PROCEEDINGS OF THE 7th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION AND IMAGE ANALYSIS: NEW INFORMATION TECHNOLOGIES (PRIA-7-2004)
Mathematical Methods in Pattern Recognition

Computational Complexity of the Minimal Committee Problem and Adjacent Problems

M. Yu. Khachai

Institute of Mathematics and Mechanics, Ural Division, Russian Academy of Sciences, ul. S. Kovalevskoi 16, Yekaterinburg, 620219 Russia
e-mail: mkhachay@imm.uran.ru

Abstract—The computational complexity of two important special cases of the minimal committee problem (MC), viz., the problem on the minimal committee of finite sets (MCFS) and the problem on the minimal committee of a system of linear algebraic inequalities (MCLE), is studied. Both problems are shown to be *NP*-hard. Separately, some adjacent problems of integer optimization are shown to be intractable. The efficient approximability threshold is estimated for the MCFS problem, the estimates being allied to the results known for the set cover problem. The intractable and polynomially solvable subclasses of the MCLE problem are given. The problem of the minimal affine separation committee (MASC) is considered in conclusion; the results obtained earlier for the MCLE problem are shown to be valid for this problem as well.

DOI: 10.1134/S1054661806040195

INTRODUCTION

In this work, we consider a problem of combinatorial optimization, viz., the problem of the minimal committee (MC). The MC problem is closely connected with three areas of operations research: voting theory, optimization, and pattern recognition. Voting theory studies collective decision-making procedures based on different logics of counting votes (democracies). A committee is a mathematical model of voting on the basis of simple majority rule.

The so-called perceptron algorithms, which go back to Rozenblatt's works, are rather important in pattern recognition theory. It is not difficult to prove that a two-layer perceptron with the additional condition of nonnegativity of the weights of the neuron at the second layer and the appropriate threshold value is mathematically equivalent to a committee.

Finally, so-called improper problems are rather common in optimization theory [1]. There are several conditions under which an optimization problem is improper. In terms of linear programming, for instance, a problem is improper if the direct, dual, or both systems of restrictions are inconsistent.

Improper problems can be corrected by several strategies. One of them suggests minimal perturbation of the initial problem parameters so that the resulting problem becomes solvable. Another strategy changes the very notion of the solution to an improper problem. Instead of one element of the reference space satisfying all restrictions, a committee of "pseudosolutions" is introduced, each of which satisfies a sufficiently large part of constraints of the initial problem. Committee solution theory [2] is one of mathematical formalizations of this approach.

By virtue of a number of objective reasons, when solving all of the described above problems in the framework of committee solution theory, it is desirable to obtain committee solutions of the simplest structure. This results in stating the minimal committee problem. In this work, we study its computational complexity.

This paper consists of four sections.

In the first section, we state the minimal committee problem for an abstract system of inclusions and the adjacent problems of integer optimization. The second section deals with a special case of the problem, in which the inclusions are given by finite sets. In this case, the problem is known to be *NP*-hard. We show that the adjacent problems of integer linear programming possess a similar property. We also estimate the approximation threshold for this problem, i.e., the maximum admissible approximation quality of the optimum of the problem, for which an efficient approximate algorithm can be constructed.

In the third section, we consider another important special case of the problem, viz., the minimal committee problem of an inconsistent system of linear algebraic inequalities. We show this problem to be *NP*-hard as well. Its adjacent problem of the minimal affine separation committee of majority possesses the similar property. Further, we list the known polynomially solvable subclasses of the problem.

In the fourth and final section, we study the computational complexity of the problem of the minimal affine separation committee for two finite sets. We show that this problem is polynomially equivalent to the MCLE problem and, hence, the results similar to those obtained for the MCLE problem are valid for the MASC problem as well.

Received April 4, 2005

1. STATEMENT OF THE MINIMAL COMMITTEE PROBLEM

Let a set X and a set $D_1, D_2, \ldots, D_m$ of its nonempty subsets be given. We consider a system of abstract inclusions

$$x \in D_j, \quad (j \in \mathbb{N}_m = \{1, 2, \ldots, m\}). \tag{1}$$

System (1) is not necessarily consistent; i.e., the relation $\cap D_j = \emptyset$ can be fulfilled. As usual (see, e.g., [2]), by a committee solution of q elements of system (1) (or just the committee) we mean a finite sequence $Q = (x^1, x^2, \ldots, x^q)$ satisfying the condition

$$\left|\{i : x^i \in D_j\}\right| > \frac{q}{2}$$

for every $j \in \mathbb{N}_m$.

We state several problems of combinatorial optimization.

The Minimal Committee Problem (MC):

*Let a set X and a set of its nonemp*ty subsets D_1, D_2, …, D_m be given. Find the committee solution to system (1) with minimum possible q (or prove that there are no committee solutions to this system).

Following [3], it is convenient to restate the MC problem in terms of integer linear programming. Let J_1, J_2, …, J_T be the index sets of all consistent subsystems of system (1) that are maximal with respect to inclusion. Obviously, a system is consistent if and only if $T = 1$; otherwise, $1 < T < 2^m$. We define two $m \times T$ incidence matrices A and B by the rule

$$a_{ji} = 1, \quad b_{ji} = 1 \text{ if } j \in J_i;$$
$$a_{ji} = 0, \quad b_{ji} = -1, \text{ otherwise},$$

and consider the problems of integer linear programming

$$\min\{(e, t) | Bt \geq f, t \in \mathbb{Z}_+^T\} \tag{2}$$

and

$$\min\left\{s: \begin{array}{ll} At \geq sf, & t \in \mathbb{Z}_+^T \\ (e, t) \leq 2s - 1, & s \in \mathbb{N} \end{array}\right\}. \tag{3}$$

Here, e and f are the vectors that consist of units and belong to spaces E_T and E_m, respectively. The following theorem is known.

Theorem 1 ([4]). The MC problem and problems (2) and (3) are simultaneously either solvable or unsolvable. The sets of optimal solutions to problems (2) and (3) are isomorphically embedded in the set of solutions to the MC problem (with the committee solutions consisting of minimal number of elements).

Along with the MC problem, we consider two special cases of it.

(1) The MC problem with a finite set X (and all its subsets D_j) (we refer to it as the MCFS problem). In Section 2, we show that Theorem 1 can be refined for this problem. We also show that this special case of the minimal committee problem is *NP*-hard, which, in particular, results in intractability of the MC problem in the general case. Moreover, we show that solving it approximately with a given accuracy is an intractable problem as well.

(2) The MC problem, in which the set X is a finite-dimensional number space $\mathbb{Q}^n$ and the subsets D_j are its half-spaces. We refer to this problem as the problem of the minimal committee of a system of linear inequalities or the MCLE problem. In Section 3, we show it to be *NP*-hard as well and list some of its polynomially solvable subclasses.

2. COMPUTATIONAL COMPLEXITY OF THE MINIMAL COMMITTEE PROBLEM

In this section, we justify the intractability of the MC problem. To do this, we show its special case—the MCFS problem with all inclusions given by finite sets—to be intractable. Then, we show that this problem and problems (2) and (3) corresponding to it are in a sense "equally intractable." In conclusion, we show that, under reasonable assumptions, this problem cannot be approximately solved with sufficient accuracy in polynomial time and give the corresponding approximability threshold.

2.1. The Intractability of the MCFS Problem

Thus, in this section, we mainly study the problem of the minimal committee of finite sets (MCFS):

Let a set $X = \{x^1, x^2, \ldots, x^p\}$ and a set of its subsets $D_1, D_2, \ldots, D_m$ be given. Find the committee of system (1) with the minimal number of elements (or show that there are no committee solutions to this system).

We choose a method for coding the problem conditions and solution. Let the conditions of the particular MCFS problem be given by an $m \times p$ matrix C filled with 1 to –1, and

$$c_{ji} = \begin{cases} 1, & \text{if } x^i \in D_j, \\ -1, & \text{otherwise.} \end{cases} \tag{4}$$

Without loss of generality, we can suppose that, for an arbitrary committee $Q = (y^1, \ldots, y^q)$, there can be found natural numbers $k \leq p$,

$$q_1, q_2, \ldots, q_k: \; q_1 + q_2 + \ldots + q_k = q,$$

and numbers

$$i_1, i_2, \ldots, i_k: \; 1 \leq i_1 < i_2 < \ldots < i_k \leq p,$$

such that

$$y^1 = y^2 = \ldots = y^{q_1} = x^{i_1},$$

$$y^{q_1+1} = y^{q_1+2} = \dots = y^{q_1+q_2} = x^{i_2},$$
$$\dots \quad \dots$$
$$y^{q_1+\dots+q_{k-1}+1} = \dots = y^{q} = x^{i_k}.$$

In what follows, the order of the elements in the sequence Q is of no interest to us; therefore, we agree to represent it in the form

$$Q = \{(x^{i_1}, q_1), (x^{i_2}, q_2), \dots, (x^{i_k}, q_k)\},$$

which is characteristic of multisets. Here, the numbers q_j stand for the multiplicities of the elements x^{i_j}.

Theorem 2. The MCFS problem is *NP*-hard.

Proof. It is sufficient to prove that the following problem of property recognition is *NP*-complete.

The Committee Problem (COM):

Let subsets $D_1, D_2, \dots, D_m$ *of* a finite set X and a number $k \in \mathbb{N}$ be given. Determine if there exists a committee of system (1) consisting of at most $2k - 1$ elements.

To prove the theorem, we apply the standard technique and show that the hitting set problem, which is proved to be *NP*-complete [5], can be polynomially reduced to the stated (according to Karp) problem.

The Hitting Set Problem:

Let subsets $C_1, C_2, \dots, C_n$ of a finite set S and a number k be given. Determine whether, for these subsets, there exists a system M of representatives with no more than k elements.

To prove the polynomial reducibility, it is sufficient, for an arbitrary finite set S, the set of its subsets $C_1, \dots, C_n$, and the number k, to indicate—in the time polynomial in the description length of the initial problem—a set X and its subsets $D_1, \dots, D_m$ such that the sets $C_1, \dots, C_n$ have a system of representatives with at most k elements if and only if system (1) has a committee whose cardinality does not exceed $2k - 1$.

Let $S = \{s^1, s^2, \dots, s^t\}$, and $C_1, \dots, C_n \subset S$ and number $k \in \mathbb{N}$ be given. We choose $s^0 \notin S$ and put $X = S \cup \{s^0\}$ and $m = n + 1$. We put

$$\begin{cases} S_j = C_j \cup \{s^0\} \ \ (j \in \mathbb{N}_n), \\ D_{n+1} = S. \end{cases} \tag{5}$$

This procedure can obviously be fulfilled in time $O(m + t)$. Indeed, without loss of generality, we can suppose that the hitting set is given by n t-bit binary numbers (it is also necessary to code the number k; however, this is not important in our case). To code the conditions of the committee problem according to (5), it is sufficient to upend one bit to each of them and add one $(t + 1)$-digit number with all digits set except for this of the lowest order.

Let M be the hitting system for the sets $C_1, \dots, C_n$ and $L(M) = \{s^{i_1}, \dots, s^{i_l}\}$ be the set of its elements, $l \le k$. It is not difficult to see that the sequence

$$K = (\underbrace{s^0, \dots, s^0}_{l-1}, s^{i_1}, \dots, s^{i_l})$$

is the committee of system (1) by virtue of the construction and that it consists of

$$2l - 1 \le 2k - 1$$

elements.

Inversely, let

$$K = (\underbrace{s^0, \dots, s^0}_{r}, s^{i_1}, \dots, s^{i_l})$$

be the committee of system (1). Without loss of generality, we can suppose that its cardinality is an odd number. Let

$$r + l = 2k' - 1 \le 2k - 1.$$

Since $D_m = S$ and $s_0 \notin S$, it follows that $r \le l - 1$ and, hence, $l \le k'$. Since K is the committee, there can be found an element $s^{i(j)} \in C_j$ among $s^{i_1}, \dots, s^{i_{k'}}$ for every $j \in \mathbb{N}_n$. Hence, the sequence

$$M = (s^{i(1)}, s^{i(2)}, \dots, s^{i(n)})$$

is a hitting system for the sets $C_1, \dots, C_n$. By virtue of the construction, M has at most $k' \le k$ elements.

Thus, the hitting set problem is shown to be polynomially reducible to the committee problem. Hence, the latter is *NP*-complete and the problem of searching the minimal committee of finite sets (MCFS) is *NP*-hard.

Remark 1. The proof of this theorem was first published in [6].

Remark 2. It is known [5] that the hitting set problem remains *NP*-complete even if we add the extra condition

$$|C_j| \le 2 \quad j \in \mathbb{N}_n.$$

Using a line of reasoning similar to that in the proof of the theorem, we can verify that the committee problem remains *NP*-complete if the condition $|D_j| \le 3$ is met for all sets D_j except for, perhaps, one. Thus, the MCFS problem remains *NP*-hard under the same condition.

Remark 3. The minimal committee problem (MC) of arbitrary system (1) with known maximum consistent subsystems is polynomially equivalent to the above-considered problem with finite sets D_j and is therefore *NP*-hard as well.

Let us establish the correspondence between the MCFS problem and problems of integer optimization (2) and (3) introduced earlier.

Theorem 3. The MCFS problem and problems (2) and (3) are polynomially equivalent.

Proof. We show that MCFS and (2) are mutually Turing reducible; the equivalence of (2) and (3) is obvious. Suppose that an $m \times p$ matrix C determines the instance of MCFS problem according to (4) and that the matrix B consists of its pairwise nondominated columns. Without loss of generality, we assume the columns of B to be the first T columns of the matrix C. Problem (2) induced by the matrix B is the one sought. Indeed, let the vector $\bar{t} = \{\bar{t}_1, \bar{t}_2, \ldots, \bar{t}_T\}$ be the optimal solution to problem (2). Then, by virtue of the construction of the matrix B, the sequence

$$\bar{Q} = \{(x^i, \bar{t}_i) \big| \bar{t}_i > 0, i \in \mathbb{N}_T\}$$

is the optimal solution to the MCFS problem (the minimal committee of system (1)).

Conversely, let the matrix B give the conditions of particular problem (2). Consider the MCFS problem, where $C = B$, and consider an arbitrary optimal solution to the problem, viz., the sequence

$$\bar{Q} = \{(x^{i_1}, q_1), (x^{i_2}, q_2), \ldots, (x^{i_k}, q_k)\}.$$

We put it into correspondence with the vector $\bar{t} \in \mathbb{Z}_+^T$ by the rule

$$t_\lambda = \begin{cases} q_j, & \text{if } \lambda = i_j, \\ 0, & \text{otherwise.} \end{cases}$$

By virtue of the construction, the vector $\bar{t}$ is the optimal solution to problem (2).

The theorem is proved.

Corollary 1. Problems (2) and (3) are NP-hard.

2.2. The Effective Approximability Threshold for the MCFS Problem

In Theorem 2, we proved the MCFS problem to be intractable. There arises the natural questions whether it can be solved approximately in polynomial (in the description length of its conditions) time and, if so, what the accuracy is. As shown below, under the assumption that $P \neq NP$ (or close to it), there is no polynomial approximate algorithm with any constant accuracy for this problem. To prove it, we use the similar results proved for the known set cover problem in [7, 8].

The Set Cover Problem:

Let a set S, $|S| = m$ and a nonempty set of its subsets $C = \{c_1, c_2, \ldots, c_l\}$ be given. Find the subset $C' \subseteq C$ that is minimal in cardinality and that covers S (for which $\cup C' = S$ is valid).

The following theorems are known.

Theorem 4 ([7]). If $P \neq NP$ holds, there is no approximate algorithm with the approximation accuracy $\frac{1}{4}\log m$ for the set cover problem.

Theorem 5 ([8]). For an arbitrary $\varepsilon > 0$, if there exists an approximate algorithm with the approximation accuracy $(1 - \varepsilon)\ln m$ for the set cover problem, then

$$NP \subseteq TIME(n^{O(\log\log n)}).$$

From this point on, we use the standard designation $\ln m$ for the natural logarithm and $\log m$ for the logarithm to base 2.

Theorem 5 is remarkable since it actually justifies the optimality of the known greedy algorithm for the set cover problem, whose approximation accuracy is $\ln m$ [9]. In what follows, we prove the similar propositions for the MCFS problem (this result was first published in [10]).

Lemma. If there exists an approximate algorithm with the approximation accuracy r for the MCFS problem, then there also exists a similar algorithm with the same accuracy for the set cover problem.

Proof. 1. We reduce the instance the set cover problem to the appropriate instance of the MCFS problem. Let the sets $S = \{s_1, s_2, \ldots, s_m\}$ and $C = \{c_1, c_2, \ldots, c_l\} \subseteq 2^S$ be fixed. It is sufficient to find, in time bounded from above by a polynomial of m and l, an appropriate statement of the particular MCFS problem and, then, prove that, for a natural k, the subset $C' \subseteq C$, $|C| \leq k$ is a cover of S if and only if the MCFS problem has an admissible committee solution consisting of $2k - 1$ elements.

We consider the $m \times l$ incidence matrix A of the sets S and C. As before, we put

$$a_{ji} = \begin{cases} 1, & \text{if } s_j \in c_i, \\ 0, & \text{otherwise.} \end{cases}$$

We put it into correspondence with the $(m + 1) \times (l + 1)$ matrix A' obtained from A by bordering it with the row and column of units (see the table). We put the element in the lower right corner of the constructed matrix equal to zero. Consider the MCFS problem corresponding to

Construction of the MCFS problem

		x^1 c_1	x^2 c_2	… …	x^l c_l	x^{l+1} S
D_1	s_1					1
D_2	s_2					1
⋮	⋮			A		
D_m	s_m					1
D_{m+1}	C	1	1	…	1	0

the matrix A. Let the set X equal $\{x^1, x^2, \dots, x^{l+1}\}$ and the subsets D_j be defined by the formulas

$$D_j = \{x^{l+1}\} \cup \{x^i: s_j \in c_i, i \in \mathbb{N}_l\} \quad (j \in \mathbb{N}_m)$$
$$D_{m+1} = \{x^1, x^2, \dots, x^l\}.$$

Let $C' = \{c_{i_1}, c_{i_2}, \dots, c_{i_k}\}$ be a cover of the set S, i.e., for every $j \in \mathbb{N}_m$, there exists a number

$$\mu(j) \in \mathbb{N}_k: s_j \in c_{i_{\mu(j)}},$$

which, in turn, implies $x^{i_{\mu(j)}} \in D_j$ by virtue of the construction. Then, the sequence

$$Q = (x^{i_1}, x^{i_2}, \dots, x^{i_k}, \underbrace{x^{l+1}, \dots, x^{l+1}}_{k-1})$$

is a committee of system (1) since every set D_j contains at least k elements of Q.

On the other hand, consider an admissible committee solution Q of system (1) consisting of $2k - 1$ elements:

$$Q = (x^{i_1}, x^{i_2}, \dots, x^{i_{2k-1-\lambda}}, \underbrace{x^{l+1}, \dots, x^{l+1}}_{\lambda}).$$

According to the choice of the set D_{m+1}, we have $\lambda < k$. Consider the subsequence $(x^{i_1}, \dots, x^{i_k})$. By the committee definition, for every $j \in \mathbb{N}_m$, there exists a number

$$\mu(j) \in \mathbb{N}_k: x_{i_{\mu(j)}} \in D_j.$$

By construction of the set D_j, the inclusion $s_j \in c_{i_{\mu(j)}}$ is valid for this number. Therefore, the set

$$C' = \{c_{i_\mu}: \mu \in \mathbb{N}_k\}$$

is the sought cover of the set S.

2. We prove that, if there exists an approximate algorithm for the MCFS problem with approximation accuracy r, then there also exists an approximate algorithm with the same accuracy estimate for the set cover problem. Suppose the approximate algorithm $\mathcal{A}$ with the approximation accuracy r is found for the MCFS problem.

Consider the particular set cover problem; let t be the cardinality of the minimal cover (the optimum of the problem). According to the rules described above, we put it into correspondence with an appropriate MCFS problem.

As proved above,

(1) the minimal committee of the MCFS problem has $2t - 1$ elements;

(2) every committee solution to the MCFS problem that consists of $2k - 1$ elements is in correspondence with a cover whose cardinality does not exceed k and that can be restored in polynomial time using the known committee.

Suppose the committee solution found by the approximate algorithm $\mathcal{A}$ has $2k - 1$ elements (as mentioned above, the committee of $2k$ elements can be transformed into a committee of $2k - 1$ elements by excluding an arbitrary element). According to the assumption,

$$1 \le \frac{2k-1}{2t-1} \le r;$$

hence, we can estimate the accuracy of the corresponding approximate algorithm for the set cover problem as

$$\frac{k}{t} \le r\left(1 - \frac{1}{2t}\right) + \frac{1}{2t} \le r\left(1 - \frac{1}{2t}\right) + \frac{r}{2t} \le r.$$

The following theorems immediately follow from the lemma and Theorems 4 and 5.

Theorem 6. If the hypothesis $P \neq NP$ holds, there is no approximate algorithm with the approximation accuracy $\frac{1}{4}\log(m-1)$ for the MCFS problem.

Proof. We assume the opposite: there exists an approximate algorithm $\mathcal{A}$ that finds an admissible solution to the MCFS problem with m inclusions with accuracy $\frac{1}{4}\log(m-1)$. Then, by the lemma, there exists an approximate algorithm for the set cover problem that constructs an admissible cover whose cardinality does not exceed $\frac{1}{4}\log(m-1)$ times that of the optimal one (if $|S| = m - 1$), which, by Theorem 4, implies the equality $P = NP$. This contradiction proves the theorem.

One can similarly prove the following theorem.

Theorem 7. If the condition

$$NP \not\subseteq TIME(n^{O(\log\log n)})$$

is met, then, for an arbitrary $\varepsilon > 0$, there is no approximate algorithm for solving the MCFS problem with the approximation accuracy $(1 - \varepsilon)\ln(m - 1)$.

Note that the technique of reducing the cover problem to the minimal committee problem, which was used in the proof of the lemma, can be applied when solving the problem of the committee division of sets. For instance, consider the following statement. Suppose subsets A and B of the main set X and a class of decision rules

$$\mathcal{F} = \{f(x, \alpha) | \alpha \in \Lambda\} \subseteq \{X \longrightarrow \{0, 1\}\}$$

are given. It is required to construct the committee decision rule (over the class $\mathcal{F}$) that accurately separates the sets A and B. In other words, it is necessary to find a sequence $(\alpha^1, \alpha^2, \dots, \alpha^q)$ such that

$$\operatorname{sgn}\left(\sum_{i=1}^{q} f(x, \alpha^i) - \frac{q}{2}\right) = \begin{cases} 1, & x \in A, \\ -1, & x \in B. \end{cases}$$

The following proposition holds.

Proposition 1. Let parameters $\alpha^0, \alpha^1, \dots, \alpha^k$ be fixed so that

$$f(a, \alpha^0) = 1 \quad (a \in A),$$
$$f(b, \alpha^i) = 0 \quad (b \in B, i \in \mathbb{N}_k)$$

and for every $a \in A$ there can be found a number $i = i(a)$ such that $f(a, \alpha^{i(a)}) = 1$. Then, the sequence

$$(\underbrace{\alpha^0, \dots, \alpha^0}_{k-1}, \alpha^1, \dots, \alpha^k)$$

defines the committee decision rule, which accurately separates the sets A and B.

Proof. Indeed, the resulting sequence consists of $q = 2k - 1$ elements. For an arbitrary point $a \in A$, by hypothesis, we have $f(a, \alpha^0) = 1$, and a number $i = i(a)$ can be found such that $f(a, \alpha^{i(a)}) = 1$; hence, $\sum_{i=1}^{k} f(a, \alpha^i) + (k-1)f(a, \alpha^0) \geq k > \frac{q}{2}$. Similarly, by hypothesis, for every $b \in B$ and $i \in \mathbb{N}_k$, the relation $f(b, \alpha^i) = 0$ is valid, which results in $\sum_{i=1}^{k} f(b, \alpha^i) + (k-1)f(b, \alpha^0) \leq k - 1 < \frac{q}{2}$. Q.E.D.

3. THE MINIMAL COMMITTEE PROBLEM OF A SYSTEM OF LINEAR INEQUALITIES

Let the set X coincide with the set $\mathbb{Q}^n$ of vectors with rational coefficients and the subsets D_j be the half-spaces

$$D_j = \{x \in X | (a_j, x) > 0\} \quad 0 \neq a_j \in X.$$

Then, system (1) has the form

$$(a_j, x) > 0 \quad (j \in \mathbb{N}_m). \tag{6}$$

There is another important special case of the MC problem, which is the minimal committee problem of a system of linear inequalities.

The Minimal Committee Problem of a System of Linear Inequalities (MCLE):

Let natural numbers m and $n > 1$ and vectors

$$a_1, a_2, \dots, a_m \in \mathbb{Q}^n$$

be given. Find the committee solution (the committee) of system (6) consisting of the minimum number of elements (or prove that there are no committee solutions of this system).

The MCLE problem is interesting for at least two reasons. On the one hand, as discussed in the next section, its application to pattern recognition learning is obvious. The approach to pattern recognition learning, which is connected with minimization of class capacity (VC-dimension) of linear (affine) committee decision rules, results in the MCLE problem.

On the other hand, unlike the MCFS problem studied above, the MCLE problem cannot be studied using the conventional approach based on reducing the problem to the equivalent problem of integer optimization (2) and studying the latter's properties. It is not difficult to see that the reducibility of the MCLE problem to (2) (or (3)) is not polynomial. Indeed, to pass to problem (2), we need to find all maximal consistent subsystems of system (6). However, the problem of searching for all maximal consistent subsystems of a system of linear inequalities is known to be intractable. We consider the following problem of combinatorial optimization.

The Maximal Consistent Subsystems Problem (Densest Hemisphere):

Let numbers $n > 1$ and m and vectors

$$a_1, a_2, \dots, a_m \in \mathbb{Q}^n$$

be given. Find the maximal consistent subsystem of system (6) with maximal cardinality.

The following theorem holds.

Theorem 8 ([11]). The densest hemisphere problem is *NP*-hard.

Thus, the conventional scheme of studying the computational complexity of the MCLE problem is not efficient in the general case. Below, we show the MCLE problem to be *NP*-hard and describe an approximate polynomial algorithm for solving it on only the basis of the geometric properties of the family of half-spaces in the finite dimensional vector space. Note that the conventional approach based on analyzing problems (2)–(3) can still be applied when studying the following problem of combinatorial optimization, which is adjacent to the MCLE problem.

The Optimal Committee Improvement Problem (COMIMP):

Let there be given natural numbers $n > 1$, m, and q and vectors

$$a_1, a_2, \dots, a_m, x^1, x^2, \dots, x^q \in \mathbb{Q}^n$$

such that the sequence $Q = (x^1, x^2, \dots, x^q)$ is the committee of system (6). Find the committee $Q' = (y^1, y^2, \dots, y^{q'})$ with minimum possible $q' \leq q$, in which

$$y^i \in \{x^1, x^2, \dots, x^q\} \quad (i \in \mathbb{N}_{q'}).$$

Indeed, we use a line of reasoning similar to that in Section 1. We consider $m \times q$ incidence matrices A' and B', whose elements obey the rule

$$a'_{ji} = 1, \quad b'_{ji} = 1, \text{ if } (a_j, x^i) > 0,$$
$$a'_{ji} = 0, \quad b'_{ji} = -1, \text{ otherwise.}$$

We obtain the matrices A and B by eliminating the pairwise dominated columns of the matrices A' and B'. Let τ be the number of their columns. Consider the problems

$$\min\{(e, t) | Bt \geq f, t \in \mathbb{Z}_+^\tau\} \tag{7}$$

and

$$\min\left\{s: \begin{array}{ll} At \geq sf, & t \in \mathbb{Z}_+^\tau \\ (e, t) \leq 2s - 1, & s \in \mathbb{N} \end{array}\right\}, \tag{8}$$

which are similar to problems (2) and (3), respectively. Taking into account the fact that the constructions in hand can be carried out in the time polynomial in the description length of the COMIMP problem, we obtain the following proposition, which is proved similar to Theorem 3.

Proposition 2, The COMIMP problem and problems (7) and (8) are polynomially equivalent.

3.1. The Computational Complexity of the MCLE Problem

In this subsection, we show that the MCLE problem given above is intractable in the general case. We also describe some of its special cases when the problem is polynomially solvable.

Theorem 9. The MCLE problem is *NP*-hard.

The proof of this theorem follows from two auxiliary propositions with several extra combinatorial problems to be considered.

The Problem of the Committee of Three Elements of a System of Linear Inequalities (3-COMLE):

Let natural numbers m and $n > 1$ and vectors

$$a_1, a_2, \ldots, a_m \in \mathbb{Q}^n$$

be given. Is there a committee solution of system (6) that consists of three elements?

Proposition 3. The 3-COMLE problem is Turing reducible to the MCLE problem.

Proof. Consider an arbitrary particular case of the 3-COMLE problem, viz., the problem of searching for a committee of three elements for fixed system (6). We put it into correspondence with a particular MCLE problem, viz., the minimal committee problem of the same system of inequalities. We apply an arbitrary algorithm of solving the MCLE problem. If there are no committee solutions of the system, the answer to the initial problem is also negative. Let $Q = (x^1, x^2, \ldots, x^q)$ be the minimal committee of system (6); if $q > 3$, then the answer to the initial problem is obviously negative. If $q = 3$, then Q is the required solution to the initial 3-COMLE problem (in this case, the answer to it is positive). We consider the case $q = 1$, where $Q = (x^1)$, separately (as mentioned above, the minimal committee cannot consist of an even number of elements). We find the vector z from the condition

$$(a_j, z) \neq 0 \quad (j \in \mathbb{N}_m).$$

Obviously, one can find such a vector in time polynomial in the description length of the problem. The committee $Q' = (x^1, z, -z)$ is the required solution to 3-COMLE problem. The proposition is proved.

The 3-COMLE problem belongs to the *NP* class, since it takes the time polynomial in the description length of the problem to check whether a fixed sequence $Q = (x^1, x^2, x^3)$ is the committee of system (6).

Consider the auxiliary combinatorial problem.

The Problem of Coloring a Graph with Three Colors (3-COLORABILITY)

Let a finite graph $G = (V, H)$, $V = \{v_1, v_2, \ldots, v_n\}$ be given. Determine whether it can be colored in three colors or, in other words, if there exists the function $\varphi: V \longrightarrow \{1, 2, 3\}$ such that for an arbitrary $u, v \in V$ the condition $(\{u, v\} \in E) \Rightarrow (\varphi(u) \neq \varphi(v))$ is met.

It is known [5] that the 3-COLORABILITY problem is *NP*-complete. In what follows, we show it to be reduced polynomially (according to Karp) to the 3-COMLE problem.

Proposition 4. The 3-COMLE problem is *NP*-complete.

Proof. Let the graph $G = (V, H)$ used to state the 3-COLORABILITY problem be given. Without loss of generality, we assume that $n > 0$ and $V = \mathbb{N}_n$. We put the graph G into correspondence with a system of linear inequalities over $\mathbb{Q}^n$ as follows

$$\begin{cases} x_i + x_j > 0 & (\{i, j\} \in E) \\ \quad x_i < 0 & (i \in V). \end{cases} \tag{9}$$

This can be constructed in the time bounded from above by a polynomial of n. It is not difficult to check that system (9) is consistent if and only if its corresponding graph G is the graph of isolated vertices. We show that the graph is 3-colorable if and only if system (9) has a committee solution consisting of three elements.

Let the partition $V_1 \cup V_2 \cup V_3$ give the coloring of the graph G in three colors. We prove that the sequence $Q = (x^1, x^2, x^3)$, where

$$x_i^k = \begin{cases} 2, & i \in V_k \\ -1, & i \notin V_k \end{cases} \quad (k \in \mathbb{N}_3),$$

is a committee solution of system (9). Let $i \in V_1$ (the cases of V_2 and V_3 can be considered similarly). The sequence Q was chosen so that $x_i^2 < 0$ and $x_i^3 < 0$; thus, Q is the committee solution of the subsystem

$$x_i < 0 \quad (i \in V).$$

Now, consider an arbitrary edge $\{i, j\} \in E$. Without loss of generality, we can assume that $i \in V_1$ and $j \in V_2$. Hence, the choice of Q results in

$$x_i^1 + x_j^1 = 1 > 0 \text{ and } x_i^2 + x_j^2 = 1 > 0.$$

Thus, we showed Q to be the committee solution of the subsystem

$$x_i + x_j > 0 \quad (\{i, j\} \in E),$$

and, hence, of system (9) as a whole.

On the other hand, let the sequence $Q = (x^1, x^2, x^3)$ be an arbitrary committee solution of system (9). We give the sets V_1, V_2, and V_3 by the rule

$$V_k = \{i \in V | x_i^p < 0 (p \in \mathbb{N}_3 \backslash \{k\})\} \quad (k \in \mathbb{N}_3). \quad (10)$$

Since Q is the committee of system (9), the equality holds

$$V_1 \cup V_2 \cup V_3 = V.$$

Without loss of generality, we can assume that subsets (10) form a partition of the set V (i.e., are nonempty and pairwise disjoint). Sets (10) define the required coloring of the graph G. Indeed, suppose the contrary, i.e., an edge $\{i, j\} \subset V_1$ exists (the cases for V_2 and V_3 can be considered similarly). By virtue of the construction of the set V_1,

$$x_i^2 < 0, \quad x_j^2 < 0,$$
$$x_i^3 < 0, \quad x_j^3 < 0,$$

hence,

$$x_i^2 + x_j^2 < 0 \text{ and } x_i^3 + x_j^3 < 0.$$

On the other hand, by the definition of a committee, at least one of the inequalities

$$x_i^2 + x_j^2 > 0 \text{ or } x_i^3 + x_j^3 > 0.$$

holds. The obtained contradiction proves that the coloring is correct. The proposition and, hence, Theorem 9 are proved.

Remark 4. It follows from the proof of the proposition that the MCLE (3-COMLE) problem remains *NP*-hard (*NP*-complete) if we restrict ourselves to considering systems of homogeneous inequalities with all coefficients belonging to the set $\{-1, 0, 1\}$, each inequality having at most three nonzero coefficients.

Remark 5. The proposition holds if n can take arbitrarily large values. If, additionally, the value n is bounded from above, then it may turn out that the 3-COMLE and MCLE problems are polynomially solvable. For instance, it is known (see [3]) that there is a polynomial algorithm for the MCLE problem when $n = 2$.

In [12], we describe an approximate algorithm for solving the MCLE problem, whose properties are given in the following theorem.

Theorem 10 ([12]). Let $m = 2k + n - 1$ for a natural k in system (6) and each subsystem of n inequalities be consistent.

(1) The algorithm described above is correct and has no more than

$$\left\lceil \frac{k}{n-1} \right\rceil$$

iterations, the complexity of each being equivalent to the complexity of the problem of searching for a solution of the consistent subsystem of the initial system.

(2) Let the cardinality of the maximal consistent subsystem of system (6) be no more than $k + (n - 1) + t$ for a natural t. Then, the approximation accuracy r of the algorithm satisfies the relation

$$1 \le r \le \frac{2\left\lceil \frac{k}{n-1} \right\rceil + 1}{2\left\lceil \frac{k-t}{2t+n-1} \right\rceil + 1} \approx 1 + \frac{2t}{n-1}.$$

In the same study, we prove that the algorithm is accurate in the class of uniformly distributed (according to Gale) systems of inequalities, which implies the polynomial solvability of the MCLE problem in this particular case.

We show that several problems adjacent to the MCLE problem are intractable.

The Problem of the Committee of a System of Affine Inequalities (COMLE):

Let there be given natural numbers $n > 1$ and m, vectors

$$a_1, a_2, \ldots, a_m \in \mathbb{Q}^n,$$

and an odd number k. Is there a committee of system (6) with no more than k elements?

Note that the COMLE problem is the MCLE problem restated as a problem of property recognition. The following corollary holds.

Corollary 2. The COMLE problem is polynomially solvable for $k \ge m$ or $k \le 2$ and *NP*-complete otherwise.

Proof. By Mazurov's theorem [3], system (6) has a committee solution whose cardinality is odd and does not exceed M if and only if each of its subsystems of two inequalities is consistent. This condition can be checked, as well as the committee be constructed, in the time bounded from above by a polynomial in m and n.

Checking whether a committee solution exists consisting of at most two elements is obviously equivalent to the problem of searching for a solution of system (6), which is known to be polynomially solvable. The intractability of the problem for $k < m$ follows from the NP-hardness of the MCLE problem.

Consider the problem of searching for solutions of the system

$$\begin{cases} (a_j, y^i) + \xi_i^j > 0 \\ \xi_1^j + \xi_2^j + \ldots + \xi_{2s+1}^j \le s \quad (i \in \mathbb{N}_{2s+1}, j \in \mathbb{N}_m). \\ \xi_i^j \in \{0, 1\}, \quad y^i \in \mathbb{Q}^n \end{cases} \quad (11)$$

Theorem 11. For $k = 2s + 1$ and $|a_j| < 1$, the COMLE problem and problem (11) are polynomially equivalent.

Proof. Since the hypothesis of the appropriate particular problems for COMLE and (11) are determined by the vectors $a_1, a_2, \ldots, a_m$ and number k, one can, without loss of generality, suppose that both problems have the same description of initial data. Hence, to prove the theorem, it is sufficient to show that both problems have the positive answer for the same initial data.

Consider a pair of particular problems of COMLE and (11). Let the answer to the COMLE problem be positive and a sequence $(x^1, x^2, \ldots, x^q)$ be the committee of system (6) for $q \le k = 2s + 1$. Without loss of generality, we assume that $|x^i| \le 1$ for every $j \in \mathbb{N}_q$. By the committee definition, the inequality

$$\left|\{i \in \mathbb{N}_q : (a_j, x^i) > 0\}\right| \ge s + 1$$

holds for all $j \in \mathbb{N}_m$. Moreover, the inequality $(a_j, x^i) > -1$ holds for every i and j by assumption. Hence, the sequence $(y^1, y^2, \ldots, y^q, \xi^1, \ldots, \xi^m)$, where $y^i = x^i$, $\xi^j \in \{0, 1\}^q$, and

$$\xi_i^j = \begin{cases} 0, & \text{if } (a_j, x^i) > 0 \\ 1, & \text{otherwise} \end{cases}$$

is the solution of system (11).

On the other hand, let the sequence $(y^1, y^2, \ldots, y^q, \xi^1, \ldots, \xi^m)$ be a solution of system (11). Substituting it into the system, we verify that the inequality

$$\left|\{y^i : (a_j, y^i) > 0\}\right| \ge s + 1$$

holds for every j. Hence, the sequence $(y^1, \ldots, y^q)$ is a committee of system (6), i.e., a solution to the COMLE problem. Thus, the sets of solutions of system (11) and the COMLE problem are connected by the natural bijection

$$(y^1, y^2, \ldots, y^q, \xi^1, \ldots, \xi^m) \longrightarrow (y^1, \ldots, y^q),$$

which can be computed in the time polynomial in the description of initial data. The theorem is proved.

Corollary 3. The problem of searching for a solution of system (11) is polynomially solvable for $2s + 1 \ge m$ or $s = 0$ and NP-complete otherwise.

Remark 6. Similarly to Remark 5, note that problem (11) is polynomially solvable for $n = 2$ and arbitrary $s \in \mathbb{Z}_+$. Whether the problem is polynomially solvable for fixed $n \ge 3$ remains to be studied.

4. THE PROBLEM OF MINIMAL AFFINE SEPARATION COMMITTEE

In this section, we consider the problem of minimal affine separation committee (MASC), which is adjacent to the MCLE problem of the minimal committee of a system of linear inequalities discussed above.

We fix sets A and B in $\mathbb{R}^n$. As usual [2], by *an affine separation committee* for the sets A and B we mean a finite sequence of functions $(f^1, f^2, \ldots, f^q)$ such that

$$f_i(x) \equiv (\beta^i, x) + \gamma^i \quad (i \in \mathbb{N}_q).$$

For the sets with finite cardinality, the following existence criteria of the dividing committee are known.

Theorem 12 ([14]). An affine separation committee for finite sets $A, B \subset \mathbb{R}^n$ exists if and only if $A \cap B = \emptyset$. The minimal affine committee consists of at most $|A \cup B|$ elements.

We put the system of inequalities

$$\begin{cases} (\beta, a) + \gamma > 0 \quad (a \in A) \\ (\beta, b) + \gamma < 0 \quad (b \in B) \end{cases} \quad (12)$$

into correspondence with the problem of searching for the affine separation committee. We denote the rank of system (12) by r. The following sufficient existence conditions of the affine separation committee directly follow from the sufficient existence conditions of the committee solution to the system of linear inequalities [2, 6].

Proposition 5. Let every subsystem of system (12) of rank k, $0 < k < r$, have a committee solution consisting of at most q elements. Then, there exists an affine separation committee for the sets A and B with the number of elements bounded from above by

$$2q\left\lceil \frac{\lfloor (m-1)/2 \rfloor}{k} \right\rceil + 1.$$

Proposition 6. Let every subsystem of system (12) consisting of $k + 1$ inequalities, where $0 < k < r$, be consistent. Then, there exists an affine separation committee for the sets A and B consisting of at most

$$2\left\lceil \frac{\lfloor (m-k)/2 \rfloor}{k} \right\rceil + 1$$

elements.

Proposition 7. Let, in addition to the hypothesis of the previous proposition, a subsystem of system (12) of cardinality μ with a committee solution consisting of $2q - 1$ elements can be found. Then, there exists an affine committee dividing the sets A and B and consisting of at most

$$2q\left(1 + \left\lceil \frac{m-\mu}{k} \right\rceil\right) - 1$$

elements.

In what follows, we assume that $A, B \subset \mathbb{Q}^n$ are finite.

The Problem of the Minimal Affine Separation Committee (MASC):

Let a natural number $n > 1$ and sets A, $B \subset \mathbb{Q}^n$, where

$$A = \{a_1, a_2, \ldots, a_{m_1}\}, \quad B = \{b_1, b_2, \ldots, b_{m_2}\},$$

be given. Find an affine separation committee of the sets A and B with the minimal number of elements (or prove that the sets cannot be divided by a linear committee).

The Problem of Affine Separation Committee of Three Elements (3-ASC):

Let there be given a natural number $n > 1$ and sets A, $B \subset \mathbb{Q}^n$, where

$$A = \{a_1, a_2, \ldots, a_{m_1}\}, \quad B = \{b_1, b_2, \ldots, b_{m_2}\}.$$

Is there an affine separation committee for the sets A and B that consists of three elements?

It is not difficult to prove the following proposition. Here, as before, we assume that two problems are polynomially equivalent if they are mutually Turing reducible to each other.

Proposition 8. (1) The MCLE and MASC problems are polynomially equivalent. (2) The 3-COMLE and 3-ASC problems are also polynomially equivalent.

Taking into account Proposition 4 and Theorem 9 proved above, we state the following corollary.

Corollary 4. (1) The MASC problem is *NP*-hard. (2) The 3-ASC problem is *NP*-complete.

Remark 7. Similarly to Remark 4, note that the MASC (3-ASC) problem remains *NP*-hard (*NP*-complete) even under the condition $A \cup B \subset \{-1, 0, 1\}^n$, and every point contains at most three nonzero coordinates.

It is not difficult to check that the approximate algorithm mentioned in the previous section and applied to search for the committee solution of system (12) is also an approximate algorithm for the MASC problem. The following theorem holds.

Theorem 13. Let $n > 2$, $|A \cup B| = 2k + n$ and every subsystem of system (12) consisting of $n + 1$ inequalities be consistent. The algorithm finds the affine separation committee for the sets A and B in no more than $\left\lceil \frac{k}{n} \right\rceil$ iterations. If, moreover, the cardinality of the maximal consistent subsystem of system (12) does not exceed $k + t + n$ for some natural number t, then the approximation accuracy r of the algorithm satisfies the relation

$$1 \le r \le \frac{2\left\lceil \frac{k}{n} \right\rceil + 1}{2\left\lceil \frac{k-t}{2t+n} \right\rceil + 1} \approx 1 + \frac{2t}{n}.$$

CONCLUSIONS

In this work, we studied the computational complexity of two special cases of the minimal committee problem—the MCFS and MCLE problems. Both problems are shown to be intractable. As a corollary, we obtained some results on the computational complexity of the problem of the minimal affine separation committee (MASC).

ACKNOWLEDGMENTS

This study was financially supported by the Russian Foundation for Basic Research (project nos. NSh 792.2003.1, 04.01.96104, 04.01.00108) and Grant of the President of the Russian Federation no. MD-6768.2006.1.

REFERENCES

1. I. I. Eremin, *Theory of Linear Optimization* (BMV, 2002).
2. V. D. Mazurov, M. Yu. Khachai, and A. I. Rybin "Committee Constructions for Solving Problems of Selection, Diagnostics, and Prediction," in *Proc. of the Steklov Institute of Mathematics*, Suppl. 1, 67–101 (2002).
3. V. D. Mazurov, *Committee Method in Optimization and Classification Problems* (Nauka, Moscow, 1990) [in Russian].
4. I. I. Yeremin and V. D. Mazurov, *Nonstationary Processes of Mathematical Programming* (Nauka, Moscow, 1979) [in Russian].
5. M. Garey and D. S. Johnson, *Computer and Intractability: A Guide to the Theory of NP-completeness* (W.H. Freeman, San Francisco, 1979).
6. M. Yu. Khachai and A. I. Rybin, "On the Committee Solution with Minimal Number of Terms of System of Linear Inequalities," in *Proc. of XI Int. Baikal School–Seminar on Optimization Methods and Their Applications* (ISE of SD RAS, Irkutsk, 1998), pp. 26–40.
7. C. Lund and M. Yannakakis, "On the Hardness of Approximationg Minimization Problems," in *Proc. of the 33rd IEEE Symposium on Foundations of Computer Science* (1992), pp. 960–981.
8. U. Feige, "A Threshold of ln n for Approximating Set Cover," J. of the ACM **45** (4) (1998).
9. D. S. Johnson, "Approximation Algorithms for Combinatorial Problems," J. Computer and System Sci. **9** (3), 256–278 (1974).
10. M. Yu. Khachay, "On Computational Complexity of the Minimal Committee of Finite Sets Problem," in *Proc. of*

the 2nd Int. Workshop on Discrete Optimization Methods in Production and Logistics, Omsk–Irkutsk (2004), pp. 176–179.

11. D. S. Johnson and F. P. Preparata, "The Densest Hemisphere Problem," Theoretical Computer Sci. **6**, 93–107 (1978).
12. I. Dinur, O. Regev, and C. Smyth, "The Hardness of 3-Uniform Hypergraph Coloring," in *Proc. of the 43rd Annual IEEE Symposium on Foundations of Computer Science*, November 2002.
13. M. Yu. Khachay, "On Approximate Algorithm of a Minimal Committee of a Linear Inequalities System," Pattern Recognition and Image Analysis **13** (3), 459–464, (2003).
14. V. D. Mazurov, "Committees of Systems of Inequalities and Recognition Problem," Kibernetika no. 3, 140–146, (1971).

Mikhail Yur'evich Khachai. Born 1970 in Krasnotur'insk, Sverdlovsk region. Graduated from the Faculty of Mathematics and Mechanics of Ural State University in Yekaterinburg in 1993 (Mathematics). Received his candidate's degree (Mathematical Cybernetics) in 1996. Since 1994, he has been at the Institute of Mathematics and Mechanics of the Ural Division of the Russian Academy of Sciences. Since 1997, he has headed the Department of Pattern Recognition at the same institute. Scientific interests: theory and methods of pattern recognition learning, committee (perceptron) decision rules, and theory and methods of improper optimization problems. Author of more than 25 publications, including five in PRIA.