INFORMATICA, 2009, Vol. 20, No. 2, 217–234

Complexity and Approximability of Committee Polyhedral Separability of Sets in General Position

Michael KHACHAY, Maria POBERII
Institute of Mathematics and Mechanics, Ural Branch of Russian Academy of Sciences
S.Kovalevskoy 16, 620219 Ekaterinburg, Russia
e-mail: mkhachay@imm.uran.ru

Received: August 2008; accepted: May 2009

Abstract. It is known that the minimum affine separating committee (MASC) combinatorial optimization problem, which is related to some machine learning techniques, is NP-hard and does not belong to Apx class unless $P = NP$. In this paper, it is shown that the MASC problem formulated in a fixed dimension space within $n > 1$ is intractable even if sets defining an instance of the problem are in general position. A new polynomial-time approximation algorithm for this modification of the MASC problem is presented. An approximation ratio and complexity bounds of the algorithm are obtained.

Keywords: polyhedral separability, computational complexity, approximation algorithms, committees.

1. Introduction

The minimum affine separating committee (MASC) problem (Khachai, 2006a, 2006b) considered in this paper arose in machine learning at the training stage in the class of committee piece-wise linear decision rules. Actually, the MASC problem is closely connected with two well-known NP-hard problems: the training problem for a simplest classical perceptron (Lin, 1991) and the problem of polyhedral separability (Megiddo, 1988). According to the traditional approach to the analysis of subclasses of intractable problems, it seems important to study computational complexity and approximability of the MASC problem, as a particular case of the mentioned-above problems.

It is known (Khachai, 2006a) that the MASC problem is NP-hard and is hardly approximable (Khachai, 2008) in the general case. Also, it is known (Khachai, 2008) that the problem remains intractable being formulated in a space of an arbitrary fixed dimension $n > 1$. Nevertheless, the proofs of all these intractability results are significantly based on considering degenerate (specifically constructed) instances of the problem. A natural question arises whether the MASC problem remains intractable if these instances are explicitly excluded from consideration. For this exclusion, it is sufficient to assume that a finite set (from an n-dimensional space) defining the instance of the MASC problem is in general position, i.e., each subset of this set containing $n + 1$ elements is affinely independent (in this case, these elements are vertices of an n-dimensional nondegenerate simplex).

Up to now, the question on intractability of the MASC problem under this additional constraint was still open. In addition, the question on shortening the gap between the known (Khachai, 2008) approximation threshold $\mathrm{O}(\log\log\log m)$ of the problem and the best approximation ratio of $\mathrm{O}(m)$ of known polynomial-time approximation algorithms (Khachay, 2006b), where m is the cardinality of a finite set defining the instance of the MASC problem, was still open as well. Answers to these questions are given in the present paper. Actually, the paper contains the following results.

1. It is proved that the MASC problem formulated in a space of an arbitrary fixed dimension $n > 1$ remains NP-hard under an additional constraint on general position (Section 3, this result was previously announced at Khachay and Pobery (2008).
2. A new polynomial-time approximation algorithm having the approximation ratio of $\mathrm{O}(\log m)$ under some natural assumption is proposed (Section 4).

2. Definitions, Problems, and Known Results

Let $\mathbb{R}$, $\mathbb{Q}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the sets of real, rational, integer, and natural numbers, respectively. Let $\mathbb{R}^n$, $\mathbb{Q}^n$, and $\mathbb{Z}^n$ denote the corresponding n-dimensional vector spaces, and $\mathbb{N}_m = \{1, \ldots, m\}$. A function $f\colon \mathbb{R}^n \to \mathbb{R}$ having the form $f(x) = c^T x - d$, where $c \in \mathbb{Q}^n$, $d \in \mathbb{Q}$, is called *an affine function* (with rational coefficients).

DEFINITION 1. Let $f_1, \ldots, f_q$ be affine functions, and A, B be finite subsets of $\mathbb{R}^n$. A finite sequence

$$Q = (f_1, \ldots, f_q)$$

is called *an affine committee separating the sets A and B* if

$$|\{i \in \mathbb{N}_q\colon f_i(a) > 0\}| > \frac{q}{2} \quad (a \in A),$$
$$|\{i \in \mathbb{N}_q\colon f_i(b) < 0\}| > \frac{q}{2} \quad (b \in B).$$

Here the number q is called the number of elements of the committee Q, and the sets A and B are called *separable* by this committee.

According to the Mazurov criterion (Mazurov, 1971), sets A and B can be separated by an affine committee if and only if $A \cap B = \emptyset$. However, for many reasons (model simplification, VCD minimization, etc.), of particular interest are separating committees with the minimum number of elements, which are called *minimum committees*.

PROBLEM 1 (Minimum affine separating committee (MASC)). Let finite sets $A, B \subset \mathbb{Q}^n$ be given. It is required to find an affine committee Q with the minimum number of elements q that separates the sets A and B.

The MASC problem is related to some combinatorial optimization problems originated from machine learning and computational geometry. Let us discuss them briefly. A simplest perceptron is a 2-layered feed-forward neural network without hidden layers with q input neurons and a single output (with number $q + 1$). An activation function of the ith neuron has the classical form

$$\varphi_i(x) = \begin{cases} 1, & \text{if } c_i^T x - d_i > 0, \\ -1, & \text{otherwise}, \end{cases} \tag{1}$$

where $x \in \mathbb{R}^n$, $c_i \in \mathbb{Q}^n$, and d_i is a rational number. Thus, the perceptron realizes the decision rule

$$F(x|\,(c_1, d_1), \ldots, (c_{q+1}, d_{q+1}))\colon\ \mathbb{R}^n \to \{-1, 1\},$$

parameterized by the pairs (c_i, d_i). This rule assigns a pattern x to the first or the second class if $F(x) = 1$ or $F(x) = -1$, correspondingly.

Given a training sample

$$(a_1, \ldots, a_{m_1}, b_1, \ldots, b_{m_2}), \quad (a_i, b_j \in \mathbb{Q}^n), \tag{2}$$

consisting of precedents a_i of the first class and b_j of the second class, one can design a learning procedure for the perceptron, i.e., a procedure of fitting parameters c_i and d_i in order to

$$F(a_i) > 0 \quad (i \in \mathbb{N}_{m_1}), \qquad F(b_j) < 0 \quad (j \in \mathbb{N}_{m_2}). \tag{3}$$

Any "trained" perceptron, which parameters are specified as an arbitrary feasible solution of system (3), is called *correct* on sample (2). The following combinatorial problems are closely connected to this learning procedure.

PROBLEM 2 (Training (loading) a perceptron). Let a natural number q and training sample (2) be given. Does there exist a correct perceptron on sample (2) with at most q input neurons?

PROBLEM 3 (Optimal correct perceptron (OCP)). Let training sample (2) be given. It is required to construct a correct on sample (2) perceptron with the minimum number of input neurons.

It is known (Blum, 1992) that the former problem is NP-complete and remains such for any fixed $q \geqslant 2$, while the latter is NP-hard (Lin, 1991). As can be seen, the MASC problem is a special case of the OCP problem, in which the output neuron has parameters $c_{q+1} = [1, \ldots, 1]^T$, $d_{q+1} = 0$ according to the simple majority voting rule.

Other problems that related to the MASC problem originate from computational geometry and concern with constructing optimal piece-wise linear separating surfaces for sets having intersecting convex hulls. Following (Megiddo, 1988), consider the formulations of these problems. To each hyperplane $H = \{x \in \mathbb{R}^n : c^T x = d, c \neq 0\}$, assign a predicate $\Pi[H]$: $\mathbb{R}^n \to \{\textit{true, false}\}$ by the rule similar to (1):

$$\Pi[H](x) = \begin{cases} \textit{true}, & \text{if } c^T x > d, \\ \textit{false}, & \text{otherwise.} \end{cases}$$

Let finite sets $A, B \subset \mathbb{R}^n$ and Boolean formula $\varphi(\xi_1, \ldots, \xi_k)$ be given. Hyperplanes $H_1, \ldots, H_k$ are said to separate the sets A and B by the rule (formula) φ if

$$\varphi(\Pi[H_1](a), \ldots, \Pi[H_k](a)) = \textit{true} \quad (a \in A),$$
$$\varphi(\Pi[H_1](b), \ldots, \Pi[H_k](b)) = \textit{false} \quad (b \in B).$$

Consider the following combinatorial problems.

PROBLEM 4 (k-Polyhedral separability with a given Boolean formula). Let finite sets $A, B \subset \mathbb{Q}^n$, $A = \{a_1, \ldots, a_{m_1}\}$, $B = \{b_1, \ldots, b_{m_2}\}$ and a Boolean function $\varphi(\xi_1, \ldots, \xi_k)$ be given. Do there exist hyperplanes $H_1, \ldots, H_k$ separating the sets A and B by the rule φ?

When the separation rule φ is not known a priory, one can formulate the following, more general, problem.

PROBLEM 5 (k-Polyhedral separability). Let finite sets $A, B \subset \mathbb{Q}^n$ $A = \{a_1, \ldots, a_{m_1}\}$, $B = \{b_1, \ldots, b_{m_2}\}$ and a natural number k be given. Do there exist hyperplanes $H_1, \ldots, H_k$ separating the sets A and B by the following rule: for each pair (a, b), where $a \in A$ and $b \in B$, there is an appropriate hyperplane H_j, $j = j(a, b)$, such that $\Pi[H_j](a) = true$ and $\Pi[H_j](b) = false$?

As is known, the latter problem has a positive answer if and only if there is an appropriate formula φ for which the former problem has also a positive answer. The following theorem summarizes the complexity results obtained in Megiddo (1988).

Theorem 1.

1. *Both above-formulated problems are NP-complete and remain intractable for any fixed* $k \geqslant 2$.
2. *Problem 5 remains NP-complete for any fixed* $n > 1$.
3. *Problem 5 for arbitrary fixed* k *and* n *is polynomially solvable.*

The MASC problem is a particular case of the optimization version of the problem of k-polyhedral separability with a given Boolean formula (which takes the value *true* if and only if the major part of its arguments take the value *true*).

The following theorem characterizes computational complexity of the MASC problem in the general case.

Theorem 2 (Khachai, 2006a). *The MASC problem is NP-hard and remains such under the additional constraint* $A \cup B \subset \{x \in \{0,1,2\}^n \colon \|x\|_2 \leqslant 2\}$.

In correspondence to the standard approach to the analysis of NP-hard problems, one can be interested in constructing polynomial-time approximation algorithms and finding lower bounds of efficient approximation ratios for these problems. In Khachay (2006b), an approximation algorithm having the ratio of $\mathrm{O}(m)$ is proposed. This ratio is the best known result to date for the MASC problem in the general case. Also, this paper contains the description of a non-trivial polynomially solvable subclass of the MASC problem for which this algorithm is exact. On the other hand, there are some *negative* results concerning approximability of the MASC problem.

Theorem 3 (Khachai, 2008). *The MASC problem does not belong to* Apx *complexity class unless* $P = NP$. *Moreover, if* $NP \not\subset DTIME(2^{poly(\log n)})$ *then there exists a constant* $D > 0$ *such that the MASC problem has no approximation algorithms with an approximation ratio better then* $D \log\log\log m$.

In Khachai (2008), a particular case of the MASC problem formulated in a fixed dimension space is considered. This case is important for applications, especially in machine learning.

PROBLEM 6 (Minimum affine separating committee in a space of fixed dimension n (MASC(n))). Let finite sets $A = \{a_1, \dots, a_{m_1}\}$ and $B = \{b_1, \dots, b_{m_2}\}$, $A, B \subset \mathbb{Q}^n$, be given, where n is fixed. It is required to find an affine committee Q with the minimum number of elements that separates the sets A and B.

It is known (Mazurov, 1971) that the MASC(n) problem is polynomially solvable for $n = 1$. For an arbitrary $n > 1$ this problem is NP-hard. The proof of this result is obtained as a consequence of intractability of the following problem.

PROBLEM 7 (Planar affine separating committee (PASC)). Let finite sets $A = \{a_1, \dots, a_{m_1}\}$ and $B = \{b_1, \dots, b_{m_2}\}$, $A, B \subset \mathbb{Q}^2$, and a natural number t be given. Does there exists an affine committee Q with at most t elements that separates the sets A and B?

It can be easily verified that the PASC problem is just a decision version of the MASC(2) problem and belongs to NP. The proof of intractability of the problem PASC follows from polynomial reduction of the well-known NP-complete problem of covering a finite set (of points) on the plane by a set of straight lines, also known as the point covering (PC) problem.

DEFINITION 2. A set of lines $L = \{l_1, \dots, l_k\}$, where $l_j = \{x \in \mathbb{R}^2 \colon c_j^T x = d_j\}$ and $c_j \neq 0$, is called a *cover* of a set P if for every $p \in P$ there is a line $l = l(p) \in L$ such that $p \in l$.

PROBLEM 8 (Point covering (PC)). Let a finite set P on the plane and a natural number s be given. Does there exist a cover L of the set P such that $|L| \leqslant s$?

If a set P is in general position, i.e., there is no any straight line containing arbitrary three points of P, the PC problem has a trivial solution ("Yes" if and only if $s \geqslant \lceil |P|/2 \rceil$) that can be found in polynomial time. Nevertheless, in the general case the PC problem is intractable.

Theorem 4 (Megiddo, 1982). *The PC problem is NP-complete in the strong sense.*

Hereinafter, the following notation is used:

$$B(x_0, \varepsilon) = \{x \in \mathbb{R}^2 \colon \|x - x_0\|_2 \leqslant \varepsilon\}$$

is the circle centered at point x_0 with radius ε, $\operatorname{aff}(P)$ is the affine hull of the set P, and $\dim$ is the dimension of an affine (or linear) manifold. Also, the following proposition is necessary.

PROPOSITION 1 (Megiddo, 1982). Let a set $P = \{p_1, \ldots, p_k\} \subset \mathbb{Z}^2$, numbers $\rho = \max\{\|p\|_2 \colon p \in P\}$ and $\varepsilon \in (0, \frac{1}{6(2\rho+1)})$, and nonempty subset $J \subset \mathbb{N}_k$ be given. A straight line $l = l(J)$ such that

$$B(p_j, \varepsilon) \cap l \neq \emptyset \quad (j \in J) \tag{4}$$

exists if and only if the condition

$$\dim \operatorname{aff}(\{p_j \colon j \in J\}) \leqslant 1$$

holds.

Further, let an instance of the PC problem be defined by a set $P = \{p_1, \ldots, p_k\} \subset \mathbb{Z}^2$ and a natural number s. Let us determine numbers ρ and ε by the formulas

$$\rho = \max\{\|p\|_2 \colon p \in P\}, \quad \varepsilon = \frac{1}{6(2\rho + 1) + 1}. \tag{5}$$

Fix a vector σ, $\|\sigma\|_2 = 1$ such that, for any $\{i, j\} \subset \mathbb{N}_k$, the line segments $[p_i - \varepsilon\sigma, p_i + \varepsilon\sigma]$ and $[p_j - \varepsilon\sigma, p_j + \varepsilon\sigma]$ do not lie on the same line. Assign to the original instance of the PC problem an appropriate instance of the PASC problem by the equations

$$A = P, \quad B = (P - \varepsilon\sigma) \cup (P + \varepsilon\sigma) \quad \text{and } t = 2s + 1.$$

One can easily verify that the above-described procedure can be carried out in time bounded from above by a polynomial in the length of a setting of the instance of PC.

Theorem 5 (Khachai, 2008). *The set P has a cover consisting of s lines if and only if the sets*

$$A = P \quad \textit{and} \;\; B = (P - \varepsilon\sigma) \cup (P + \varepsilon\sigma) \tag{6}$$

are separable by an affine committee of $2s + 1$ elements.

COROLLARY 1. The PASC problem is NP-complete in the strong sense. The ASC(n)[1] problem within an arbitrary $n > 1$ is NP-complete in the strong sense as well.

COROLLARY 2. The MASC(n) problem within an arbitrary $n > 1$ is NP-hard.

COROLLARY 3. The problem of k-polyhedral separability with a given formula formulated in $\mathbb{Z}^n$ within a fixed $n > 1$ is NP-complete in the strong sense.

3. The MASC Problem: Case of General Position

Theorem 5 has been proved under the implicit assumption on possible degeneracy of the set $A \cup B$ (i.e., when this set in not in general position). In this section, a similar result obtained without this assumption is proved.

DEFINITION 3. A set $Z \subset \mathbb{R}^n$, $|Z| > n$, is said to be in *general position* if the equality $\dim \operatorname{aff}(Z') = n$ holds for each subset $Z' \subseteq Z$, $|Z'| = n + 1$.

Particularly, a set $Z \subset \mathbb{R}^2$ is in general position (by the definition) if there is no any subset $Z' = \{z_1, z_2, z_3\} \subseteq Z$ such that points z_1, z_2, and z_3 belong to the same straight line. It is evident that the instances of the PASC problem such that the set $A \cup B$ is not in general position are very rare (one can even say that they constitute a set of measure zero). So, it could be interesting to investigate computational complexity of the PASC problem, wherein the set $A \cup B$ taken according to Definition 3.

PROBLEM 9 (Planar affine separating committee for sets in general position (PASC-GP)). Let finite sets $A = \{a_1, \ldots, a_{m_1}\}$ and $B = \{b_1, \ldots, b_{m_2}\}$, $A, B \subset \mathbb{Q}^2$, such that the set $A \cup B$ is in general position, and a natural number t be given. Does there exists an affine committee Q with at most t elements that separates the sets A and B?

Along with the PASC-GP problem, consider a modification of the MASC(n) problem taking into account the additional constraint on general position.

PROBLEM 10 (Minimum affine separating committee in a space of fixed dimension n for sets in general position (MASC-GP(n))). Let finite sets $A = \{a_1, \ldots, a_{m_1}\}$ and

[1] ASC(n) is a decision version of the MASC(n) problem.

$B = \{b_1, \ldots, b_{m_2}\}$, $A, B \subset \mathbb{Q}^n$, such that the set $A \cup B$ is in general position be given. It is required to find an affine committee Q with the minimum number of elements that separates the sets A and B.

Similarly to the considerations from the previous section, assume that an instance of the PC problem is defined by a finite set P of k integer points and some natural number s. Let numbers ρ and ε be determined by formulas (5). Let us fix 2-dimensional vectors σ and τ such that $\|\sigma\|_2 = \|\tau\|_2 = 1$, $\sigma^T\tau = 0$, and for any $\{i, j\} \subset \mathbb{N}_k$ the pairs of line segments $[p_i - \varepsilon\sigma, p_i + \varepsilon\sigma]$ and $[p_j - \varepsilon\sigma, p_j + \varepsilon\sigma]$ and also $[p_i - \varepsilon\tau, p_i + \varepsilon\tau]$ and $[p_j - \varepsilon\tau, p_j + \varepsilon\tau]$ do not lie on the same line. To the original instance of the PC problem assign the instance of the PASC-GP problem (Fig. 1), which is defined by the equalities

$$A = \{p \pm \frac{\varepsilon(p)}{M}\tau\colon\ p \in P\}, \quad B = \{p \pm \varepsilon(p)\sigma\colon\ p \in P\}, \quad \text{and } \ t = 2s + 1.$$

Here the numbers $\varepsilon(p)$ and $M > 0$ are chosen in such a way that the following inequality

$$\max_{p \in P} \frac{\varepsilon(p)}{M} < \min_{p \in P} \varepsilon(p)$$

is valid and the set $A \cup B$ is in general position. As well as in the case of the PASC problem, it is easily verified that the above-described reduction can be done in polynomial time.

Theorem 6. *The set $P = \{p_1, \ldots, p_k\} \subset \mathbb{Z}^2$ has a cover of s straight lines if and only if the sets $A = \{p \pm \frac{\varepsilon(p)}{M}\tau\colon\ p \in P\}$ and $B = \{p \pm \varepsilon(p)\sigma\colon\ p \in P\}$ are separable by an affine committee of $2s + 1$ elements.*

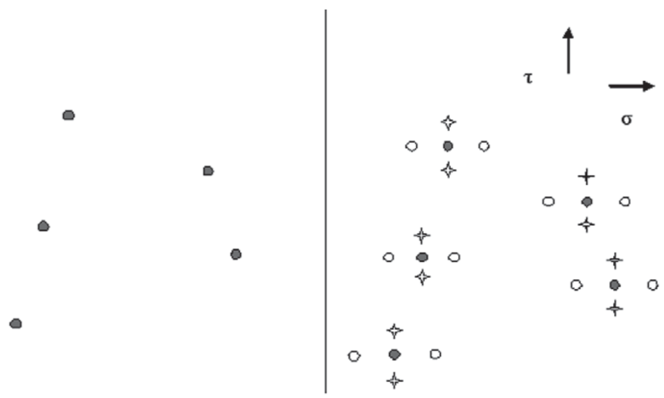

Fig. 1. Reduction of PC to PASC-GP.

Proof. 1. Consider an arbitrary cover L of the set P by straight lines. To each line $l_j \in L$, $l_j = \{x \in \mathbb{R}^2\colon\ c_j^T x = d_j\}$, assign the subset $P(j) = P \cap l_j$ and subsets

$$A(j) = \left\{p \pm \frac{\varepsilon(p)}{M}\tau\colon\ p \in P(j)\right\} \quad \text{and } B(j) = \big\{p \pm \varepsilon(p)\sigma\colon\ p \in P(j)\big\}.$$

For any point $p \in P(j)$, the appropriate elements $p - \varepsilon(p)\sigma$ and $p + \varepsilon(p)\sigma$ of the set $B(j)$, obviously, satisfy the inequality

$$\big(c_j^T(p - \varepsilon\sigma) - d_j\big)\big(c_j^T(p + \varepsilon\sigma) - d_j\big) < 0, \tag{7}$$

i.e., these points are located on the different sides of the line l_j. Take an arbitrary number δ_j, satisfying the condition

$$\max_{p \in P} \frac{\varepsilon(p)}{M} < \delta_j < \min_{p \in P} \varepsilon(p),$$

in such a way that for the functions f_{2j-1}, f_{2j}, defined by the formulas

$$f_{2j-1}(x) = c_j^T x - d_j + \delta_j, \quad f_{2j}(x) = -c_j^T x + d_j + \delta_j, \tag{8}$$

the inequalities

$$\left.\begin{aligned} &f_{2j-1}(p - \varepsilon(p)\sigma) \cdot f_{2j}(p - \varepsilon(p)\sigma) < 0,\\ &f_{2j-1}(p + \varepsilon(p)\sigma) \cdot f_{2j}(p + \varepsilon(p)\sigma) < 0,\\ &f_{2j-1}(p - \tfrac{\varepsilon(p)}{M}\tau) \cdot f_{2j}(p - \tfrac{\varepsilon(p)}{M}\tau) > 0,\\ &f_{2j-1}(p + \tfrac{\varepsilon(p)}{M}\tau) \cdot f_{2j}(p + \tfrac{\varepsilon(p)}{M}\tau) > 0 \end{aligned}\right\} \quad (p \in P(j))$$

are valid. By virtue of (7), such a construction is possible.

By the choice of ε, the inequalities

$$\begin{aligned} &f_{2j-1}(a) > 0, f_{2j}(a) > 0, && (a \in A(j)),\\ &f_{2j-1}(x) \cdot f_{2j}(x) < 0, && (x \in A \cup B \setminus (A(j) \cup B(j)) \end{aligned}$$

are valid as well.

The constructed finite sequence of functions $(f_1, ..., f_{2s})$ possesses the property

$$\begin{aligned} &|\{i\colon\ f_i(a) > 0\}| \geqslant s + 1,\ (a \in A),\\ &|\{i\colon\ f_i(b) < 0\}| = s, \qquad (b \in B). \end{aligned}$$

Let us supplement this sequence with an arbitrary affine function f_0 satisfying the condition

$$f_0(x) < 0, \quad (x \in A \cup B) \tag{9}$$

that is feasible by virtue of finiteness of the set $A \cup B$. The sequence

$$Q = (f_0, f_1, \dots, f_{2s})$$

is the required affine committee of $2s+1$ elements that separates the sets A and B (Fig. 2).

2. Let the sequence $Q = (f_1, ..., f_q)$ be an arbitrary affine committee separating the sets A and B. By the definition of an affine separating committee, for each point $p \in P$ and for each pair $\{1,2\}$, $\{2,3\}$, $\{3,4\}$ and $\{1,4\}$ (Fig. 3), there is an appropriate element of Q that classifies this pair correctly. Let p be an arbitrary element of P and $f(x) = c^T x - d$ be a committee member classifying the points $p - \varepsilon(p)\sigma$ and $p + \frac{\varepsilon(p)}{M}\tau$ correctly, i.e., in such a way that the inequalities

$$\begin{aligned} &f\left(p + \frac{\varepsilon(p)}{M}\tau\right) = c^T p + \frac{\varepsilon(p)}{M} c^T \tau - d > 0, \\ &f(p - \varepsilon(p)\sigma) = c^T p - \varepsilon(p) c^T \sigma - d < 0 \end{aligned}$$

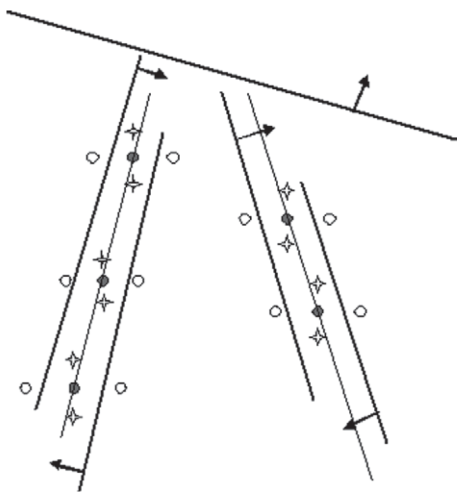

Fig. 2. The construction of a separating committee in the case of general position.

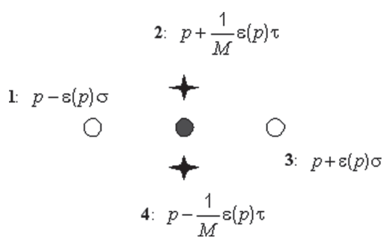

Fig. 3. The numeration of the points.

are valid. Whence,

$$c^T \left(\sigma + \frac{1}{M}\tau \right) > 0. \tag{10}$$

Let us prove that for any $p' \in P$ the points $p' + \varepsilon(p')\sigma$ and $p' - \frac{\varepsilon(p')}{M}\tau$ cannot be correctly classified by the same function f. Indeed, if, on the contrary, the inequalities

$$f\left(p' - \frac{\varepsilon(p')}{M}\tau\right) = c^T p' - \frac{\varepsilon(p')}{M} c^T \tau - d > 0,$$
$$f(p' + \varepsilon(p')\sigma) = c^T p' + \varepsilon(p') c^T \sigma - d < 0$$

are valid, then

$$c^T \left(\sigma + \frac{1}{M}\tau \right) < 0.$$

This contradicts (10).

For each point $p \in P$, let us denote by $I_1(p)$ the set of indices of the elements of Q classifying correctly the pair $\{1, 2\}$, and by $I_2(p)$ the same for the pair $\{3, 4\}$. Further, define the sets I_1 and I_2 by the equalities

$$I_1 = \bigcup_{p \in P} I_1(p), \quad I_2 = \bigcup_{p \in P} I_2(p).$$

From the above proof, the constructed sets I_1 and I_2 are nonempty and $I_1 \cap I_2 = \emptyset$. Consider the set $P'(i) = \{p \in P\colon\ i \in I_1(p)\}$ for an arbitrary $i \in I_1$. Obviously, $\bigcup_{i \in I_1} P'(i) = P$. Further, the straight line $f_i(x) = 0$ intersects the neighborhood $B(p, \varepsilon)$ for any point $p \in P'(i)$. Hence, $\dim \operatorname{aff}(P'(i)) \leqslant 1$, by the choice of ε and Proposition 1. Therefore, the set P has a cover consisting of $|I_1|$ lines and, consequently, $|I_1| \geqslant s$, by the choice of s. The inequality $|I_2| \geqslant s$ can be proved by analogy. Since $I_1 \cap I_2 = \emptyset$, the condition

$$q \geqslant |I_1| + |I_2| \geqslant 2s \tag{11}$$

is valid.

It can be easily verified that the stronger inequality $q \geqslant 2s + 1$ holds as well. Indeed, otherwise the committee Q of $2s$ elements can be transformed by eliminating an arbitrary element (see, for example, Mazurov, 1971), into an affine committee separating the sets A and B and consisting of $2s - 1$ elements. But this contradicts (11). The theorem is proved.

COROLLARY 4. The PASC-GP problem is NP-complete in the strong sense. The MASC-GP(n) and the MASC-GP problems[2] are NP-hard.

[2]Similarly to MASC-GP(n), the MASC-GP problem is a modification of the MASC problem with an additional constraint on general position.

COROLLARY 5. The problem of k-polyhedral separability with a given formula formulated in $\mathbb{Z}^n$ within an arbitrary fixed $n > 1$ is NP-complete in the strong sense even under an additional constraint on general position of sets to be separated.

COROLLARY 6. The OCP problem is NP-hard even under the following additional constraints:

1. the dimension of the input feature space is an arbitrary fixed $n > 1$;
2. an activation function of the output neuron is defined by the formula

$$f_{q+1}(x) = \begin{cases} 1, & \text{if } \sum_{i=1}^n x_i > 0, \\ -1, & \text{otherwise;} \end{cases}$$

3. the set of elements of training sample (2) is in general position.

4. Approximation Algorithm

In this section, a new approximation algorithm for the MASC-GP(n) problem is presented. Actually, this algorithm is a modification of the known approximation algorithm (Khachay, 2006b) for the more general MASC problem. This modification effectively uses additional constraints to improve its approximation ratio.

Introduce some necessary notation and definitions. Let an instance of the MASC-GP(n) problem be defined by some finite set $Z = A \cup B \subset \mathbb{Q}^n$.

DEFINITION 4. A subset $Z' = A' \cup B'$, where $A' \subseteq A$, $B' \subseteq B$, is called *an affine separable subset* (of the set Z), if there exist a vector $c \in \mathbb{R}^n$ and a number $d \in \mathbb{R}$ such that

$$\begin{cases} c^T a - d > 0, & (a \in A'), \\ c^T b - d < 0, & (b \in B'). \end{cases} \tag{12}$$

Denote a solution set of system (12) by $\mathfrak{S}(Z')$.

DEFINITION 5. An affine separable subset Z' is called *a maximum (by inclusion) affine separable subset* of the set Z, if for each $z \in Z \setminus Z'$

$$\mathfrak{S}(Z' \cup \{z\}) = \emptyset.$$

Denote by $\mathfrak{M}(Z)$ the set of maximum affine separable subsets of the set Z. One of the alternatives below is valid.

1. The set Z is affine separable, i.e., $\mathfrak{M}(Z) = \{Z\}$. In this case, for an arbitrary pair $(c, d) \in \mathfrak{S}(Z)$, the finite sequence $Q = (c^T x - d)$ (consisting of a single element) is a solution of the problem MASC-GP(n).

2. There exists a subset $Z' \subset Z$, $Z' \neq Z$ such that $Z' \in \mathfrak{M}(Z)$. In this case, any minimum affine committee separating the sets A and B consists of more than one element.

The following proposition is valid.

PROPOSITION 2. Let an affine committee separating the sets A and B, consists of p elements. There exist subsets $Z'_1, \dots, Z'_p$, $Z'_i \in \mathfrak{M}(A \cup B)$ (not necessary mutually different) such that, for any pairs $(c_i, d_i) \in \mathfrak{S}(Z'_i)$, the finite sequence $(f_1, \dots, f_p)$, where $f_i(x) = c_i^T x - d_i$, is also an affine committee separating the sets A and B.

This proposition implies that it is convenient to describe solving techniques for the MASC-GP(n) problem in terms of a graph of maximum affine separable subsets of the set $Z = A \cup B$.

DEFINITION 6. A finite graph $G_Z = (V, E)$ is called *a graph of maximum affine separable subsets* of the set Z, if $V = \mathfrak{M}(Z)$ and, for each $\{Z'_1, Z'_2\} \subset V$,

$$\{Z'_1, Z'_2\} \in E \iff Z'_1 \cup Z'_2 = Z.$$

For every set Z being in general position such that $Z = A \cup B$, the graph G_Z can be constructed according to the following simple algorithm.

Algorithm 1. **Constructing the graph** G_Z

1. Initiate $V = \emptyset$ and $E = \emptyset$.
2. For each $\zeta \subset Z$, $|\zeta| = n$ (since Z is in general position, by the condition, such a subset ζ exists),
   (a) Construct a hyperplane containing the subset ζ. Let it be defined by the equation $H(x) = 0$. Because of general position of the set Z, such a hyperplane is unique, and

   $$H(x) \neq 0 \quad (x \in Z \setminus \zeta).$$

   (b) Define sets X_1 and X_2 by the equalities

   $$X_1 = \zeta \cup \{a \in A \mid H(a) > 0\} \cup \{b \in B \mid H(b) < 0\},$$
   $$X_2 = \zeta \cup \{a \in A \mid H(a) < 0\} \cup \{b \in B \mid H(b) > 0\}.$$

   (c) For each $i \in \{1, 2\}$,
      i. exclude from V all the elements $Y \in V$ such that $Y \subset X_i$;
      ii. if there are no elements $Y \in V$ such that $X_i \subseteq Y$, then set

      $$V = V \cup \{X_i\}.$$

3. For each pair $e = \{Z'_1, Z'_2\} \subset V$ such that $Z'_1 \cup Z'_2 = Z$, set $E = E \cup \{e\}$.

Correctness of the presented above algorithm easily follows from the boundary solutions principle (Eremin, 1975) for systems of linear inequalities. Particularly, if the set Z is affine separable then the resulted graph has the form $G_Z = (\{Z\}, \emptyset)$, as expected.

Let us go through the main algorithm. Let $Z = A \cup B \subset \mathbb{Q}^n$ and $|Z| = m$.

Algorithm 2. **Greedy Committee**

1. Construct the graph $G_Z = (V, E)$.
2. If $V = \{Z\}$ then define a finite sequence $K = (Z)$, $q_{\min} = 1$ and go to Step 4, otherwise define $q_{\min} = \infty$.
3. For each $\zeta \in V$,
   3.1 define the finite sequence $K(\zeta)$, the set J, and the number $q(\zeta)$ by the equalities $K(\zeta) = (\zeta)$, $J = Z \setminus \zeta$, and $q(\zeta) = 1$;
   3.2 while $J \neq \emptyset$, repeat the following steps
      3.2.1 find $\{Z', Z''\} = \text{argmax}\,\{|X_1 \cap X_2 \cap J| \colon \{X_1, X_2\} \in E\}$;
      3.2.2 add the sets Z' and Z'' to the sequence $K(\zeta)$, redefine
$$J = J \setminus (Z' \cap Z'') \text{ and } q(\zeta) = q(\zeta) + 2.$$
   3.3 if $q(\zeta) < q_{min}$ then set $K = K(\zeta)$ and $q_{min} = q(\zeta)$.
4. Let $K = (Z'_1, \ldots, Z'_{q_{\min}})$. For each $i \in \mathbb{N}_{q_{\min}}$ define a function $f_i(x) = c_i^T x - d_i$, where (c_i, d_i) is an arbitrary element of $\mathfrak{S}(Z'_i)$. The finite sequence $Q = (f_1, \ldots, f_{q_{\min}})$ is a required affine committee separating the sets A and B.

The following assumption is necessary.

ASSUMPTION 1. Let for every affine inseparable set $Z = A \cup B$ for some t there exist subsets $Z'_0, Z'_1, \ldots, Z'_{2t} \in V$ (not necessary mutually different) such that

$$\{Z'_{2j-1}, Z'_{2j}\} \in E \quad (j \in \mathbb{N}_t)$$

and, for an arbitrary $(c_i, d_i) \in \mathfrak{S}(Z'_i)$, $i = 0, 1, \ldots, 2t$, the sequence

$$Q = (c_0^T x - d_0, c_1^T x - d_1, \ldots, c_{2t}^T x - d_{2t})$$

is a minimum affine committee separating the sets A and B.

PROPOSITION 3. Let a set $Z = A \cup B \subset \mathbb{Q}^n$, $|Z| = m$, define an instance of the MASC-GP(n) problem. Complexity of the Greedy Committee algorithm is

$$O\left(\binom{m}{n}^3 + \Theta m\right),$$

where Θ is the complexity bound of solving a feasible system of at most m linear inequalities of $n+1$ variables. The approximation ratio of the algorithm is $\mathrm{O}(m/n)$.

If Assumption 1 for the set Z is valid then the approximation ratio of the algorithm is $\mathrm{O}(\log m)$.

Proof. 1. Complexity of the algorithm is defined by complexities of Steps 4 and 4, which are $\mathrm{O}(\binom{m}{n}^3)$ and $\mathrm{O}(\Theta m)$, correspondingly.

2. Let us estimate the approximation ratio. If the set $A \cup B$ is affine separable then the algorithm solves the MASC-GP(n) problem exactly. Otherwise, the algorithm finds an affine committee of at most

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

elements separating these sets. Indeed, by the construction, there exists $X \in V$ such that $|X| \geqslant (m+n)/2$. Further, on each Step 3.2.2. of the algorithm, cardinality of the set J decreases at least by n.

Let Assumption 1 be valid. Then the MASC-GP(n) problem is equivalent to the corresponding instances of the known Set Cover problem [Johnson, 1974]. In this case, for each $\zeta \in V$, it is required to find a cover of the set $J = Z \backslash \zeta$ by subsets $Z'_i \cap Z'_j \cap J$ belonging to $\{Z'_i \cap Z'_j \cap J:\ \{Z'_i, Z'_j\} \in E\}$, where $G_Z = (V, E)$. Actually, in Step 4 of the algorithm, an appropriate instance of the Set Cover problem is solved by the known Greedy (Johnson, 1974) algorithm, which has the approximation ratio of $\mathrm{O}(\log |J|) = \mathrm{O}(\log m)$. The proposition is proved.

Assumption 1 seems to be too tight but it should be noted that instances of the MASC problem constructed in the known proofs of its hardness (particularly, in the proofs of Theorems 5 and 6) are in agreement with this assumption.

Indeed, without loss of generality, one can assume that the set Z is not affine separable. Let $Q = (f_0, f_1, \ldots, f_{2s})$ be a committee constructed at the first part of the proof of Theorem 6. Let us denote by $A'(i) \cup B'(i)$, $i = 0, \ldots, 2s$, the subset of the set $Z = A \cup B$ correctly separated by the hyperplane $f_i(x) = 0$. By the construction (see Fig.2), for each $j \in \mathbb{N}_s$,

$$(A'(2j-1) \cup B'(2j-1)) \cup (A'(2j) \cup B'(2j)) = Z.$$

Let a subset $Z'(i)$ be any maximum affine separable subset of the set Z such that $A'(i) \cup B'(i) \subseteq Z'(i)$. Since $Z'(2j-1) \cup Z'(2j) = Z$ for each $j \in \mathbb{N}_s$, then $\{Z'(2j-1), Z'(2j)\}$ is an edge of the graph G_Z, and Assumption 1 holds.

REMARK 1. Results of the present section can be easily reformulated in terms of two-layered perceptrons and the OCP problem.

5. Conclusions

As shown above, the minimum affine committee (MASC) problem is intractable not only in the general case (when the dimension n is unbounded) but also in the case of a space of fixed dimension $n > 1$ even under an additional constraint on general position of the sets A and B. Particularly, the problem formulated on the plane is NP-hard. Further, the

Table 1

The MASC problem: results

Problem	Compl. status	Approx. threshold	Approx. ratio
MASC	NP-hard	$\mathrm{O}(\log\log\log m)$	$\mathrm{O}(m)$
MASC(n), $n > 1$	NP-hard	?	$\mathrm{O}(m/n)$
PASC	NP-complete (in the strong sense)	—	—
MASC-GP	**NP-hard**	?	$\mathrm{O}(m)$
MASC-GP(n), $n > 1$	**NP-hard**	?	$\mathbf{O}(\mathbf{log\, m})$
PASC-GP	**NP-complete (in the strong sense)**	—	—

MASC problem is hardly approximable. If $NP \not\subset DTIME(2^{poly(\log n)})$, the approximation ratio of any polynomial-time algorithm can not be better than $\mathrm{O}(\log\log\log m)$ (in the worst case). But if the problem is formulated in a space of fixed dimension with an additional constraint on general position and Assumption 1 is valid, then this problem can be solved in polynomial time with approximation ratio of $\mathrm{O}(\log m)$.

The results concerning the MASC problem and its particular cases are enlisted in Table 1. Results obtained in this paper are marked by bold font, open questions are denoted by the question sign. As corollaries, some new results concerning the known OCP and k-polyhedral separability combinatorial optimization problems are also obtained.

Acknowledgements

Research was supported in part by the Foundation of the President of Russian Federation (grants No. MD-370.2008.1 and NS-2081.2008.1), by the Ural-Siberian interdisciplinary project, and by the Russian Foundation for Basic Research (grant No. 07-07-00168).

References

Blum, A.L., and R.L. Rivest (1992). Training a 3-node neural network is NP-complete. *Neural Networks*, **5**, 117–127.

Eremin, I.I., and Astafiev N.N. (1975). *Introduction into the Theory of Linear and Convex Programming*. Phys. Mat. Lit., Moscow, 192 pp. (in Russian).

Johnson, D.S. (1974). Approximation algorithms for combinatorial problems. *Journal of Computer and System Sciences*, **9**(3), 256–278.

Khachai, M.Yu. (2006a). On the computational complexity of minimum committee problem and related problems. *Doklady Mathematics*, **73**(1), 138–141.

Khachay, M.Yu. (2006b). On the computational and approximational complexity of the minimum affine separating committee problem. *Tavricheskii Vestnik Informatiki i Matematiki*, **1** 34–43 (in Russian).

Khachai, M.Yu. (2008). Computational and approximational complexity of combinatorial problems related to the committee polyhedral separability of finite sets. *Pattern Recognition and Image Analysis*, **18**(2), 237–242.

Khachai, M.Yu., and M.I. Pobery (2008). Computational complexity and approximability of combinatorial optimization problems connected with committee poplyhedral separability of finite sets. In *20th Intern. Conf. EURO Mini Conf. "Continuous Optimiz. & Knowledge-Based Technolog. (EurOPT-2008)"*. Selected papers. Technika, Vilnius. pp. 42–47.

Lin, J.H., and J.S. Vitter (1991). Complexity results on learning by neural nets. *Machine Learning*, **6**, 211–230.

Mazurov, Vl.D. (1971). Committees of inequalities systems and the pattern recognition problem. *Kibernetika*, **3**, 140–146 (in Russian).

Megiddo, N., and A. Tamir (1982). On the complexity of locating linear facilities in the plane. *Operations Research Letters*, **5**(1), 194–197.

Megiddo, N. (1988). On the complexity of polyhedral separability. *Discrete and Computational Geometry*, **3**, 325–337.

M. Khachay was awarded the candidate of mathematical and physical sciences degree at the Institute of Mathematics and Mechanics, Ural Branch of RAS (IMM UB of RAS), Ekaterinburg in 1996, doctor degree of mathematical and physical sciences at Dorodnicyn Computing Center RAS, Moscow, in 2005. At present, he is a head of the Mathematical Programming Department of IMM UB of RAS and a professor of the Ural State University at the Mathematical Economy Chair. His research interests include problems of combinatorial optimization, complexity, and machine learning.

M. Poberii is a post-graduate student at the Institute of Mathematics and Mechanics, Ural Branch of RAS, Ekaterinburg. Her scientific interests include questions of theoretical computer science.

Aibių komitetų poliedrinio atskiriamumo ir aproksimacijų sudėtingumas

Michael KHACHAY, Maria POBERII

Gerai žinoma, kad minimaliai afininė komitetų atsikriamumo (the minimum affine separating committee – MASC – angl.) problema yra NP-sudėtingumo kombinatorinis uždavinys, pritaikomas kai kuriuose mašininio mokymo algoritmuose. Šiame darbe parodyta, kad MASC problema fiksuoto matavimo erdvėje, kai $n > 1$, yra bendru atveju eksponentinio sudėtingumo. Yra pasiūlytas naujas polinominio sudėtingumo aproksimacinis algoritmas, nustatant aproksimacijos eilę bei sudėtingumo ribas.