# Submodular Set Functions and Monotone Systems in Aggregation Problems. I *

I. B. Muchnik and L. V. Shavartser UDC 62-506.1

A relationship is established between two types of set functions – submodular functions and functions determining the extremal properties of monotone systems. It is shown that this relationship may be utilized in applied combinatorial optimization problems, in particular, for identifying the structure of empirical information.

## 1. Introduction

Data aggregation problems are often stated in terms of extremizing some criteria, which measure the quality of the sought result [1,2]. Problems of this kind are reducible to combinatorial optimization. Traditionally, these problems were solved by local optimization schemes, which produced only approximate solutions [1,2]. In recent years, it became clear that some aggregation problems may be treated as extremizing a submodular function (SF) [3] or a search for a core of a monotone system (MS) [4,7]. These two systems of set functions have many properties, which make it possible to find their global extrema. The work on submodular functions and monotone systems has pursued different paths different disciplines. Yet there is a way to link these disciplines, since the derivatives of SF, defined in a certain way, are the monotone functions entering the definition of MS. The authors of [8], who proposed this definition of derivatives, also showed that any set function with a monotone derivative is submodular. However, not being familiar with [9], they could not deal with MS as an independent object with its own specific extremal properties, nor could they of course consider the dependence between the properties of these two systems of set functions.

The aim of the present paper is to investigate this dependence. The main focus is on topics, which are directly related to the development of effective extremum-seeking algorithms for these functions.

Section 2 of Part I analyzes the dependence between submodular set functions and monotone systems. Necessary and sufficient conditions are derived for a monotone system to be a derivative of a submodular function. It is shown that the cores of monotone systems associ-

* Moscow. Translated from Avtomatica i Telemekhanika, No. 5, pp. 135 – 148, May, 1987. Original article submitted June 2, 1986.

ated with the rank functions of matroids have a graphic interpretation; the analysis suggests effective solubility of one maximin problem on the set of vertices of a convex polyhedron generated by SF. In this section, we present a duality theory of MS, which is the basis for the derivation of bounds of the following algorithms. It is also shown that a stronger form of the duality theorem makes it possible to link the theory of monotone systems with an important class of applied scheduling problems [10,11]. Section 2 reviews maximin-seeking methods for SF, both global maxima and maxima under inequality constraints. It is shown that these methods essentially utilize the monotonicity of the derivative of the SF. Known methods are generated and extended.

Problems with equality constrains so far have not been considered in the literature. The solution of such problems is particularly difficult, because a set of fixed-cardinality subsets is not a lattice. However, our approach suggests a modification of the branch-and-bound algorithm for the global optimization problem, capable of solving optimization problems with equality constraints. A description of the maximum-seeking problem for SF on fixed-cardinality subsets concludes Sec. 3.

Part II will deal with minimization of SF, and also with applications of the various problems to empirical data processing. The analysis of these problems essentially utilizes the extremal properties of MS. Proofs of theorems are collected in the Appendices.

## 2. Monotone Systems as Derivatives of Submodular Functions

Let a number $\pi^+(i,H)$ be associated with the subset $H \subseteq W$ $\left(|W| = N\right)$ and the element $i \in H$, such that $\forall\, i,H_1,H_2 :\ i \in H_1 \subseteq H_2 \subseteq W$ we have

$$\pi^+(i,H_1) \geq \pi^+(i,H_2)\,. \tag{1}$$

Define the set function

$$F(H) = \max_{i \in H} \pi^+(i,H),\ H \subseteq W\,. \tag{2}$$

The construction (1), (2) is denoted by $\langle W, \pi^{+}, F\rangle$ and is called ⊕-MS [9]. A different construction $\langle W, \pi^{-}, F\rangle$, defined similarly but with $\geq$ replaced by $\leq$ and $+$ by $+$ by $-$ in (1) and $max$ replaced by $min$ in (2), is called ⊖-MS. In what follows, we deal with ⊕-MS, and the sign ⊕ is omitted. [1]

An effective procedure is available [9], which finds the global minimum of the function $F(H)$ on the set of all subsets of $W$ and simultaneously the maximum core – $min\, F(H)$ with argument of maximum cardinality. [2] In order to describe the procedure, we introduce the notion of interval – a family $[A,B]$, $A \subseteq B \subseteq W$ of subsets $H \subseteq W$ such that $A \subseteq H \subseteq B$.

The core extraction procedure is the following. Starting with interval $[\varnothing, W]$, successively reduce the interval from the right, $[\varnothing, B] \rightarrow [\varnothing, B \setminus \{i_B\}]$, where $\pi(i_B, B) = \max_{i \in B} \pi(i, B)$. In each reduction step, record the omitted element $i_B$ and the value $\pi(i_B, B)$. Continue the procedure until $W$ has been exhausted. The sequence of omitted elements $\langle i_{B_1}, i_{B_2}, \ldots\rangle$, where $B_{k+1} = B_k \setminus \{i_B\}$, is called maximum defining sequence. We denote it by $I$. Then in the sequence $B_I = \langle B_1, B_2, \ldots\rangle$ of the right ends of the intervals, take the first subset $B$ with $\min_{B \in B_I} \pi(i_B, B)$; this is the maximum core of the MS.

Let $\Omega = \{J\}$ be the set of all orders on $W$ ($J$ is an order) and let

$$\Pi(J, H) = \pi(i_H(J), H), \tag{3}$$

where $H \subseteq W$, $i_H(J)$ is the first element of the set $H$ in the order $J$. For instance, if the order is the maximum defining sequence $I$, then the function $\Pi(I, H)$ coincides with the values of the function $F(H)$ on the sets $H$ from the sequence $B_I$. The following theorem holds.

---

[1] It is easily seen that all the properties of ⊕-MS discussed below are symmetrical to the properties of ⊖-MS with $\geq$ replaced by $\leq$, $+$ by $-$, and $max$ by $min$.

[2] The arguments of $min\, F(H)$ are called the cores of MS. It is shown in [9] that the cores of a MS form a semilattice – it is closed under the operation of union. Therefore, in particular, the maximum core is a priori the largest, i.e., includes any other core.

Theorem 1 (duality).[3] *For the subset* $H^*$ *to be a solution of the problem*

$$F(H) \to \min_{H \in W},$$

*it is necessary and sufficient that there exists an order* $J^* \in \Omega$ *such that the function* $\Pi(J,H)$ *on the pair* $(J^*,H^*)$ *satisfies the condition*

$$\Pi(J,H^*) \le \Pi(J^*,H^*) \le \Pi(J^*,H),$$

*i.e., the point* $(J^*,H^*)$ *is a saddle point of the function* $\Pi(J,H)$ *on the set* $\{(J,H)) | J \in \Omega, H \subseteq W\}$.

Corollary of Theorem 1.

$$\max_{J \in \Omega} \min_{H \subseteq W} \Pi(J,H) = \min_{H \subseteq W} \max_{J \in \Omega} \Pi(J,H).$$

Note that

$$\max_{J \in \Omega} \Pi(J,H) = \max_{i \in H} \pi(i,H) = F(H),\ H \subseteq W,$$

where it follows that if there exists an effective procedure to compute the right-hand side of the equality in the Corollary of Theorem 1, then it also determines the value of the left-hand side.

Remark. The duality theorem can be strengthened as follows. Let $\Omega' \subseteq \Omega$. Introduce a restriction of the function $F(H)$ on the set $\Omega'$,

$$F_{\Omega'}(H) = \max_{J \in \Omega'} \Pi(J,H).$$

It is easy to show that the algorithm minimizing the function $F_{\Omega'}(H)$ in this case differs from the core-seeking algorithm described above only in that it uses a different construction of the sequence $\langle i_{B_1}, i_{B_2}, \ldots \rangle$. Specifically, let $\Omega'_H$ be the set of orders induced by the set $\Omega'$ on $H \in W$. Then in each step of the construction of the defining sequence, in addition to the condition $\pi(i_{B_j}, B_j) = \max_{i \in B_j} \pi(i, B_j)$, $j = \overline{1,N}$, we should also have $\langle i_{B_j}, i_{B_{j+1}}, \ldots, i_{B_N} \rangle \in \Omega'_{B_j}$.

[3] This duality property has no relation to the duality considered in [9].

Theorem 1a. For the set $H^*$ to be a solution of the problem

$$F_{\Omega'}(H) = \min_{H \subseteq W},$$

it is necessary and sufficient that there exists an order $J^* \in \Omega'$ such that $\Pi(J^*, H^*)$ is a saddle point of the function $\Pi(J, H)$ on the set $\{(J,H) | J \in \Omega', H \subseteq W\}$.

The proof of the theorem is similar to the proof of Theorem 1, with $\Omega$ replaced by $\Omega'$ in all statements.

If $\Omega'$ is the set of orders consistent with some acyclic digraph, then we obtain the result of [11]. Thus, there is a definite link between the theory of MS and problems of maximum-penalty minimization in single-stage scheduling systems [10].

In addition to the general analogy, this link also applies to particular cases. For instance, a special class of MS includes systems with separable variables ($\pi(i,H) = g(i) + f(H)$, see [5] ), and the maximum core seeking algorithm is the simplest for these functions. Similarly, the scheduling theory of one-stage systems considers penalty functions with separable variables ($\varphi_i(t) = \varphi(t) + a_i$, see [10] ), which are analogs of the functions $\pi(i,H)$ in MS. These functions lead to the simplest optimal scheduling algorithm.

Let us now consider MS as the derivatives of SF.

A submodular set function is the function $P(H)$, $H \subseteq W$, with the property [4]

$$P(A) + P(B) \geq P(A \cup B) + P(A \cap B) \ \ \forall\, A, B \subseteq W. \tag{4}$$

A derivative of the function $P(H)$ on the element ("in the direction") $i \in H$ is the function [5]

$$\pi(i,H) = P(H) - P(H \setminus \{i\}). \tag{5}$$

For $H_1 \subset H_2 \subseteq W$, and $H_2 \setminus H_1 = \{i_1, \ldots, i_k\}$, we have the obvious expansion

$$P(H_2) = P(H_1) + \pi(i_1, H_1 \cup \{i_1\}) + \pi(i_2, H_1 \cup \{i_2, i_1\}) + \ldots + \pi(i_k, H_1 \cup \{i_k, i_{k-1}, \ldots, i_1\}). \tag{6}$$

---

[4] The function $(-P(H))$, where $P(H)$ satisfies (4), is called supermodular. A function, which is both sum- and supermodular, is called modular.

[5] To simplify the discussion, both the function $\pi(i,H)$ and the associated monotone system $\langle W, \pi, F \rangle$ are treated in what follows as the derivatives of some submodular function.

**Theorem** [8]. The following statements are equivalent:

1) $P(H)$ is submodular function.
2) $\pi(i,H_1) \geq \pi(i,H_2)$ for all $i$, $H_1$, $H_2$, such that $i \in H_1 \subseteq H_2 \subseteq W$ [for supermodular function, respectively, $\pi(i,H_1) \leq \pi(i,H_2)$].

Thus, the derivative of a SF is monotone decreasing in the set argument and, conversely, a set function whose derivative is monotone increasing in the set argument is supermodular. However, not every MS is a derivative of some SF.

Theorem 2. A MS is a derivative of a SF if and only if for all $i, j$: $i, j \in H \subseteq W$, we have

$$\pi(i,H) - \pi(i,H \setminus \{j\}) = \pi(j,H) - \pi(j,H \setminus \{i\}). \tag{7}$$

Sufficiency of Theorem 2 follows from the fact that, when (7) holds, the sought SF may be defined by (6) – the result is well-defined by (7), since the value of SF is independent of the integration path of its derivative.

Corollary of Theorem 2. Let $\langle W, \pi, F \rangle$ be the derivative of an antitone submodular function $P(H)$ $(H' \subseteq H \Rightarrow P(H') \geq P(H))$. Then if the element $j$ is removed from the set $H$, the weights $\pi(i,H)$ of all the remaining elements increase at most by absolute value of the weight of the element $j$ on the set $H$:

$$\pi(i,H \setminus \{j\}) - \pi(i,H) \leq |\pi(j,H)|.$$

Consider another special class of MS, which provides additional opportunities for generating SF.

Let $\langle W, \pi^M, F \rangle$ is a MS with a monotone derivative, then the function

$$P(H) = \sum_{i \in H} \pi^M(i,H), \; H \subseteq W \text{ is submodular.}$$

In Part II we will identify a system of SF composed in this way from MS with a constant derivative on $H$ ( $\pi(i,H) - \pi(i,H \setminus \{j\}) = f(i,j)$ is independent of $H$ ). We will show that this system plays an important role in aggregation problems.

In conclusion of this section, let us consider two examples, which demonstrate the use of these relationships between MX and SF.

Example 1. We know [12] that the rank function $r(H)$ of the matroid [6] $(W,\mathsf{F})$, defined on the set $W$ is submodular $(H \subseteq W)$. Let $\langle W,\pi^{\mathrm{r}},F^{\mathrm{r}}\rangle$ be the MS, which is the derivative of the function $r(H)$.

Theorem 4. For the matroid $(W,\mathsf{F})$ to have a cycle it is necessary and sufficient that $F^{\mathrm{r}}$ vanish on the core $H^*$ of the MS $\langle W,\pi^{\mathrm{r}},F^{\mathrm{r}}\rangle$,

$$F^{\mathrm{r}}(H^*)=0. \tag{8}$$

The set $H' \subseteq W$ is a cycle of the matroid $(W,\mathsf{F})$ if and only if it is minimal (by inclusion) core of the MS $\langle W,\pi^{\mathrm{r}},F^{\mathrm{r}}\rangle$.

Theorem 5. If the matroid $(W,\mathsf{F})$ contains at least one cycle, the largest core of the system $\langle W,\pi^{\mathrm{r}},F^{\mathrm{r}}\rangle$ is the union of all the minimal cores of the system $\langle W,\pi^{\mathrm{r}},F^{\mathrm{r}}\rangle$, i.e., all the cycles of the matroid $(W,\mathsf{F})$.

We can interpret these theorems in two ways: on one hand, they provide a new characterization of the set of cycles of a matroid in terms of the properties of MS; on the other hand, they identify a new object in the analysis of MS – cores minimal by inclusion. For instance, in the MS with $\pi(i,H)=\sum_{j\in H}a_{ij}$ studied in [4], practical considerations indeed suggest that we look for minimal cores. It is easy to show that if all $a_{ij}>0$, then this core is unique, and a simple algorithm finds it.

Example 2. With each SF $P(H)$, $H \subseteq W$, we can associate a convex polyhedron [13],

$$M(P)=\left\{ x \,\middle|\, x\in \Re^W, \forall\, H \subseteq W : x(H)=\sum_{i\in H}x(i)\le P(H) \right\},$$

where $\Re^W$ is the set of vectors $x=\{x(i), i\in W\}$, $x(i)\in\Re$. From $M(P)$ we can isolate a convex set of $B(P)=\left\{x \,\middle|\, x\in M(P), x(W)=P(W)\right\}$, which is bounded [13].

Let $V(P)$ be the vertex set of $B(P)$, and consider the following problem: find a vertex $x^*\in V(P)$ and a coordinate $i^*\in W$, such that

$$x^*(i^*)=\max_{x\in V(P)}\min_{i\in W}x(i). \tag{9}$$

[6] All the necessary definitions are given in Appendix 2. Note that by the definition of matroid, the empty set is always independent, and therefore a cycle, if it exists, is always a nonempty set.

Theorem 6. The problem (9) reduces to finding the maximum-cardinality core of the MS $\langle W, \pi^{P}, F^{P} \rangle$ – the derivative of the generating function $P(H)$.

By Theorem 6, the problem (9) is solved in time $O(N^2/2)$ $(|W| = N)$.

## 3. Monotonicity of the Derivative and Maximum-Seeking Algorithms for SF

Consider the following problems:

$$P(H) \to \max_{H \subseteq W}, \tag{10}$$

$$P(H) \to \max_{H \subseteq W, |H| \le k \le N}, \tag{11}$$

$$P(H) \to \max_{H \subset W, |H| \ge N-k,\, k \le N}, \tag{12}$$

$$P(H) \to \max_{H \subseteq W, |H| = k < N}. \tag{13}$$

The problem (10) and its algorithm were first proposed in [14]. Then this method was improved in [15-19]. A modified form of these algorithms is also applicable for the problem (11) and (12). All these algorithms are variants of the branch-and-bound method. A number of suboptimal algorithms were proposed in [9]: these algorithms can be used to estimate the quality of the approximate solutions, and these estimates lead to a stronger form of the branch-and-bound algorithms. [7]

The algorithms are based on the properties of SF, which yield sufficient conditions for reduction of the localizing solution of the set-theoretic interval or rejection of unpromising intervals. Different authors develop their own schemes of the algorithm, utilizing only part of the SF properties. Yet all these properties stem from the same basic fact, namely that the derivatives of SF form MS. Therefore, by treating these properties in a unified framework, we will not only incorporate in one algorithm all the various sufficient conditions for interval reduction or rejection, but also find ways for development of algorithms for new problems. In

[7] Note that the "greedy" algorithm is in a certain sense the best among the suboptimal algorithms proposed in [8].

particular, proceeding along these lines, we have constructed an algorithm for the problem (13), which cannot be solved by available algorithms. Let us now present this unified framework [8] and use it to examine the algorithms for problems (10)-(13).

Definition. A point of local maximum of the function $P(H)$ is the subset $H^\circ \subseteq W$, such that

$$\pi(i,H^\circ) \geq 0 \quad \forall i \in H^\circ,$$

$$\pi(i,H^\circ \cup \{i\}) \leq 0, \ \forall i \in W \setminus H^\circ,$$

where $\langle W,\pi,F \rangle$ is the MS of the derivatives of the function $P(H)$.

Cherenin's Theorem. [9] Let $J = \langle i_1,...,i_N \rangle$ be an order on the set $W$. If the set $H^{**} = \{ i_k, i_{k+1},...,i_N \}$ is a point of local maximum of $P(H)$, then

$$\pi(i_j, \{i_j, i_{j+1},...,i_N\}) \leq 0, \ j \leq k,$$

$$\pi(i_j, \{i_j, i_{j+1},...,i_N\}) > 0, \ j > k.$$

Consider the sequence of nested sets $\langle W, W \setminus \{i_1\}, W \setminus \{i_1 \cup i_2\},..., H = \{i_k,...,i_N\} \rangle$ generated by successively dropping the elements of $W$, taken one by one in the order $J$. This sequence is called the chain induced by the order $J$. Then, Cherenin's Theorem implies that $H^{**}$ is the point of global maximum of $P(H)$ on the entire chain induced by the order $J$.

The proof of the theorem easily follows from monotonicity of the derivatives of the function $P(H)$.

Cherenin's Theorem indicates that a SF does not have too many maxima, and the global maximum can be chosen by direct enumeration. It provides two branch-rejection rules reducing this enumeration. Let the local maximum point $H^\circ$ of function $P(H)$ belong to the interval $[A,B]$.

[8] In Part II, we will show that this unified framework leads to a symmetric scheme for the solution of minimization problems of submodular functions.

[9] This is a restatement of the theorem from [17], which in its original form does not utilize the notion of MS; all the known facts, including the maximum-seeking algorithms, and in particular, the enumeration-reducing rules described below, are presented in a similarly restated form.

First Rejection Rule. If $\pi(i,B)>0$ for some $i\in B\setminus A$, then $i\notin H^\circ$ ($H^\circ\in[A\cup\{i\},B]$ is the left-reduction of the interval).

Second Rejection Rule. If $\pi(i,A\cup\{i\})<0$ for some $i\in B\setminus A$, then $i\in H^\circ$ ($H^\circ\in[A\cup\{i\},B]$ is the right-reduction of the interval).

These rules are clearly equivalent to (yet another) restated form of Cherenin's Theorem.

Cherenin's algorithm based on these two rules is a branch-and-bound algorithm with the following scheme.

1. Construct the solution tree, starting with the vertex $[A,B]$, $A=\varnothing$, $B=W$ and $[A\cup\{i\},B]\leftarrow[A,B]\rightarrow[A,B\setminus\{i\}]$, where $i\in B\setminus A$; to each vertex assign the best value so far $r=max\left(P(A),P(B)\right)$. The leaves of this tree are the intervals $[A,B]$ m where $A=B$.
2. Traverse the tree: each step involves successive reduction of the interval $[A,B]$ by the first or the second rule, until no further reduction is possible. If the conditions of the rules are not satisfied for any element $i\in B\setminus A$, descend along the leftmost of those branches, which have not been traversed.
3. The leaf with the maximum best value $r$ so far is the solution.

The algorithm is effective because when it reaches the vertex $[A,B]$ where either $A$ or $B$ is strict [10] local extremum, the next step (by the rejection rule) is a leaf. [11]

Third Rejection Rule. Let $r^{max}$ be the value of $r$ attained by the given step of the algorithm. let

$$P_1^{A,B}=P(A)+\sum_{i\in B\setminus A}\pi(i,A\cup\{i\}),\ P_2^{A,B}=P(B)-\sum_{j\in B\setminus A}\pi(j,B).$$

[10] The case with nonstrict inequalities was considered in [18].

[11] It is noted in [8] that the search even for local extremum may have exponential complexity. Yet in practice, the proposed algorithm converges in $O(N^3)$ steps [20].

Suppose that the interval $[A,B]$ cannot be reduced further by the first two rules. Then, if $P_1^{A,B} < r^{max}$ or $P_2^{A,B} < r^{max}$, the solution of (10) is not contained in the interval $[A,B]$.

Proof of the Third Rule. We will show that $\forall C,\ C \in [A,B]$ we have

$$P_1^{A,B} \geq P(C), \tag{14}$$

$$P_2^{A,B} \geq P(C). \tag{15}$$

Use the expansion (6)

$$P(C) = P(A) + \pi(i_1, A \cup \{i_1\}) + \pi(i_2, A \cup \{i_1 \cup i_2\}) + ,..., + \pi(i_k, A \cup \{i_1,...,i_k\}),$$

$$P(B) = P(C) + \pi(j_1, C \cup \{j_1\}) + \pi(j_2, A \cup \{j_1 \cup j_2\}) + ,..., + \pi(j_m, C \cup \{j_1,..., j_m\}),$$

where $\langle i_1,...,i_k \rangle$ is some ordering of the elements from $C \setminus A$; $\langle j_1,..., j_m \rangle$ is an ordering of the elements from $B \setminus C$. By (7),

$$P(C) \leq P(A) + \sum_{i \in C \setminus A} \pi(i, A \cup \{i\}), \tag{16}$$

$$P(C) \leq P(B) + \sum_{j \in B \setminus C} \pi(j, B). \tag{17}$$

If the first rule cannot be applied to $[A,B]$, then $\pi(j,B) \leq 0\ \ \forall j \in B$, if the second rule is inapplicable, then $\pi(i, A \cup \{i\}) \geq 0\ \ \forall i \in B \setminus A$. From (16) and (17) we thus obtain (14) and (15).

By the third rule, the algorithm is augmented, first, by calculation of the bounds $P_1^{A,B}$ and $P_2^{A,B}$ at the vertex $[A,B]$ and, second, by rejection of unpromising intervals $[A,B]$ (branches of the solution tree). [12]

The original [17] and the modified [18,19] algorithms may be applied to solve the problem (11). In this case, the leaves are the intervals $[A,B]$ with either $|A| = k$ or $|B| = k$. The problem (12) is solved by passing to a different [13] SF [8], $\overline{P}(H) = P(W \setminus H)$.

---

[12] Other authors also pursue the logic applying special rules to an “unreduced” (in Cherenin’s sense) interval. Thus, current local updating of the data matrix is proposed in this case for local problems in [21].

[13] The problem (12) for $P(H)$ is transformed to the problem (11) $\overline{P}(H)$.

Choice of the starting best value for the solution of the problem (11) with $k = N$ and for the solution of the problem (10) may be performed by the "greedy" procedure [14] [8]: starting with $[\varnothing, W]$, successively reduce the interval "from the left" $[A, W] \to [A \cup \{i_A\}, W]$, with $i_A$ chosen from $\pi(i_A, A \cup \{i_A\}) = \max_{i \in W \setminus A} \pi(i, A \cup \{i\})$, if $\pi(i_A, A \cup \{i_A\}) \geq 0$; otherwise (if $\pi(i_A, A \cup \{i_A\}) < 0$, the algorithm stops (it also stops when $|A \cup \{i_A\}| = k$ ).

Proof of choice of the starting value. The bound on the approximate solution obtained by this procedure (its "closeness" to the exact solution) has the form [8]

$$\frac{P^{max} - P^G}{P^{max} - P(\varnothing) + k \cdot \theta} \geq \left( \frac{k-1}{k} \right)^k, \tag{18}$$

where $P^{max}$ is the exact solution of the problem (11), $P^G$ is its "greedy" solution, $\theta$ is a constant such that $\pi(i, H) \geq -\theta \quad \forall H \subseteq W$, $\forall i \in H$.

This bounding method is based on the relationship

$$\forall S, T \subseteq W \quad P(T) \leq P(S) + \sum_{i \in T \setminus S} \pi(i, S \cup \{i\}) - \sum_{i \in S \setminus T} \pi(i, S \cup T),$$

which, in turn, directly follows from the expansion (6) using the monotonicity of the derivative (7). [15]

These algorithms will not solve the problem (13). [16] However, they may be modified in order to solve this problem. Specifically, using (16), (17), we obtain the following bounds $\forall H$, $H \in [A, B]$, $|H| = k$:

$$P(H) \leq P(A) + (k - |A|) \cdot \max_{i \in B \setminus A} \pi(i, A \cup \{i\}), \tag{19}$$

$$P(H) \leq P(B) - (|B| - k) \cdot \min_{j \in B \setminus A} \pi(j, B). \tag{20}$$

As a result, we obtain the following branch-and-bound algorithm.

---

[14] The proposed "greedy" procedure, unlike the standard version (see Appendix 2), executes, not on an arbitrary system of independent subsets, but on the set of all subsets of $W$. It is designed to find, not an exact extremum of a modular function, but an approximate extremum of a submodular function. The procedure correspondingly necessitates checking the condition $(A \cup \{i\}) \in \mathrm{F}$ and the stopping rules in each step.

[15] It is also shown in [8] that this inequality may be chosen as one of the equivalent definitions of SF.

[16] If $P(H)$ is a submodular function, which is monotone nondecreasing with the extension of the set argument, the "greedy" procedure described above also gives an approximate solution of the problem (13) [with the bound (18)].

1. Construct the solution tree as in Cherenin's algorithm (without evaluating $r$ in each step); the starting best value may be chosen as the value of $P(H)$ on any flexible $H \in [A,B]$; the leaves are $(|A| = k) \vee (|B| = k)$; update the best value so far when a leaf is reached.

2. Traverse the tree along the leftmost of the remaining branches, reaching "promising" vertices for which the bounds (19) and (20) are greater (not less) than the maximum of the best values so far. [17]

   With each interval $[A,B]$, associate the SF $P_{AB}(H)$ defined on the set of subsets of $B \setminus A$, $P_{AB}(H) = P(H \cup A)$, $H \subseteq B \setminus A$. Then (18) for $P_{AB}(H)$ takes the form

$$P_{AB}^{max} \le \frac{P_{AB}^{G} - P_{AB}(\varnothing)\cdot\alpha^{k'} + k'\cdot\theta\cdot\alpha^{k'}}{1-\alpha^{k'}} = \tilde{P}_{AB}\,, \tag{21}$$

Clearly, (21) suggests further rejection of unpromising intervals for the case [18] $P_{AB} < r^{max}$. Note, however, that this bound requires $O\left(|B \setminus A|^2 / 2\right)$ computations of $\pi(i,H)$, whereas the previous bounds required only $O(|B \setminus A|)$ such computations each.

## APPENDIX 1

Proof of Theorem 1. Sufficiency. We will show that if $(J^*, H^*)$ is a saddle point of the function $\Pi(J,H)$, then

$$F(H^*) \le F(H),\; H \subseteq W\,. \tag{A.1}$$

Indeed,

$$\Pi(J^*,H^*) = \min_{H \subseteq W}\; \max_{J \in \Omega}\; \Pi(J,H) = \min_{H \subseteq W} F(H)\,.$$

On the other hand, $\Pi(J^*,H^*) = \max_{J \in \Omega} \Pi(J,H^*) = \max_{i \in H^*} \pi(i,H^*) = F(H^*)$, which proves (A.1).

[17] In Part II, the bound (20) will be strengthened with the aid of the extremal properties of the derivatives of SF.
[18] The feasibility of this bound has been noted in [8].

Necessity. We will show that if $H^*$ is a solution of the problem $F(H) \to \min_{H \subseteq W}$, then there exists a sequence $J^* \in I$ such that

$$\Pi(J,H^*) \le \Pi(J^*,H^*) \le \Pi(J^*,H). \tag{A.2}$$

The proof of this relies on the following lemma.

Let $J = \langle j_1, j_2, ..., j_N \rangle \in \Omega$. Denote $\mathrm{H}_J = \left\langle H \mid H = \{j_k, j_{k+1}, ..., j_N\}, k = \overline{1,N} \right\rangle$.

LEMMA.

$$\min_{H \in \mathrm{H}_J} \Pi(J,H) = \min_{H \subseteq W} \Pi(J,H).$$

Proof of Lemma 1. Let $H \notin \mathrm{H}_J$. Then there exists a set $H' \in \mathrm{H}_J$, such that its first element in the sequence $J$ coincides with the first element of the set $H$ in the sequence $J$; a priori, $H' \supseteq H$. By monotonicity of the function $\pi(i,H)$ and from the definition of $\Pi(J,H)$, we have $\Pi(J,H') \le \Pi(J,H)$, which proves Lemma 1.

Let us now prove (A.2). Take $J^*$ as the defining sequence. Then by Mullat's theorem [9] and by the definition of $\Pi(J,H)$ and $F(H)$, we have

$$\Pi(J^*,H^*) = F(H^*) = \max_{J \in \Omega} \Pi(J,H^*), \tag{A.3}$$

$$\Pi(J^*,H^*) = \min_{H \in \mathrm{H}_{J^*}} \Pi(J^*,H). \tag{A.4}$$

The left-hand side of (A.2) clearly follows from (A.3). By Lemma 1, from (A.4) we obtain the right-hand side of (A.2). ■

Proof of Theorem 2. Necessity. By the theorem [8], the derivative of the SF $P(H)$ is the MS $\langle W,\pi,F \rangle$. From the definition of the derivative (5), it follows that $\forall\, i,j \in H$, $H \subseteq W$

$$\begin{aligned}
&\pi(i,H) + \pi(j,H \setminus \{i\}) = \\
&P(H) - P(H \setminus \{i\}) + P(H \setminus \{i\}) - P(H \setminus \{i,j\}) = \\
&P(H) - P(H \setminus \{i,j\}),
\end{aligned}$$

$$\begin{aligned}
&\pi(j,H) + \pi(i,H \setminus \{j\}) = \\
&P(H) - P(H \setminus \{j\}) + P(H \setminus \{j\}) - P(H \setminus \{i,j\}) = \\
&P(H) - P(H \setminus \{i,j\}),
\end{aligned}$$

whence we obtain necessity of the condition (7).

Sufficiency. Let $H' \subseteq H \subseteq W$ . Let $J_{H\backslash H'} = \langle i_1, i_2, ..., i_d \rangle$ be an order on $H \setminus H'$ . Denote

$$\varphi( J_{H\backslash H'} ) = \pi( i_1, \{i_1\} \cup H' ) + \pi( i_2, \{i_2, i_1\} \cup H' ) + ... + \pi( i_d, \{i_d, i_{d-1}, ..., i_1\} \cup H' ).$$

We will show that for any two orders $J^1_{H\backslash H'} = \langle i^1_1, i^1_2, ..., i^1_d \rangle$ and $J^2_{H\backslash H'} = \langle i^2_1, i^2_2, ..., i^2_d \rangle$, we have

$$\varphi( J^1_{H\backslash H'} ) = \varphi( J^2_{H\backslash H'} ). \tag{A.5}$$

This will be proved by mathematical induction. For $d = 2$, (A.5) is (7). Let (A.5) hold for all $d = 2, ..., k$ . We will show that then it also holds for $d = k + 1$.

Consider the elements $j^1 = i^1_{k+1}$ and $j^2 = i^2_{k+1}$. If $j^1 = j^2 = j$ , then

$$\varphi( J^1_{H\backslash H'} ) = \varphi( i^1_1, i^1_2, ..., i^1_k ) + \pi( j, H \setminus \{ j\}),$$
$$\varphi( J^2_{H\backslash H'} ) = \varphi( i^2_1, i^2_2, ..., i^2_k ) + \pi( j, H \setminus \{ j\}),$$

and the first terms in the right-hand sides of these equalities are also equal by the inductive hypothesis. Therefore, (A.5) holds. Now let $j^1 \neq j^2$ . Let

$$J^3_{H\backslash H'} = \langle i_1, i_2, ..., i_{k-1}, j^1, j^2 \rangle,$$
$$J^4_{H\backslash H'} = \langle i_1, i_2, ..., i_{k-1}, j^2, j^1 \rangle,$$

where $\langle i_1, i_2, ..., i_{k-1} \rangle$ is an ordering of the elements from $H \setminus \left(H' \cup \left\{ j^1, j^2 \right\}\right)$. Then, as is easily seen,

$$\varphi( J^4_{H\backslash H'} ) = \varphi( \langle i_1, i_2, ..., i_{k-1}, j^2 \rangle ) + \pi( j^1, H \setminus \left\{ j^1 \right\})$$

should be equal to

$$\varphi( J^1_{H\backslash H'} ) = \varphi( \langle i^1_1, i^1_2, ..., i^1_{k-1}, j^2 \rangle ) + \pi( j^1, H \setminus \left\{ j^1 \right\}),$$

since $\langle i_1, i_2, ..., i_{k-1}, j^2 \rangle = \langle i^1_1, i^1_2, ..., i^1_k \rangle = H \setminus \left\{ j^1 \right\}$, and therefore by the inductive hypothesis $\varphi( \langle i_1, i_2, ..., i_{k-1}, j^2 \rangle ) = \varphi( \langle i^1_1, i^1_2, ..., i^1_k \rangle )$. Similarly, $\varphi( J^3_{H\backslash H'} )$ should be equal to $\varphi( J^2_{H\backslash H'} )$.

Finally

$$\varphi( J^3_{H\backslash H'} ) = \varphi( \langle i_1, i_2, ..., i_{k-1} \rangle ) + \pi( j^1, H \setminus \left\{ j^1, j^2 \right\}) + \pi( j^2, H \setminus \left\{ j^2 \right\})$$

and

$$\varphi( J^4_{H\backslash H'} ) = \varphi( \langle i_1, i_2, ..., i_{k-1} \rangle ) + \pi( j^2, H \setminus \left\{ j^1, j^2 \right\}) + \pi( j^1, H \setminus \left\{ j^2 \right\})$$

should be equal by assumption. Therefore, (A.5) holds.

We have thus proved (A.5). This means that the MS $\langle W,\pi,F\rangle$ uniquely determines the function $P(H)$ in accordance with the expansion (6). It is easy to see that the system $\langle W,\pi,F\rangle$ is the derivative of this function, which by the theorem of [8] proves its submodularity. ■

Proof of Theorem 3. Consider

$$\pi(i,H)=P(H)-P(H\setminus\{i\})=\pi^M(i,H)+\sum_{j\in H\setminus\{i\}}\left(\pi^{\mathrm{M}}(j,H)-\pi^M(j,H\setminus\{i\})\right). \quad \text{(A.6)}$$

The first term is monotone in $H$ since $\langle W,\pi,F\rangle$ is a MS, the second term is monotone by monotonicity of the derivative $\langle W,\pi,F\rangle$. ■

## APPENDIX 2

The necessarily definitions [8,12]. Let $W$ be a finite set, $\mathsf{F}$ is a nonempty set of subsets $F$ of the set $W$. The system $(W,\mathsf{F})$ is matroid if a) $(F_1\in\mathsf{F})\wedge(F_2\subset F_1)\Rightarrow F_2\in\mathsf{F}$, b) for any $H\subseteq W$, all the maximal (by inclusion) elements of set $\mathsf{F}(H)=\{F \mid F\in\mathsf{F}, F\subseteq H\}$ have the same cardinality.

The rank function $r(H)$, $H\subseteq W$ of a matroid is the cardinality of the maximal (by inclusion) elements of $\mathsf{F}(H)$. The rank function is submodular [12]. The elements of the system $\mathsf{F}$ are called independent subsets. The subsets of the set $W$ not included in $\mathsf{F}$ are called dependent. A minimal (by inclusion) dependent subset is called a cycle of the matroid.

If to each element $i$ of the set $W$ is assigned a weight $\omega(i)$ and a modular function $P(H)=\sum_{i\in H}\omega(i)$, is defined, then the problem of finding an independent subset which maximizes the function $P(H)$ on $\mathsf{F}$ is solved by the "greedy" algorithm [12]: starting with $[\varnothing,W]$, successively reduce the interval from the left $[A,W]\to[A\cup\{i_A\},W]$, where $i_A$ is chosen such that $i_A\in W\setminus A$, $(A\cup\{i_A\})\in\mathsf{F}$.

The solution is clearly $A^*$, such that $\forall\, i\in W\setminus A^*$, $(A\cup\{i\})\in\mathsf{F}$.

Proof of Theorem 4. Note that the function $\pi^{\mathrm{r}}(i,H)$ may only take two values, $0$ and $1$, i.e.,

$$\pi^{\mathrm{r}}(i,H)=\begin{cases}1, & \text{if } r(H)>r(H\setminus\{i\}),\\ 0, & \text{if } r(H)=r(H\setminus\{i\}).\end{cases} \tag{A.7}$$

Also note that if $H$ is an independent subset, then

$$F^{\mathrm{r}}(H)=\max_{i\in H}\pi^{\mathrm{r}}(i,H)=1.$$

Necessity. Let $H'$ be a cycle. Then for $\forall\, i\in H'$ the subset $H'\setminus\{i\}$ is independent, but $H'$ is dependent. Hence $\forall\, i\in H'$ we have

$$r(H')=r(H'\setminus\{i\}).$$

Thus, $r(H')-r(H'\setminus\{i\})=\pi(i,H')=0$, i.e., $F(H')=0$ and $H'$ is a core. We will show that it is a minimal core. Since $H'$ is a cycle, i.e., a minimal dependent set, for $\forall\, H\subset H'$, we have that $H$ is dependent and therefore not a core. This implies that $H'$ is a minimal core. This completes the proof of necessity.

Sufficiency. Now let $H^*$ be a minimal (by inclusion) core of the system $\langle W,\pi^{\mathrm{r}},F^{\mathrm{r}}\rangle$. By condition (8) of Theorem 4, $F(H^*)=0$, which implies that $H^*$ is a dependent set. Assume that it contains another dependent set as a proper subset. Then it contains a cycle, i.e., it is not a minimal (by inclusion) core. The contradiction proves sufficiency. ■

Proof of Theorem 5. We first prove the following lemma.

LEMMA 2. Let $i\in W$ be such that it is not included in any cycle of the matroid $(W,\mathrm{F})$. Then $\pi^{\mathrm{r}}(i,H)=1$ for all $H$ such that $i\in H$ $(H\subseteq W)$.

Proof of Lemma 2. Let $H'$ be a maximal (by inclusion) independent subset of the set $H$ and let $i\notin H'$, $i\in H$. Consider the set $H'\cup\{i\}$. It is dependent (by maximality of $H'$ in $H$) and therefore contains a cycle. But this contradicts the independence of $H'$ and the fact that $i$ is not included in any cycle. The contradiction shows that $i$ is included in all the maximal independent subsets of the set $H$, which proves Lemma 2.

Let us now prove Theorem 5. By the proof of [9], the set of all cores is closed under union. Therefore, the union of all cycles is a core. We will show that it is a maximum core. Consider any set $H$, which contains the union of all cycles as a proper subset. This subset contains the element $i$, which is not included in any cycle, and therefore by Lemma 2, $F^{\mathrm{r}}(H)=1$, i.e., by the existence of a cycle (8), the subset $H$ is not a core of the MS $\langle W, \pi^{\mathrm{r}}, F^{\mathrm{r}} \rangle$. ■

<u>Corollary of Theorem 5</u>. The algorithm seeking the maximum core of the MS $\langle W, \pi^{\mathrm{r}}, F^{\mathrm{r}} \rangle$ operates in the following way: calculate all $\pi^{\mathrm{r}}(i,W)$, $i \in W$, and a) if $\pi^{\mathrm{r}}(i,W)=1$, $i \in W$, then the sought core is $W$; b) otherwise, drop the elements with $\pi^{\mathrm{r}}(i,W)=1$.

<u>Proof of Theorem 6</u>. It is shown in [13] that the point $x$ is a vertex of the polyhedron $B(P)$ if and only if there exists an order $J$ on the set $W$ such that $x(i_J^k)=\pi(i_J^k, H_k^J)$, $k=1,\dots,N$, where $H_1^{\mathrm{r}} \subseteq H_2^{\mathrm{r}} \subseteq \dots \subseteq H_N^{\mathrm{r}} = W$, is the chain of subsets of $W$ induced by the order $J$; $i_J^k$ is the first element of the set $H_k^J$ in the order $J$.

Now, using (3), Theorem 1, and the effective procedure for isolating the largest core, we obtain the proposition of Theorem 6. ■

## LITERATURE CITED

1. E. M. Braverman and I. B. Muchnik, Structural Methods of Empirical Data Processing [in Russian], *Nauka*, Moscow (1983).
2. B. G. Mirkin, Analysis of Qualitative Features and Structures [in Russian], *Statistika*, Moscow (1980).
3. I. A. Rubina, “Application of the method of successive calculations to the problem of automatic classification,” *Avtomat. Telemekh., No. 1*, 126-132 (1986).
4. E. N. Kuznetsov, “Analyzing the structure of the coupling matrix with the aid of the associated monotonic system,” *Avtomat. Telemekh., No. 7*, 128-136 (1980), http://www.datalaundering.com/download/couplem.pdf.
5. I. B. Muchnik, N. F. Chkuaseli, and L. V. Schvartser, “Linguistic analysis of 0-1 Matrices matrices using monotone systems,” Avtomat. Telemekh. No. 4, 132-139 (1986), http://www.datalaundering.com/download/linguist.pdf .

6. L. K. Vykhandu, "On some methods to order objects and variables in data systems," *Tr. Tallinsk. Poytekh. Inst., Tallinn, No. 482*, 43-50 (1980).
7. L. K. Vykhandu, "A fast search on bit matrices," *Tr. Tallinsk. Poytekh. Inst., Tallinn, No. 554*, 49-60 (1983).
8. G. L. Nemhauser, L.A. Wolsey, and M. L. Fisher, "An analysis of approximations for maximaizing submodular set functions, I," *Math. Progr., No. 14*, 265-294 (1978).
9. I. E. Mullat, "Extremal subsystems of monotonic systems, I, II," *Avtomat. Telemekh. No. 5*, 130-139; *No. 8*, 169-179 (1976), http://www.datalaundering.com/download/extrem01.pdf, http://www.datalaundering.com/download/extrem02.pdf.
10. V. S. Tanaev, V. S. Gordon, and Ya. M. Shafranskii, Theory of Schedules. One-Stage Systems [in Russian], *Nauka*, Moscow (1984).
11. S.A Beletskii, "On minimization of the maximum penalty on permutations," *Kibernetika, No. 1*, 70-72. (1986).
12. H. Papadimitriu and K. Steiglitz, Combinatorial Optimization. Algorithms and Complexity [Russian translation], Mir, Moscow (1983).
13. S. Fujishige, "Submodular systems and related topics," *Math. Progr. Study, No. 22*, 113-131 (1984).
14. A. P. Petrov, and V. P. Cherenin, "Improvement of the method for planning of train marshalling," *Zhelezn. Transport, No. 3*, 49-66 (1948).
15. V. P. Cherenin, "Mechanized calculations of train marshalling plans," Tekh. Zhelezny. Dorog., No. 1, 79-96 (1954).
16. V. P. Cherenin, "Optimal scheduling of one-group train marshalling using electronic computers," *Vestnik Vses. Nauch.-Issl. Inst. Zhelzn. Transport., No. 1*, 51-87 (1961).
17. V. P. Cherenin, "Solution of some combinatorial problems of optimal scheduling by the method of successive computations," *Proc. of Conf. on Experience and Prospective Application of Mathematical Methods in Planning [in Russian], Isd. SO AN SSSR, Novosibirsk* (1962), p.111-113.
18. B. R. Khachaturov, Some Questions and applications of the Method of Successive Computations to the solution of Plant Location Problems [in Russian], *Dissertation, TsEMI AN SSSR*, Moscow (1968).
19. B. R. Khachaturov, "Combinatorial approximation method and some applications," *Zh. Vychisl. mat. mat. Fiz., 14*, No. 6, 1464-1487 (1974).
20. A. A. Korbut and Yu.Yu. Finkel'shtein, Discrete Programming [in Russian], *Nauka*, Moscow (1969).
21. B. I. Gol'dengorin, "A correcting algorithm for the solution of the location problems," *Avtomat. Telemekh., No. 5*, 53-62 (1984).