ISSN 1064–5624, Doklady Mathematics, 2006, Vol. 73, No. 1, pp. 138–141. © Pleiades Publishing, Inc., 2006.
Original Russian Text © M.Yu. Khachai, 2006, published in Doklady Akademii Nauk, 2006, Vol. 406, No. 6, pp. 742–745.

COMPUTER SCIENCE

Computational Complexity of the Minimum Committee Problem and Related Problems

M. Yu. Khachai

Presented by Academician I.I. Eremin June 21, 2005

Received July 19, 2005

DOI: 10.1134/S1064562406010376

Given a set X and a collection $D_1, D_2, \ldots, D_m$ of its nonempty subsets, consider the 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 case $\cap D_j = \emptyset$ is admissible. A finite sequence $Q = (x^1, x^2, \ldots, x^q)$ satisfying the condition

$$\left|\{i\colon\ x^i \in D_j\}\right| > \frac{q}{2}, \quad j \in \mathbb{N}_m$$

is called a committee solution with q elements of system (1) (or a committee) [1].

Minimum committee (MC) problem. Given a set X and subsets $D_1, D_2, \ldots, D_m \neq \emptyset$, find a committee solution to system (1) with the least possible q (or show that the system has no committee solutions).

Following [2], it is convenient to restate the MC problem in terms of integer linear programming. Let $J_1, J_2, \ldots, J_T$ be the index set of all maximal (under inclusion) consistent subsystems (MCSs) of system (1). Obviously, the system is consistent if and only if $T = 1$; otherwise, $1 < T < 2^m$. Define two $m \times T$ incidence matrices A and B according to the rule

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

$$a_{ji} = 0, \quad b_{ji} = -1 \text{ otherwise}$$

also consider the integer linear programs

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

and

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

Here, e and f are all-one vectors from the spaces E_T and E_m, respectively.

The following result is well known.

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

We consider the following two special cases of the MC problem.

1. The MC problem in which X is finite (referred to as the MCFS problem). We show that Theorem 1 can be refined in this case. It is also shown that the MCFS problem is *NP*-hard, which implies that the MC problem is generally *NP*-hard as well. Moreover, the threshold of effective approximability for the MCFS problem is estimated.

2. The MC problem in which $X = \mathbb{Q}^n$ for arbitrary $n > 1$ and the subsets D_j are open half-spaces. This problem is referred to as the minimum committee problem for a system of linear inequalities (abbreviated as MCLE). It is shown that the problem is also *NP*-hard, and some of its polynomially solvable subclasses are presented.

Minimum committee of finite sets (MCFS). Given a set $X = \{x^1, x^2, \ldots, x^p\}$ and a collection of its subsets $D_1, D_2, \ldots, D_m$, find the committee of system (1) with a minimum number of elements (or show that the system has no committee solutions).

Theorem 2 [4]. *The MCFS problem is NP-hard.*

Remark 1. The MCFS problem remains NP-hard if $|D_j| \leq 3$ for all D_j, except for possibly one.

Remark 2. *The MC problem, which is an extension of the MCFS problem, is NP-hard.*

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

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

Theorem 2 states that the MCFS problem is *NP*-hard. Below, we give an estimate for the threshold

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

of its effective approximability (the proofs can be found in [5]). These results are related to those derived for the set cover problem in [6, 7].

Lemma. *The existence of an approximation algorithm with an performance guarantee of r for the MCFS problem implies the existence of a similar algorithm with the same accuracy for SET COVER.*

Theorem 4. *If* $P \neq NP$, *then there is no approximation algorithm with the performance guarantee* $\frac{1}{4}\log_2(m-1)$ *for the MCFS problem.*

Theorem 5. *If* $NP \not\subseteq \mathrm{TIME}(n^{O(\log_2\log_2 n)})$, *then, for every* $\varepsilon > 0$, *there is no approximation algorithm with the performance guarantee* $(1-\varepsilon)\ln(m-1)$ *for the MCFS problem.*

Note that the technique used to reduce the set cover problem to the minimum committee problem in the proof of the lemma can also be used to solve the problem of committee separation of sets. Consider, for example, the following setting. Given subsets A and B of X and the class of decision rules $\mathcal{F} = \{f(x, \alpha) | \alpha \in \Lambda\} \subseteq \{X \to \{0, 1\}\}$, construct a committee decision rule (over $\mathcal{F}$) that correctly separates A and B. In other words, it is necessary to find a sequence $(\alpha^1, \alpha^2, \ldots, \alpha^q)$ such that

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

Proposition 1. *Let the parameters* $\alpha^1, \alpha^2, \ldots, \alpha^k$ *be fixed so that*

$$f(a, \alpha^0) = 1 \ (a \in A),$$

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

and, for every $a \in A$, *there be a number* $i = i(a) \in \mathbb{N}_k$ *such that* $f(a, \alpha^{i(a)}) = 1$.

Then the sequence $(\underbrace{\alpha^0, \ldots, \alpha^0}_{k-1}, \alpha^1, \ldots, \alpha^k)$ *defines a committee decision rule that separates A from B without errors.*

In what follows, let $X = \mathbb{Q}^n$ for $n > 1$, and let $D_j = \{x \in X | (a_j, x) > 0, a_j \neq 0\}$. System (1) then becomes

$$(a_j, x) > 0 \ (j \in \mathbb{N}_m). \qquad (4)$$

Minimum committee of a system of linear inequalities (MCLE). Given positive integers m and $n > 1$ and vectors $a_1, a_2, \ldots, a_m \in \mathbb{Q}^n$, find a committee solution (committee) of system (4) with the smallest number of elements (or show that the system has no committee solutions).

The MCLE problem is of interest for at least two reasons. On the one hand, it has an obvious application to pattern recognition learning. This problem arises in the capacity minimization method (VC-dimension) in the class of linear (affine) committee decision rules. On the other hand, the traditional approach (effective for the MCFS problem), in which the MCSs of system (4) are enumerated and the problem is reduced to equivalent problem (2) (or (3)), is not rational in this case. Indeed, consider the following combinatorial optimization problem.

Largest MCS problem (DENSEST HEMISPHERE). Given scalars $n > 1$ and m and vectors $a_1, a_2, \ldots, a_m \in \mathbb{Q}^n$, find a maximum-cardinality MCS of system (4).

Theorem 6 [8]. *DENSEST HEMISPHERE is an NP-hard problem.*

Note that the traditional approach based on analysis of problems (2) and (3) is not effective for MCLE but can still be used to study the following combinatorial optimization problem, which is similar to MCLE.

Optimal committee improvement (COMIMP). Given positive integers $n > 1$, m, and q and vectors $a_1, a_2, \ldots, a_m, x^1, x^2, \ldots, x^q \in \mathbb{Q}^n$ such that the sequence $Q = (x^1, x^2, \ldots, x^q)$ is a committee of system (4), find a committee $Q' = (y^1, y^2, \ldots, y^{q'})$ with the least possible $q' \leq q$ such that

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

Indeed, let A and B be the $m \times q$ matrices defined according to the rule

$$a_{ji} = 1, \quad b_{ji} = 1 \text{ if } x^i \text{ satisfies the } j\text{th inequality,}$$
$$a_{ji} = 0, \quad b_{ji} = -1 \text{ otherwise.}$$

Without loss of generality, we assume that the columns of A and B are pairwise nondominated. Consider problems (2) and (3) they define. It is easy to see that these constructions can be performed in polynomial time with respect to the input length of the COMIMP problem.

Returning to the MCLE problem, we formulate the basic theorem.

Theorem 7. *The MCLE problem is NP-hard.*

The proof of the theorem follows from two auxiliary statements, which are preceded by a few additional combinatorial problems.

Three-element committee of a linear inequality system (3-COMLE). Given positive integers m and $n > 1$ and vectors $a_1, a_2, \ldots, a_m \in \mathbb{Q}^n$, does system (4) have a three-element committee solution?

Proposition 2. *The 3-COMLE problem is reducible in the sense of Turing to the MCLE problem.*

Obviously, 3-COMLE is an NP-problem, because whether a fixed sequence $Q = (x^1, x^2, x^3)$ is a committee of system (4) can be checked in polynomial time with respect to the input length of the latter.

Three-colorability of a graph (GRAPH 3-COLORABILITY). Given a graph $G = (V, E)$, determine whether it can be colored in three colors; i.e., whether there is a function φ: $V \to \{1, 2, 3\}$ such that $\{u, v\} \in E$ implies $\varphi(u) \neq \varphi(v)$.

It is well known [9] that the GRAPH 3-COLORABILITY problem is *NP*-complete.

Proposition 3. *The* GRAPH 3-COLORABILITY *problem is reducible to the* 3-*COMLE problem.*

Remark 3. Proposition 3 implies that the 3-COMLE problem is *NP*-complete. In fact, it has been shown that the MCLE (3-COMLE) problem remains *NP*-hard (*NP*-complete) for systems (4) with $a_j \in \{-1, 0, 1\}^n$ and $(a_j, a_j) \leq 2$.

Remark 4. Proposition 3 holds for arbitrarily large values of n. Under an additional restriction on n from above, the 3-COMLE and MCLE problems can become polynomially solvable. For $n = 2$, a polynomial-time algorithm is known to exist for MCLE [2].

In [10], an approximation algorithm for the MCLE problem was presented, whose properties are described by the following statement.

Theorem 8. *In system* (4), *let* $m = 2k + n - 1$ *for some positive integer k and each subsystem of n inequalities be consistent.*

(i) *The algorithm is well defined and consists of no more than* $\left\lceil \frac{k}{n-1} \right\rceil$ *iterations. The complexity of each iteration is equivalent to the complexity of solving a consistent subsystem of the original system.*

(ii) *Suppose that the cardinality of the largest consistent subsystem of system* (4) *does not exceed* $k + t + n - 1$ *for some positive integer t. Then the performance guarantee r of the algorithm satisfies*

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

It was also shown in [10] that the algorithm is accurate in the class of uniformly distributed (in the sense of Gale) inequality systems, which implies that the MCLE problem is polynomially solvable in this special case.

The MCLE problem can be restated as the problem of property recognition.

Committee of a system of linear inequalities (COMLE). Given positive integers $n > 1$ and m, vectors $a_1, a_2, \ldots, a_m \in \mathbb{Q}^n$, and an odd scalar k, is there a committee of system (4) with the number of elements not exceeding k?

Corollary 2 (to Theorem 7). *The COMLE problem is polynomially solvable if* $k \geq m$ *or* $k \leq 2$ *and is NP-complete otherwise.*

Consider the problem of a minimum affine separating committee, which is similar to the MCLE problem. An affine separating committee for finite sets $A, B \subset \mathbb{R}^n$ [1] is defined as a committee decision rule over the class of affine functions $\mathcal{F} = \{f(x, \beta, \gamma) = (\beta, x) + \gamma | \beta \in \mathbb{R}^n, \gamma \in \mathbb{R}\}$ that correctly separates these sets. The following necessary and sufficient condition for the existence of a separating committee is available.

Theorem 9 [11]. *An affine separating committee for finite sets A,* $B \subset \mathbb{R}^n$ *exists if and only if* $A \cap B = \emptyset$. *A minimum affine committee contains no more than* $|A \cup B|$ *elements.*

The problem of searching for an affine separating committee is associated with the system of inequalities

$$\begin{aligned} &(\beta, a) + \gamma > 0, \quad a \in A, \\ &(\beta, b) + \gamma < 0, \quad b \in B. \end{aligned} \tag{5}$$

Let r denote the rank of system (5). Below are several sufficient conditions for the existence of a separating committee.

Proposition 4. *Suppose that each k-rank subsystem of system* (5), *where* $0 < k < r$, *has a committee solution consisting of no more than q elements.*

Then there is an affine separating committee for A and B such that the number of its elements is bounded above by $2q\left\lceil \frac{\lfloor (m-1)/2 \rfloor}{k} \right\rceil + 1$.

Proposition 5. *Let each* $(k + 1)$-*inequality subsystem of system* (5), *where* $0 < k < r$, *be consistent.*

Then there is a separating affine committee for A and B that consists of no more than $2\left\lceil \frac{\lfloor (m-k)/2 \rfloor}{k} \right\rceil + 1$ *elements.*

Proposition 6. *Suppose that the assumption of Proposition* 5 *be satisfied and there exists a subsystem of system* (5) *with cardinality* μ *that has a committee solution consisting of* $2q - 1$ *elements.*

Then there is a separating affine committee for A and B that consists of no more than $2q\left(1 + \left\lceil \frac{m-\mu}{k} \right\rceil\right) - 1$ *elements.*

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

Minimum affine separating committee (MADC). Given a positive integer $n > 1$ and sets $A, B \subset \mathbb{Q}^n$, where $A = \{a_1, a_2, \ldots, a_{m_1}\}$ and $B = \{b_1, b_2, \ldots, b_{m_2}\}$, find an affine separating committee for A and B with the smallest number of elements (or show that the sets cannot be separated by a committee).

Three-element affine separating committee (3-ADC). Given a positive integer $n > 1$ and sets $A, B \subset \mathbb{Q}^n$, is there a three-element affine separating committee for A and B?

The result given below can be proved by the method used in the proof of Theorem 7.

Proposition 7. *The MADC problem is NP-hard. The* 3-*ADC problem is NP-complete.*

Remark 5. By analogy with Remark 3, we note that the MADC (3-ADC) problem remains *NP*-hard (*NP*-complete) even if $A \cup B \subset \{z \in \{-1, 0, 1\}^n | (z, z) \leq 2\}$.

It is easy to see that the approximation algorithm used above to search for a committee solution to system (5) is also an approximation algorithm for the MADC problem; moreover, the following result holds.

Theorem 10. *Let* $n \geq 1$, $|A \cup B| = 2k + n$, *and each* $(n + 1)$-*inequality subsystem of system* (5) *be consistent. Then the algorithm finds a separating affine committee for A and B in no more than* $\left\lceil \frac{k}{n} \right\rceil$ *iterations. If, additionally, the cardinality of the largest MCS of system* (5) *does not exceed* $k + t + n$ *for some positive integer t, then the performance guarantee r of the algorithm satisfies*

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

ACKNOWLEDGMENTS

This work was supported by the Russian Foundation for Basic Research (project nos. 04-01-96104, 04-01-00108), the program "Leading Scientific Schools" (project no. NSh-792.2003.1), and the Russian Science Support Foundation.

REFERENCES

1. M. Yu. Khachai, Vl. D. Mazurov, and A. I. Rybin, Proc. Steklov Inst. Math. Suppl., No. 1, S67–S101 (2002).
2. Vl. D. Mazurov, *Committee Method in Optimization and Classification Problems* (Nauka, Moscow, 1990) [in Russian].
3. I. I. Eremin and Vl. D. Mazurov, *Nonstationary Processes in Mathematical Programming* (Nauka, Moscow 1979) [in Russian].
4. M. Yu. Khachai and A. I. Rybin, in *Proceedings of XI International Baikal School–Seminar on Optimization Methods and Applications* (ISE SO RAN, Irkutsk, 1998), pp. 26–40.
5. M. Yu. Khachay, in *Proceedings of 2nd International Workshop 'Discrete Optimization Methods in Production and Logistics'* (Omsk, 2004), pp. 176–179.
6. C. Lund and M. Yannakakis, in *Proceedings of XXXIII IEEE Symposium on Foundations of Computer Science* (IEEE Comput. Soc., New York, 1992), pp. 960–981.
7. U. Feige, "A Threshold of ln$n$ for Approximating Set Cover," J. ACM **45** (4) (1998).
8. D. S. Johnson and F. P. Preparata, Theor. Comput. Sci., No. 6, 93–107 (1978).
9. M. R. Garey and D. Johnson, *Computers and Intractability* (Freeman, San Francisco, 1979; Mir, Moscow, 1982).
10. M. Yu. Khachay, Pattern Recogn. Image Anal. **13**, 459–464 (2003).
11. Vl. D. Mazurov, Kibernetika, No. 3, 140–146 (1971).