Submodular Set Functions and Monotone Systems in Aggregation Problems. II *

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

A relationship from Part I between submodular functions and functions determining extremal properties of monotone systems is applied to prove that, on the chain of any set-theoretical interval, the submodular function varies more slowly than the linear function of the cardinality of ordered sets. Branch-and-bound algorithms are developed for unconstrained and constrained extremization with optimal tree traversal. Applying the apparatus of combinatorial optimization of submodular functions solves some standard examples of aggregation of empirical data.

1. Introduction

In this article, which is a continuation of [1], we investigate the relationship between submodular functions and monotone systems in the context of minimization of submodular functions. Unlike the maximization problems considered in Part I, minimization of submodular functions requires utilizing the extremal properties of the associated monotone systems. Yet the overall construction for the minimization of submodular functions is entirely similar to the maximization construction.

In Sec. 2, we consider a unified minimum-seeking scheme for the global minimum and for minima under inequality and equality constrains. Sec. 3 discusses some applied topics associated with the general methods from both parts of the paper for aggregation of large arrays of empirical data. Taking one aggregation problem as an example, we show how the minimization of a general (not submodular) function can be reduces to sequential minimization of submodular functions on sets of fixed-cardinality subsets.

The proofs are collected in the Appendix.

* Moscow. Translated from Avtomatica i Telemekhanika, No. 6, pp. 138 – 147, June, 1987. Original article submitted June 4, 1986.

2. Extremal Properties of the Derivative for Minimization of submodular Functions

Consider the following problems:

$$P(H) \to \max_{H \subseteq W}, \tag{1}$$

$$P(H) \to \min_{\substack{H \subseteq W \\ |H| \le k \le N}}, \tag{2}$$

$$P(H) \to \min_{\substack{H \subseteq W \\ |H| \ge N-k,\, k \le N}}, \tag{3}$$

$$P(H) \to \min_{\substack{H \subseteq W \\ |H| = k < N}}, \tag{4}$$

where W is a finite set, k, N natural numbers, $|W| = N$, $P(H)$ is a submodular function on 2^W-set of subsets of W.

The problem (1) is remarkable in that it has a polynomial-time algorithm [2] if $P(H)$ is an integer-valued function with $P(\emptyset) = 0$. For the special case of finding a cut in a graph, this is a flow algorithm [3]. Yet efficient use of the general polynomial algorithms for the problem (1) has been insufficiently studied in [2], and the problem of finding new effective algorithms (possibly heuristic) remains quite topical. This is even more valid for the problems (2)-(4), since to the best of our knowledge no polynomial algorithms have been proposed so far for their solution.

The branch-and-bound algorithms proposed below for minimization of $P(H)$ in problem (1) are logically as simple as the algorithms for the maximization of this function in problem (I.10), described in Part I [1].[1] They are also probably as efficient, since they are entirely analogous to the maximization algorithms.

The algorithms for the solution of problem (I.10) rely on some general properties of the derivatives of submodular functions – the expansion (I.6) and the inequalities (I.7). The algorithms for the solution of problem (1), on the other hand, utilize some special (extremal) properties of these derivatives.

The algorithms for the solution of problem (1) are in effect based on the reversal of three rejection rules used during the branching of the solution tree in [1].

---

[1] The notation (I.10) is a reference to formula (10) in Part I [1]. This notation is used throughout.

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

$$\pi(i, H_\circ) \le 0 \quad \forall i \in H_\circ,$$

$$\pi(i, H_\circ \cup \{i\}) \ge 0 \quad \forall i \in W \setminus H_\circ.$$

Let $[A,B]$ be some set-theoretical interval $(A \subseteq B \subseteq W)$ and $H_\circ \in [A,B]$.

Reversal of first rejection rule. If $i \in B \setminus A$ and $\pi(i,B) > 0$, then $i \notin H_\circ$, i.e., $H_\circ \in [A, B \setminus \{i\}]$ perform right-reduction of the intervals.

Proof of the reversal of the first rejection rule. Let $i \in H_\circ \in [A,B]$. Noting that $\pi(i,B) > 0$ and using (I.7), we again obtain a contradiction with the definition of the set $H_\circ$ as a point of local minimum.

Reversal of second rejection rule. If $i \in B \setminus A$ and $\pi(i, A \cup \{i\}) < 0$ then $i \in H_\circ$ $\left(H_\circ \in [A \cup \{i\}, B]\right.$ perform left-reduction of the interval $\left.\right)$.

Proof of the reversal of the second rejection rule. Let $i \notin H_\circ \in [A,B]$. Noting that $A \cup \{i\} \subseteq H_\circ \cup \{i\}$ and using (I.7), we again obtain a contradiction, which proves that $i \in H_\circ$.

It is easily seen that if H°, $H_\circ \in [A,B]$ $\left(H^\circ\right.$ is a local maximum of $P(H)$ [1] $\left.\right)$, then those and only those elements, which are rejected "from the right" by the minimum-seeking procedure retained "from the left" by maximum-seeking procedure, while those and only those elements, which are retained "from the left" by minimum-seeking procedure are rejected "from the right" by the maximum-seeking procedure.

These two rules lead to a branch-and-bound algorithm for the problem (1), which is similar to Cherenin's algorithm [2] for the solution of the problem (I.10). However, in this case, when the algorithm hits a local minimum (a node $[A,B]$ such that either $A = H_\circ$ or $B = H_\circ$), this does not necessarily guarantee direct transition to a leaf. Therefore, two rejection rules in no way reduce the path from local minimum to a leaf, and it is essential to supplement this algorithm with bounds of the type (I.14) and (I.15).

[2] Concerning the branch-and-bound scheme (and in particular, Cherenin's algorithm), see [1].

In addition to the monotone system $\langle B,\pi,F\rangle$ generated[3] by function $P(H)$, $H\subseteq B$, we also introduce an auxiliary system $\langle B\setminus A,\pi',F'\rangle$,

$$\pi'(i,H)=\pi(i,H\cup A)\ \ \forall H\subseteq(B\setminus A)\ \ \forall i\in H\,. \tag{5}$$

If $J_{B\setminus A}$ is an order on $B\setminus A$, and $\Omega_{B\setminus A}=\{J_{B\setminus A}\}$ is the set of all orders on $B\setminus A$, then the function $\Pi'(J_{B\setminus A},H)$, $H\subseteq B\setminus A$ is defined by the system $\langle B\setminus A,\pi',F\rangle$ according to (I.3): $\Pi'(J_{B\setminus A},H)=\pi'(i_{J_{B\setminus A}},H)$, where $i_{J_{B\setminus A}}$ is the first element of the set H in the sequence $J_{B\setminus A}$.

We also introduce

$$\tilde{F}(A,B,J_{B\setminus A})=\min_{H\subseteq B\setminus A}\Pi'(J_{B\setminus A},H)\,, \tag{6}$$

$$\tilde{P}_t(A,B,J_{B\setminus A})=P(A)+\left(t-|A|\right)\cdot\tilde{F}(A,B,J_{B\setminus A})\,, \tag{7}$$

$$F(A,B)=\min_{H\subseteq B\setminus A}\ \max_{J_{B\setminus A}\in\Omega_{B\setminus A}}\Pi'(J_{B\setminus A},H)\,, \tag{8}$$

$$\tilde{P}_t(A,B)=P(A)+\left(t-|A|\right)\cdot\tilde{F}(A,B)\,,\ t\le|B|\,. \tag{9}$$

Theorem 1. 1.) The quantities (7) for various $J_{B\setminus A}\in\Omega_{B\setminus A}$ form the class of lower bounds on $P(H)$ for any $H\in[A,B]$, $|H|=t$. 2.) When the reversal of the first rejection rule is inapplicable to the interval $[A,B]$, the quantities (7) with $t=|B|$ for various $J_{B\setminus A}\in\Omega_{B\setminus A}$ form the class of lower bounds on $P(H)$ for any $H\in[A,B]$. 3.) For any fixed t such that $t\le|B|$, the quantity (9) is an unimprovable bound in the class of bounds $\tilde{P}_t(A,B,J_{B\setminus A})$, $J_{B\setminus A}\in\Omega_{B\setminus A}$, defined by (7). 4.) The algorithm evaluating bounds (9) runs in time $O(|B\setminus A|^2/2)$.

Theorem 1 provides lower bounds upon $P(H)$ on fixed-cardinality subsets in the form

$$P(H)\le P(B)-\left(|B|-k\right)\cdot\tilde{F}(A,B)\,,$$

which, in particular, constitutes an immediate improvement of (I.12).

The theorem provides the basis for the following branch-and-bound algorithm solving the problem (1).

[3] In the sense of definition (I.5).

Solution tree. The standard branching scheme is used,

$$[A \cup \{i\}, B] \leftarrow [A,B] \rightarrow [A, B \setminus \{i\}],$$

where $i \in B \setminus A$; the initial interval is $[\varnothing, W]$; the best value so far is $r = min(P(A), P(B))$.

Note that the branch-and-bound strategy [4] chooses the element i so that it produces the maximum possible increment of the bound in the course of downward motion trough the solution tree. In this case, such choice is indeed feasible.

Let $\left(G_B^A\right)'$ be the maximum-cardinality core of the monotone system $\langle B \setminus A, \pi', F' \rangle$. Let $\left(G_{B\setminus\{i\}}^A\right)'$ and $\left(G_B^{A\cup\{i\}}\right)'$ denote the corresponding maximum-cardinality cores of monotone systems on the intervals $[A, B\setminus\{i\}]$ and $[A\cup\{i\}, B]$.

Theorem 2. 1.) Let $i \in (B \setminus A) \setminus \left(G_B^A\right)'$. Then the bound $\tilde{P}_t(A,B)$ (with t fixed) is not increasing on passing to the node $[A, B\setminus\{i\}]$. 2.) Let $i \in \left(G_B^A\right)'$. Then the bound $\tilde{P}_t(A,B)$ (t fixed) may only increase on passing to the node $[A, B\setminus\{i\}]$. 3.) Let i be the last element of the maximum defining sequence of the monotone system $\langle B \setminus A, \pi', F' \rangle$ ($i \in \left(G_B^A\right)'$). Then the bound $\tilde{P}_t(A,B)$ (t fixed) may only increase on passing to the node $[A\cup\{i\}, B]$.

The first two propositions of Theorem 2 show that the choice of the element i from the core $\left(G_B^A\right)'$ for current reduction of the interval is preferred to the choice of any other element from the complement $(B \setminus A) \setminus \left(G_B^A\right)'$. The third proposition identifies a unique core element as the best.

Choice of initial value. In the initial step, the best value should be chosen as the least value obtained by executing the following three procedures.

Procedure 1. Find the "greedy" solution for $k = N$ and evaluate $P(H)$ on its complement. This procedure is based on the analogy between the rejection rules and their reversals, and also on the closeness of the greedy solution to the maximum.

Procedure 2. Apply the greedy procedure to the function $P(H)$ (see [1]) and take the value $P(H)$ on this solution. This procedure is based on the symmetry of the function $P(H)$ and $\overline{P}(H) = P(W \setminus H)$.

Procedure 3. Continue constructing the maximum defining sequence until $\pi(i_k, H_k)$ is less than zero. This procedure is based on the analogy (“reversibility”) of the greedy sequence and the maximum defining sequence (the greedy procedure extends the sequence, while the maximum defining sequence is truncated by maximum $\pi(i_k, H_k)$.

Note that if $P(H)$ is a monotone function, we can apply these heuristics so as to choose only subsets of maximum cardinality k.

Traversal of the solution tree is performed as in the algorithm of [1] (reversing the inequalities and evaluating minima instead of maxima).

Since the algorithm for the problem (I.11) and (I.12) is a modification of the algorithm for the problem (I.10), the algorithms for the problems (2) and (3) are also constructed as a similar modification of the algorithm for the problem (1). The scheme of solution of the problem (4) is symmetrical to the scheme of solution of the problem (I.13) with the necessary changes in the branching procedure.

3. Application of the Algorithms in Aggregation Problems

All the examples considered in this section reduce the aggregation problem to the solution of the problem (4).

Some of these problems are reducible to global minimization of nonsubmodular functions from the class

$$P_{\sim}(H) = \alpha(|H|) \cdot P(H) + \beta(|H|),$$

where $P(H)$ is a submodular function, $\alpha(x)$, $\beta(x)$ are arbitrary real functions, $H \subseteq W$.

Seeing that the solution subsets of the problems $P_{\sim}(H) \to \min\limits_{|H|=k\le N}$ and $P(H) \to \min\limits_{|H|=k\le N}$ coincide, global minimization of the function $P_{\sim}(H)$ can be performed by enumeration of its values on the solutions of the problem (4) for all $k=\overline{1,N}$.

Following the usual practice of aggregation problems [5], we will use “object-feature” and “object-object” (or “feature-feature”) matrices as the input.

Let X ($|X|=n$) be the set of objects, Y ($|Y|=m$) the set of features (parameters); $\Phi = \left\| \varphi_i^j \right\|$, $i=\overline{1,n}$, $j=\overline{1,m}$ the “object-feature” matrix, φ_i the i-th row and φ^j the j-th column of the matrix Φ.

Let W ($|W|=N$) be the set of elements such that either $W=X$ or $W=Y$, and $A=\left\| a_{ij} \right\|_{i=\overline{1,M}}^{j=\overline{1,M}}$ the matrix of pair association coefficients between the elements of this set.

<u>1°. Minimization of the Rank Function of a Matrix Matroid</u>. One of the central problems of factor analysis is to isolate a subspace of strongly associated (in some sense) feature [5]; in regression analysis, the corresponding problem is to select subspaces of features, which are maximally correlated with a designated output feature [6]. Problems of this kind can be formulated as follows in terms of submodular functions.

Let (Y,F) be a matrix matroid, where Y is the set of features – the columns of the matrix Φ, F the set of linearly independent subsets of the set Y. We know [4] that in this case the rank function $r(H)$ ($H \subseteq Y$) is the rank of the submatrix Φ^H of the matrix Φ. Then the problem to find H such that

$$r(H) \to \min_{|H|=k},$$

precisely defines the subspace, which allows “maxima” reduction of dimensionality (for a fixed number k of input features).

Now, in addition to Φ, we have the column vector z ($\dim z = n$). Consider the function $r'(H) = rank(\Phi^H, z)$, where (Φ^H, z) is a $\{n \times (|H|+1)\}$ matrix formed by combining the submatrix Φ^H with the column z. In this case, the solution of the problem

$$r'(H) \to \min_{|H|=k}$$

isolates a k-element subset of the input features that are maximally correlated with the output feature [4] z. Therefore, the regression equation should be constructed on the system (z, H^*), where H^* is the solution of the last problem. [5]

2°. Formation of Elementary Propositions (selection of homogeneous submatrices) in Linguistic Analysis of Data Matrices [7]. This problem is central in various methods of linguistic analysis [5,7], which however only consider crudely approximate algorithms for its solution. In terms of extremization of a supermodular function, this problem can be stated as follows.

Let Φ be a data matrix with nonnegative elements.

Let $B(X,Y) = \{H \mid H = H_1 \cup H_2, H_1 \subseteq X, H_2 \subseteq Y\}$. On $B(X,Y)$ define the function

$$P(H) = \sum_{\substack{i \in H_1 \\ j \in H_2}} \varphi_i^j .$$

This function is obviously supermodular. The solution of the problem

$$P(H) \to \max_{|H|=k}$$

is a submatrix with total number of rows and columns equal to k; this submatrix is homogenous in the sense that all its elements have relatively large (and therefore close) values. Successive extraction of such submatrices from the matrix Φ until the latter is exhausted generates a linguistic description of the matrix (i.e., a system of elementary propositions, which characterize the qualitative distribution of the empirical data in Φ).

---

4 Note that the functions $r(H)$ and $r'(H)$ are monotone ($H' \subseteq H \Rightarrow r(H) \geq r(H')$, and similarly for $r'(H)$), so that we can choose the initial best value as suggested in Sec. 2. Moreover, since the problems (3) and (4) in this case have the same solution, the interval analyzing algorithm may use the reversals of the rejection rules in addition to the bound (9).

5 As shown in [8], the monotonicity property of the set functions $P(H)$ in itself is sufficient in order to pose and effectively solve the problem of finding all the so-called V-extrema – $H^* \subseteq W$ such that $P(H^*) > V$ and $\forall H \subset H^*$, $P(H^*) \leq V$, where V is a prespecified number. It is easily seen [8] that this problem also can be interpreted as the problem of selecting a “defining” regression subspace.

3°. Aggregation Using the Association Matrix. Let $A=\|a_{ij}\|$ $\forall i,j$, $a_{ij}\geq 0$. Consider the following two set functions,

$$P_1(H)=\sum_{i,j\in H} a_{ij}\ ,\ P_2(H)=\sum_{\substack{i\in H\\ j\in\overline{H}}} a_{ij}\ ,$$

where $H\subseteq W$, $H\neq\varnothing$, $\overline{H}=W\setminus H$, and $P_1(\varnothing)=P_2(\varnothing)=0$.

The function $P_1(H)$ is clearly supermodular and $P_2(H)$ is submodular [9]. From this pair of functions, we can construct the following parametric family of functions [6]:

$$P_{\ell t}(H)=\ell\cdot P_1(H)+t\cdot P_2(H),\ H\subseteq W,\ \ell,t\in\Re. \qquad (10)$$

Theorem 3. For $t>2\cdot\ell$, the function $P_{\ell t}(H)$ is supermodular, and for $t<2\cdot\ell$, it is submodular; for $t=2\cdot\ell$ it is modular.

The family of functions (10) opens wide horizons for the construction of various aggregation problems. Let us consider four examples with integer parameters ℓ and t.

Example 1. We use the following method for selecting the set of class "centers" in automatic classification of objects into a given number k of classes. Let $\ell=1$ and $t=0$. Then $P_{\ell t}(H)=P_1(H)$.

For a fixed number of classes k, the set of class "centers" is the set H^* obtained by solving the problem

$$P_1(H)\to\max_{|H|=k}. \qquad (11)$$

Noting that $(-P_1(H))$ is a submodular function, this solution can be obtained by the minimization algorithm (4) described in Sec. 2.[7]

---

[6] This family of functions is generated, according to Theorem I.3, from the family of monotone systems $\pi_{\ell t}(i,H)=\ell\cdot\sum_{j\in H} a_{ij}+t\cdot\sum_{j\in\overline{H}} a_{ij}$ with constant derivative.

[7] Note that the function $P_1(H)$ is antitone (i.e., $H'\subseteq H\Rightarrow(-P_1(H'))\leq(-P_1(H))$ and therefore the problem of finding V-extrema [8] of the function $P_1(H)$ is similar of the problem (11) if we consider it with the constraint $|H|=k$.

Example 2. Now consider the automatic classification problem of objects with the matrix A into an unspecified number of classes. In this case, the set of "centers" of the sought classification is the set $H^* \subseteq W$, obtained by solving the problem [8]

$$P_3(H) = \frac{1}{|H|} \cdot \sum_{\substack{i \in H \\ j \in H}} a_{ij} \to \max_{H \subset W}.$$

Clearly, although $P_3(H)$ is not supermodular, the solution is obtained by enumeration [9] of the solutions of the problem (11) for all $k = \overline{2,N}$.

Example 3. The problem of finding the centroid component reduces [1] to maximizing the functional

$$I = \sum \sigma_i \sigma_j \cdot \rho_{\varphi^i \varphi^j} \to \max \qquad (12)$$

where $\left\| \rho_{\varphi^i \varphi^j} \right\|_M^M$ is the matrix of sample correlation coefficients between the parameters from the set Y, and the maximum is on the set of all tuples $\sigma = \langle \sigma_1, \sigma_2, ..., \sigma_M \rangle$, where $|\sigma_i| = 1$, $i = \overline{1,M}$.

Using the function $P_{\ell t}(H)$ from (10) with $\ell = 0$, and $t = 1$, i.e., $P_2(H)$, we consider the cut function

$$P_4(H) = \begin{cases} \sum\limits_{\substack{i \in H \\ j \in \overline{H}}} \rho_{\varphi^i \varphi^j}, & \text{if } H \neq \varnothing, H \neq Y, \\ 0 & \text{when } H = \varnothing \text{ or } Y = \varnothing. \end{cases}$$

Make the transformation

$$a_{ij} = \rho_{\varphi^i \varphi^j} + h,$$

where h is a number such that

$$h \geq \max_{i,j = \overline{1,M}} \left| \rho_{\varphi^i \varphi^j} \right|.$$

---

8 Other formulations of this problem are proposed in [10] and [11].

9 The function $(-P_3(H))$ is a particular case of the class of functions $P_\sim(H)$ described at the beginning of this section, which are extremized by such enumeration.

Then

$$P_2(H) = P_4(H) + |H| \cdot |\overline{H}| \cdot h,$$

where $\overline{H} = W \setminus H$. The function $P_4(H)$ belongs to the class of nonsubmodular functions $P_{\sim}(H)$. Therefore, for a fixed $|H|$, the solution of the problem $P_4(H) \to min$ coincides with the solution of the problem $P_2(H) \to min$. Hence it follows that the global minimum of $P_4(H)$ may be found by enumerating its values on the corresponding set of minimum points of $P_2(H)$; for even $|W|$, it is necessary to enumerate the minima of $P_2(H)$ for all cardinalities $|H|$ from $\{1,2,...,|W|/2\}$, whereas for odd $|W|$, for all cardinalities $\{1,2,...,(|W|+1)/2\}$.

Minimization of $P_2(H)$ for $|H| = const$ is problem (4).

<u>Example 4</u>. Aggregation of an empirical graph [5] requires algorithms for the selection of a certain subset from the vertex set of the graph. The elements of this subset should be "distant" in a certain sense from most vertices and from one another [13]. In order to formulate the corresponding problem, we construct the submodular function

$$P(H) = \sum_{\substack{i \in H \\ j \in W}} a_{ij} + \sum_{\substack{i \in W \\ j \in H}} a_{ij}, \; H \subseteq W. \tag{13}$$

Let $\|\delta_{ij}\|$ be the adjacency matrix of the original graph G, and $\|\alpha_{ij}\|$ the matrix of weights responsible for the "interaction" of vertex i with vertex j. The matrix $\|a_{ij}\|$ is defined as $a_{ij} = \delta_{ij} \cdot \alpha_{ij}$. Then sought subset H^* of vertices which play "a certain role" in the system W (of interacting elements) can be formulated as the solution of the problem

$$P(H) \to \min_{|H|=k} \tag{14}$$

Note that the first term in (13) is the function with $\ell = 1$, $t = 1$ from (10), whereas the second term is the same function on the matrix A^T. By Theorem 3, both these functions are submodular and the function (13) therefore is also submodular. This implies that the problem (14) is reducible to the problem (4).

APPENDIX

<u>Proof of Theorem 1</u>. In order to prove the first proposition of the theorem, we use the expansion (I.6) and definition (5). Then

$$P(H) = P(A) + \pi'(i_t, \{i_t\}) + \pi'(i_{t-1}, \{i_{t-1}, i_t\}) + \ldots + \pi'(i_1, \{i_1, i_2, \ldots, i_t\}), \tag{A.1}$$

where $\langle i_1, i_2, \ldots, i_t \rangle$ is a suborder of the order $J_{B\backslash A}$, $H \in [A,B]$. Then we obtain the sought proposition by (I.3) and (6).

To prove the second proposition, it suffices to note that if the reversal of the second rejection rule does not hold, then $\pi'(i, B\setminus A) \le 0 \ \ \forall i \in B\setminus A$, i.e., the same is also true for the element i_1 in the expansion (A.1). By (6), (I.3), we thus obtain $F(A,B,J_{B\backslash A}) < 0$. This immediately proves the second proposition.

The third proposition is easily obtained from (7) and (6), arbitrariness of the order $J_{B\backslash A}$ and the corollary of the duality theorem.

Finally, the fourth proposition follows directly from the existence of an effective algorithm to isolate the maximum core of a monotone system. ■

<u>Proof of Theorem 2</u>. Note that by (8), $\tilde{F}(A,B) = F((G_B^A)')$. Then the first proposition of the theorem clearly follows from the definition of the core of a monotone system. The second proposition follows from the definition of the core and from the fact that $\forall H \subseteq B\setminus A$, $\pi(j,H) \le \pi(j, H\setminus\{i\})$, $j \in B\setminus(A\setminus\{i\})$ (by definition of monotone system).

Let us prove the third proposition. We have

$$\tilde{P}_t(A,B) = P(A) + (t - |A|)\cdot \tilde{F}(A,B),$$

$$P_t(A\cup\{i\}) = P(A) + \pi'(i,\{i\}) + (t-|A|)\cdot \tilde{F}(A\cup\{i\},B) - \tilde{F}(A\cup\{i\},B),$$

whence

$$\tilde{P}_t(A\cup\{i\},B) - \tilde{P}_t(A,B) = \pi'(i,\{i\}) - \tilde{F}(A\cup\{i\},B) +$$
$$+ (t-|A|)\cdot\left[\tilde{F}(A\cup\{i\},B) - \tilde{F}(A\cup\{i\},B)\right].$$

Consider two cases:

a) $\left(G_B^A\right)' = \{i\}$.

In this case, clearly,

$$\pi'(i,\{i\}) = \tilde{F}(A,B), \text{ and}$$

$$\tilde{P}_t(A \cup \{i\},B) - \tilde{P}_t(A,B) = \left(t - |A| - 1\right) \cdot \left[\tilde{F}(A \cup \{i\},B) - \tilde{F}(A,B)\right].$$

It is also clear that $\forall\, H \in [A \cup \{i\}]$

$$F'_{A \cup \{i\},B}(H) = F'_{A,B}(H),$$

i.e., by the definition of the core and arbitrariness of H,

$$\tilde{F}(A \cup \{i\},B) = \min_{H \subseteq H \setminus (A \cup \{i\})} F'_{A \cup \{i\},B}(H) \geq F'_{A,B}(\left(G_B^A\right)') = \tilde{F}(A,B),$$

whence

$$\tilde{P}_t(A \cup \{i\},B) - \tilde{P}_t(A,B) \geq 0; \tag{A.2}$$

b) $\left(G_B^A\right)' = i$.

Then, seeing that i is the last element in the defining sequence, we conclude that the weights on the remaining elements in the new defining sequence do not change. Since by assumption $\left(G_B^A\right)' \neq \{i\}$, we have

$$\tilde{F}(A \cup \{i\},B) = \tilde{F}(A,B)$$

and therefore $\tilde{F}(A \cup \{i\},B) \leq \pi'(i,\{i\})$, whence we also obtain (A.2). ■

<u>Proof of Theorem 3</u>. Take the derivative of the function $P_{\ell t}(H)$ in the sense of definition (I.5). We have

$$\pi_{\ell t}(i,H) = 2 \cdot \left[\ell \cdot \sum_{j \in H} a_{ij} + (t - \ell) \cdot \sum_{j \in H} a_{ij}\right],\ i \in H \subseteq W.$$

Let $t - \ell > \ell$, i.e., $t > 2 \cdot \ell$. Then

$$\pi_{\ell t}(i,H) = 2 \cdot \left[\ell \cdot \sum_{j \in H} a_{ij} + (t - 2 \cdot \ell) \cdot \sum_{j \in H} a_{ij}\right]. \tag{A.3}$$

Clearly, the system (A.3) is monotone, and therefore $P_{\ell t}(H)$ is supermodular.

Arguing as above, we prove the remaining propositions. ■

LITERATURE CITED

1. I. B. Muchnik and L. V. Shvartser, “Submodular set functions and monotone systems in aggregation problems, I,” *Avtomat. Telemekh., No. 5*, 135-148 (1987), http://www.datalaundering.com/download/submod01.pdf.
2. M. Grotchel, L. Lovasz, and A. Schrijver, “The ellipsoid method and its consequences in combinatorial optimization,” *Combinatorica, No. 1*, 169-197 (1981).
3. D. T. Phillips and A. Garcia-Diaz, Network Analysis Methods [Russian translation], *Mir*, Moscow (1984).
4. H. Papadimitriu and K. Steiglitz, Combinatorial Optimization. Algorithms and Complexity [Russian translation], *Mir*, Moscow (1985).
5. É. M. Braverman and I. B. Muchnik, Structural Methods of Processing Empirical Data [in Russian], *Nauka*, Moscow (1983).
6. S. A. Dubrovskii, Applied Multivariate Statistical Analysis [in Russian], *Finansy i Statistika*, Moscow (1982).
7. É. M. Braverman, N. E. Kiseleva, I. B. Muchnik, and S. G. Novikov, “A linguistic approach to the problem of processing of large data arrays,” *Avtomat. Telemekh., No. 11*, 73-88 (1974).
8. M. V. Sapir, “Analysis of the association structure of empirical data,” *NTI, Ser. 2*, No. 8, 75-88 (1985).
9. D. M. Topkis, “Minimizing a submodular function on a lattice,” *Oper. Res., 26, No. 2*, 305-320 (1978).
10. I. A. Rubina, “Application of the method of succesive calculations to the problem of automatic classification, “*Avtomat. Telemekh., No. 1*, 126-132 (1986).
11. 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.
12. G. N. Vostrov, “On extremal clustering of parameters in problems on weighted graphs,” *Avtomat. Telemekh., No. 1*, 78-88 (1978).
13. I. B. Muchnik, “Processing of expert judgements about structured objects,” *Proc. of the Institute of Control Problems [in Russian], Inst. Probl. Upravl.*, Moscow (1982), pp. 27-32.