for location problems, European Journal of Operational Research 65 (1993) 383–399.
[8] S. Benati, An improved Branch & Bound method for the uncapacitated competitive location problem, Annals of Operations Research 122 (2003) 43–58.
[9] O. Bilde, J. Krarup, Sharp lower bounds and efficient algorithms for the simple plant location problem, Annals of Discrete Mathematics 1 (1977) 79–97.
[10] T.B. Boffey, Graph Theory in Operations Research, Academic Press, London, UK, 1982.
[11] C.P. Chambers, F. Echenique, Supermodularity and Preferences, Research Reportm Division of the Humanities and Social Sciences, California Institute of Technology, USA, 2007.
[12] V.P. Cherenin, Mechanization of calculations for drawing up train gathering plans, Tekhnika Zheleznykh Dorog 1 (1954) 22–34 (in Russian).
[13] V.P. Cherenin, Drawing up an optimal plan of gathering one-group trains using computers, Vestnik VNII Zheleznodorozhnogo Transporta 1 (1961) 21–24.
[14] V.P. Cherenin, Solving Some Combinatorial Problems of Optimal Planning by the Method of Successive Calculations, in: Proceedings of the Conference on Experiences and Perspectives of the Application of Mathematical Methods and Electronic Computers in Planning, Mimeograph, Novosibirsk, Russia, 1962 (in Russian).
[15] J. Edmonds, Matroids, submodular functions, and certain polyhedra, in: R.K. Guy, H. Hanani, N. Sauer, J. Schönheim (Eds.), Combinatorial Structures and Their Applications, Gordon and Beach, New York, 1970, pp. 69–87.
[16] M.A. Efroymson, T.L. Ray, A Branch-Bound algorithm for plant location, Operations Research 14 (1966) 361–368.
[17] U. Feige, J. Vondrák, The Allocation Problem with Submodular Utility Functions (preliminary version), 2006.
[18] A. Frank, Matroids and submodular functions, in: M. Dell'Amico, F. Maffiolli, S. Martello (Eds.), Annotated Bibliographies in Combinatorial Optimization, Wiley, New York, 1997, pp. 65–80.
[19] A.M. Frieze, A cost function property for plant location problems, Mathematical Programming 7 (1974) 245–248.
[20] M.R. Garey, D.S. Johnson, Computers and Intractability, A Guide to the Theory of NP-Completeness, Freeman, San Francisco, 1979.
[21] A. Genkin, C.A. Kulikowski, Ilya Muchnik, Set covering submodular maximization: An optimal algorithm for data mining in bioinformatics and medical informatics, Journal of Intelligent & Fuzzy Systems: Applications in Engineering and Technology 12 (1) (2002) 5–17.
[22] A.V. Genkin, I.B. Muchnik, Optimum algorithm for maximization of submodular functions, Automation and Remote Control USSR 51 (1990) 1121–1128.
[23] B.I. Goldengorin, A correcting algorithm for solving some discrete optimization problems, Soviet Mathematical Doklady 27 (1983) 620–623.
[24] B. Goldengorin, Requirements of Standards: Optimization Models and Algorithms, Russian Operations Research, Hoogezand, The Netherlands, 1995.
[25] B. Goldengorin, G. Sierksma, G.A. Tijssen, M. Tso, The data-correcting algorithm for minimization of supermodular functions, Management Science 45 (1999) 1539–1551.
[26] B. Goldengorin, G.A Tijssen, M. Tso, The Maximization of Submodular Functions: Old and New Proofs for the Correctness of the Dichotomy Algorithm, SOM Research Report 99A17, University of Groningen, 1999, p. 19.
[27] B. Goldengorin, D. Ghosh, The multilevel search algorithm for the maximization of submodular functions applied to the quadratic cost partition problem, Journal of Global Optimization 32 (1) (2005) 65–82.
[28] R.P. Grimaldi, Discrete and Combinatorial Mathematics: An Applied Introduction, third ed., Addison-Wesley Publishing Company, New York, 1994.
[29] P. Hammer, S. Rudeanu, Boolean Methods in Operations Research and Related Areas, Springer, 1968.
[30] V.R. Khachaturov, Some Problems of the Successive Calculation Method and its Applications to Location Problems, Ph.D. Thesis, Central Economics & Mathematics Institute, Russian Academy of Sciences, Moscow, 1968 (in Russian).
[31] V.R. Khachaturov, Mathematical Methods of Regional Programming, Nauka, Moscow, 1989 (in Russian).
[32] C.-W. Ko, J. Lee, M. Queyranne, An exact algorithm for maximum entropy sampling, Operations Research 43 (1995) 684–691.
[33] J. Lee, Constrained maximum-entropy sampling, Operations Research 46 (1998) 655–664.
[34] H. Lee, G.L. Nemhauser, Y. Wang, Maximizing a submodular function by integer programming: Polyhedral results for the quadratic case, European Journal of Operational Research 94 (1996) 154–166.
[35] L. Lovasz, Submodular functions and convexity, in: A. Bachem, M. Grötschel, B. Korte (Eds.), Mathematical Programming: The State of the Art, Springer-Verlag, Berlin, Germany, 1983, pp. 235–257.
[36] M. Minoux, Accelerated greedy algorithms for maximizing submodular set functions, in: J. Stoer (Ed.), Actes Congress IFIP, Springer-Verlag, Berlin, Germany, 1977, pp. 234–243.
[37] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions-I, Mathematical Programming 14 (1978) 265–294.
[38] G.L. Nemhauser, L.A. Wolsey, Best algorithms for approximating the maximum of a submodular set function, Mathematical Operations Research 3 (3) (1978) 177–188.
[39] G.L. Nemhauser, L.A. Wolsey, Maximization submodular set functions: Formulations and analysis of algorithms, in: P. Hansen (Ed.), Studies on Graphs and Discrete Programming, North-Holland Publishing Company, Amsterdam, The Netherlands, 1981, pp. 279–301.
[40] A. Petrov, V. Cherenin, An improvement of train gathering plans design methods, Zheleznodorozhnyi Transport 3 (1948) 60–71 (in Russian).
[41] T.G. Robertazzi, S.C. Schwartz, An accelerated sequential algorithm for producing D-optimal designs, SIAM Journal on Scientific and Statistical Computing 10 (1989) 341–358.
[42] A. Schrijver, A combinatorial algorithm minimizing submodular functions in strongly polynomial time, Journal of Combinatorial Theory Series B 80 (2) (2000) 346–355.
[43] J. Vondrák, Optimal Approximation for the Submodular Welfare Problem in the Value Oracle Model, Department of Mathematics, Princeton, USA (unpublished manuscript), 2007.