# An Exact Pseudopolynomial Algorithm for a Problem of Finding a Family of Disjoint Subsets

Alexandr Galashov[1] and Alexander Kel'manov[1,2]

[1] Novosibirsk State University,
2 Pirogova St., 630090 Novosibirsk, Russia
[2] Sobolev Institute of Mathematics,
4 Koptyug Ave., 630090 Novosibirsk, Russia
galashov.alexandr@gmail.com, kelm@math.nsc.ru

**Abstract.** We consider a strongly NP-hard problem of finding a family of disjoint subsets with given cardinalities in a finite set of points from Euclidean space. The minimum of the sum over all subsets from required family of the sum of the squared distances from the elements of these subsets to their centers is used as a search criterion. The subsets centers are optimizable variables defined as the mean values of the elements of these subsets. With an additional restriction on the problem that the coordinates of the input points are integer, an exact algorithm is proposed. This algorithm is pseudopolynomial in the case of fixed space dimension and of fixed number of required subsets.

**Keywords:** Euclidean space, subsets search, clustering, NP-hard problem, exact pseudopolynomial-time algorithm.

## Introduction

The subject of our study is a strongly NP-hard problem of finding a family of disjoint subsets in a finite set of points from Euclidean space. Our aim is justification of an exact pseudopolynomial algorithm for a special case of this problem.

The investigation is motivated by poor study of the problem and its importance in applications, in particular, for mathematical problems of data analysis, approximation theory and mathematical statistics. The problem models a situation where it's required to classify noisy data provided from experimental observations for the states of some material objects (see [1–4] and works cited there).

The paper is organized as follows. In the next section the formal definition of the problem under study is given; an example of application (origin) of the problem being investigated is also presented. In Section 3, we provide a review of the known results and announce the obtained algorithmic result. Basic definitions and statements, that give elements to prove the properties of the proposed algorithm, are presented in Section

In: A. Kononov et al. (eds.): DOOR 2016, Vladivostok, Russia, published at http://ceur-ws.org

4. Finally, in Section 5 we construct an exact algorithm for solving the problem under study, justify its properties and show its pseudopolynomiality for a special case of the problem.

## 1 Problem formulation and its origin

Everywhere below $\mathbb{R}$ denotes the set of real numbers, $\|\cdot\|$ denotes the Euclidean norm, and $\mathbb{Z}$ denotes the set of integer numbers.

We consider the following problem.

*Problem 1. Given* a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of points from $\mathbb{R}^q$ and some positive integers $M_1, \ldots, M_J$. *Find* a family $\{\mathcal{C}_1, \ldots, \mathcal{C}_J\}$ of disjoint subsets of $\mathcal{Y}$ such that

$$F(\mathcal{C}_1, \ldots, \mathcal{C}_J) = \sum_{j=1}^{J} \sum_{y \in \mathcal{C}_j} \|y - \overline{y}(\mathcal{C}_j)\|^2 \to \min \ , \tag{1}$$

where $\overline{y}(\mathcal{C}_j) = \frac{1}{|\mathcal{C}_j|} \sum\limits_{y \in \mathcal{C}_j} y$ is the centroid (geometrical center) of the subset $\mathcal{C}_j$, under constraints $|\mathcal{C}_j| = M_j, j = 1, \ldots, J$, and

$$\sum_{j=1}^{J} M_j \leq N \ , \tag{2}$$

on the cardinalities of the required subsets.

The Problem 1 has the following origin [5]. Given a table $\mathcal{Y} = \{y_1, \ldots, y_N\}$ containing results of multiple measurements of the $q$-dimenstional tuple $y$ of numerical characteristics corresponding to $J$ unique objects. Numbers $M_j$, $j = 1, \ldots, J$, of measurements for the $j$-th object are known. Moreover, the table includes $(N - \sum_{j=1}^{J} M_j)$ results of single measurements of the states of some random objects. Each result of measurement, presented in the table, has an error. Furthermore, the correspondence between the measurements and the objects is unknown. It's required to find a family $\{\mathcal{C}_1, \ldots, \mathcal{C}_J\}$ of disjoint subsets of the set $\mathcal{Y}$, using the minimum of squared distances criterion, where each subset contains the elements corresponding to the appropriate unique object, and estimate values $\overline{y}(\mathcal{C}_1), \ldots, \overline{y}(\mathcal{C}_J)$ corresponding to the numerical characteristics of the unique objects (assuming that the data has a measurement error).

The Problem 1 regularly appears as a mathematical problem, in particular, for Data Analysis and Pattern Recognition, and is induced [5] by the following approximation model.

Given a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of points from $\mathbb{R}^q$ and some positive integers $M_1, \ldots, M_J$. Find a family $\{\mathcal{M}_1, \ldots, \mathcal{M}_J\}$ of disjoint subsets of the set $\mathcal{N} = \{1, \ldots, N\}$ such that

$$\sum_{n \in \mathcal{N}} \|y_n - x_n\|^2 \to \min \ , \tag{3}$$

where

$$x_n = \begin{cases} w_1 \in \mathbb{R}^q, \ n \in \mathcal{M}_1, \\ \dots \\ w_J \in \mathbb{R}^q, \ n \in \mathcal{M}_J, \\ v_n \in \mathbb{R}^q, \ n \in \mathcal{N} \setminus (\cup_{j=1}^{J} \mathcal{M}_j), \end{cases} \tag{4}$$

under the same constraints on the required subsets, as in Problem 1.

This problem consists in approximation of the sequence $y_n$ by the sequence $x_n$ using the minimum of squared distances criterion under the condition that the structure of the sequence $x_n$ is described by the formula (4). In this formula, the points $w_1, \dots, w_J$ are interpreted as the numerical characteristics describing the unique objects. The points from the set $\{v_i, i \in \mathcal{N} \setminus (\cup_{j=1}^{J} \mathcal{M}_j)\}$ are considered as the collection of characteristics describing the random objects.

Having uncovered the sum (3) using (4) and grouped the terms, we can check using derivation, that for each fixed collection $\{\mathcal{M}_1, \dots, \mathcal{M}_J\}$ of sets, the values $w_j = \overline{y}(\mathcal{M}_j)$, $j = 1, \dots, J$, provide the points where the objective function (3) attains its minimum. If we put these values in the formulae (3), (4) and let $\mathcal{C}_j = \{y_i \in \mathcal{Y} \mid i \in \mathcal{M}_j\}$, then we can simply verify (see also [5]) that the model (3), (4) induces Problem 1.

A similar to Problem 1 is well-known strongly NP-hard [6] problem MSSC (Minimum Sum-of-Squares Clustering) of clustering, which is also known as $k$-means [7], [8], [9]. In this problem the objective function is the same as in Problem 1, but the cardinalities of the required clusters are optimizable variables and instead of searching a family of subsets which union could not cover all input set, we have to find a partition of this set.

Notice, that the numbers from the set $\mathcal{N} \setminus (\cup_{j=1}^{J} \mathcal{M}_j)$ in the model (3), (4) correspond to the elements from the set $\mathcal{Y} \setminus (\cup_{j=1}^{J} \mathcal{C}_j)$ in Problem 1. Elements of these sets do not appear in the formulae for the estimation of the values $w_j$, $j = 1, \dots, J$, and the corresponding centroids $\overline{y}(\mathcal{C}_j)$, $j = 1, \dots, J$, in Problem 1. Therefore, Problem 1 could be considered as a problem of Data Censoring [10]. In this applied problem, a part of the data (generally, unknown part) of obtained experimental results in a situation of random failure of measurement instument has to be excluded from the estimation procedure, because this data distorts final estimations.

## 2 Known and obtained results

The strong NP-hardness of Problem 1 is implied from the results obtained in [11], since in the cited work it was proved that the special case of Problem 1 when $J = 1$ is NP-hard in the strong sense problem.

The Problem 1 is referred to the algorithmically poorly-studied problems of the discrete optimization. In [5] for this problem, was proposed a 2-approximation algorithm which time complexity is equal to $\mathcal{O}(N^2(N^{J+1} + q))$. For the case of Problem 1 when the number $J$ of required subsets is fixed (not the input of the problem), this algorithm is polynomial. Currently, there are no other algorithmic results for Problem 1 and the known results [12–15] were obtained only for its special case when $J = 1$. These results are described below.

For Problem 1 in [12], a 2-approximation polynomial-time algorithm of complexity $\mathcal{O}(qN^2)$ was constructed.

For the variation of Problem 1 with an additional restriction that the coordinates of the input points are integer and for the case of fixed space dimension, in [13] an exact pseudopolynomial algorithm was presented, which time complexity is equal to $\mathcal{O}(N(MB)^q)$, where $B$ is the maximum absolute value of the coordinates of the input points.

Furthermore, for the case of the fixed space dimension in [14] an FPTAS was proposed. This scheme for a given relative error $\varepsilon$ finds $(1+\varepsilon)$-approximate solution in $\mathcal{O}(N^2(M/\varepsilon)^q)$ time, that is polynomial in the size of input and $1/\varepsilon$.

A PTAS of complexity $\mathcal{O}(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$, where $\varepsilon$ is a guaranteed relative error, was found in [15].

In the current work, for the variation of Problem 1 with an additional restriction that the coordinates of the input points are integer, an algorithm which finds an exact solution in $\mathcal{O}(N(N^2+qJ)(2MB+1)^{qJ}+J^2\,log^2 N)$ time is constructed, where $B$ is the maximum absolute value of the coordinates of the input points and $M$ is the least common multiple for the numbers $M_1,\ldots,M_J$. In the case of the fixed dimension $q$ of the space and of the fixed number $J$ of required subsets, the proposed algorithm is pseudopolynomial and its time complexity is bounded by $\mathcal{O}(N^3(MB)^{qJ})$.

## 3 Fundamentals of algorithm

In order to justify the properties of the algorithm, we need a few basic statements, an auxiliary problem and an exact polynomial-time algorithm finding its solution.

The following statements are the geometrical basis of the algorithm.

**Lemma 1.** *Let all conditions of Problem 1 be satisfied and $\mathcal{Y}\subset\mathbb{Z}^q$. Moreover, let*

$$B=\max_{y\in\mathcal{Y}}\max_{i\in\{1,\ldots,q\}}|(y)^i| \tag{5}$$

*be the maximum absolute value of the coordinates of the input points, where $(y)^i$ is the $i$-th coordinate of the point $y$. Then, $\overline{y}(\mathcal{C}_j)\in\mathcal{D}$, $j=1,\ldots,J$, where*

$$\mathcal{D}=\{z\in\mathbb{R}^q|\ (z)^k=\frac{1}{M}(v)^k,(v)^k\in\mathbb{Z}^q,|(v)^k|\le MB,k=1,\ldots,q\}\ , \tag{6}$$

*where $M$ is the least common multiple for the numbers $M_1,\ldots,M_J$.*

*Proof.* According to the definition of the geometrical center (as the mean over the set), we have

$$(\overline{y}(\mathcal{C}_j))^k=\frac{1}{|\mathcal{C}_j|}\sum_{y\in\mathcal{C}_j}(y)^k=\frac{p}{M_j},\ \ j=1,\ldots,J,\ \ k=1,\ldots,q\ ,$$

where $p$ is an integer such that

$$|p|=|\sum_{y\in\mathcal{C}_j}(y)^k|\le M_jB\le MB\ .$$

By the definition of the least common multiple $M$, for each $j = 1, \ldots, J$, there exists an integer $s(j)$ such that $0 < s(j) \leq M$, and $M = s(j)M_j$. Therefore,

$$\frac{p}{M_j} = \frac{s(j)p}{s(j)M_j} = \frac{s(j)p}{M},\ j = 1, \ldots, J\ .$$

To conclude, notice that $s(j)p \in \mathbb{Z}$ and

$$\frac{|s(j)p|}{M} \leq Bs(j) \leq BM\ .$$

□

The set $\mathcal{D}$ is the multi-dimensional grid with the rational step and with a center in the origin. For the cardinality of this grid the following obvious equality holds

$$|\mathcal{D}| = (2MB + 1)^q\ . \tag{7}$$

**Lemma 2.** *For each non-empty finite set $\mathcal{Z}$ of points from $\mathbb{R}^q$, the minimum over $x$ of the $\sum\limits_{z \in \mathcal{Z}} \|z - x\|^2$ is attained at $x = \overline{z} = \frac{1}{|\mathcal{Z}|} \sum_{z \in \mathcal{Z}} z$.*

It could be easily checked by the derivation.

The calculation basis of the algorithm is given by an exact polynomial-time algorithm finding the solution of the following auxiliary problem.

*Problem 2. Given* a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$, a tuple $b = (b_1, \ldots, b_J)$ of points from $\mathbb{R}^q$ and some positive integers $M_1, \ldots, M_J$. *Find* a family $\{\mathcal{B}_1, \ldots, \mathcal{B}_J\}$ of disjoint subsets of the set $\mathcal{Y}$ such that

$$G^b(\mathcal{B}_1, \ldots, \mathcal{B}_J) = \sum_{j=1}^{J} \sum_{y \in \mathcal{B}_j} \|y - b_j\|^2 \to \min\ , \tag{8}$$

under constraints $|\mathcal{B}_j| = M_j$, $j = 1, \ldots, J$, and (2) on the cardinalities of the required subsets.

The Problem 2 can be reduced [5] to the known assignment problem (see for example [16]) and its exact solution can be found [5] in $\mathcal{O}(N(N^2 + qJ))$ time. In the following, the algorithm solving Problem 2 is noted by $\mathcal{A}_1$.

The following technical lemma lies in the fundamentals of the alorithm and is given for the completeness of the description.

**Lemma 3.** *The least common multiple for the numbers $M_1, \ldots, M_J$, can be found in $\mathcal{O}(J \sum_{j=1}^{J} \log^2 M_j)$ time.*

*Proof.* Let $M = lcm(M_1, \ldots, M_J)$ be the least common multiple for the numbers $M_1, \ldots, M_J$. Using well known properties of the least common multiple, we have the following equalities.

$$lcm(M_i, M_j) = \frac{M_i M_j}{gcd(M_i, M_j)}\ , \tag{9}$$

where $gcd(M_i, M_j)$ is the greatest common divisor for the numbers $M_i, M_j$.

$$lcm(M_1, \ldots, M_J) = lcm(lcm(M_1, \ldots, M_{J-1}), M_J) \ . \tag{10}$$

It's known that the greatest common divisor for the numbers $a, b$ can be found by the Euclidean algorithm in $\mathcal{O}(log^2(max(a, b)))$ time.

From (9) and (10), the obvious recurrent algorithm of computing the least common multiple for the numbers $M_1, \ldots, M_J$, is implied. For the complexity of computing the least common multiple for the numbers $M_1, \ldots, M_J$, we have

$$\begin{aligned}\sum_{i=1}^{J-1}(\log(max(M_{J-1+1}, lcm(M_1, \ldots, M_{J-i})))^2 \leq \sum_{i=1}^{J}(\log^2(\prod_{k=1}^{J-i+1} M_k))& \\ = \mathcal{O}(J(\sum_{k=1}^{J}\log^2 M_k))& \ ,\end{aligned}$$

where we used the following obvious upper bound on the least common multiple

$$lcm(M_1, \ldots, M_{J-i}) \leq \prod_{k=1}^{J-i} M_k \ .$$

□

## 4 Exact pseudopolynomial algorithm

The idea of the proposed algorithm is as follows. In the region of the input space, defined by the maximum absolute value of the coordinates of the input points, a multidimensional grid is constructed which is uniform over each coordinate and has the rational step. By construction, all centroids including optimals belong to the grid.

For each tuple of $J$ points from the constructed grid, the auxiliary Problem 2 is solved. Its solution — a family of disjoint subsets — is included in the collection of the candidates on the solution for Problem 1.

The family of the subsets with minimum value of the objective function of Problem 2 is taken as a solution of Problem 1.

Let us formulate an algorithm which implements the described approach.

*Algorithm* $\mathcal{A}$.

*Input*: Set $\mathcal{Y}$ and positive integers $M_1, \ldots, M_J$.

`Step 1`. Find the least common multiple $M$ for the numbers $M_1, \ldots, M_J$ using (9) and (10), and the value $B$ using formula (5). Construct the grid $\mathcal{D}$ using formula (6).

`Step 2`. For every tuple $d = (d_1, \ldots, d_J) \in \mathcal{D}^J$, using Algorithm $\mathcal{A}_1$, find and memorize the exact solution $\{\mathcal{B}_1(d), \ldots, \mathcal{B}_J(d)\}$ of the auxiliary Problem 2 and the value of the objective function $G^d$, putting $b = d$ in the formula (8).

`Step 3`. In the collection $\{\mathcal{B}_1(d), \ldots, \mathcal{B}_J(d)\}$, $d \in \mathcal{D}^J$, of the solutions found on step 2, find the tuple $d^A = (d_1^A, \ldots, d_J^A)$ on which the objective function $G^d$ attains its minimum. As the solution $\{\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\}$ of Problem 1, we take constructed on step 2 subsets $\mathcal{B}_1(d^A), \ldots, \mathcal{B}_J(d^A)$ corresponding to the tuple $d^A$.

*Output*: Collection $\{\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\}$.

Next statement specifies the properties of the proposed algorithm.

**Theorem 1.** *Let in the conditions of Problem 1, the coordinates of all points from the set $\mathcal{Y}$ have integer values in the interval $[-B, B]$. Then, algorithm $\mathcal{A}$ finds an optimal solution of this problem in $\mathcal{O}(N(N^2 + qJ)(2MB + 1)^{qJ} + J^2\, log^2 N)$ time.*

*Proof.* Let the subsets $\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*$ be the optimal solution of Problem 1. According to the definition of step 3, algorithm $\mathcal{A}$ finds the solution of Problem 1 in the form:

$$\{\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\} = \{\mathcal{B}_1(d^A), \ldots, \mathcal{B}_J(d^A)\} \ , \tag{11}$$

where

$$d^A = \arg\min_{d \in \mathcal{D}^J} G^d(\mathcal{B}_1(d), \ldots, \mathcal{B}_J(d)) \ . \tag{12}$$

Thus, noticing that the elements of the optimal tuple $\overline{y}^* = (\overline{y}(\mathcal{C}_1^*), \ldots, \overline{y}(\mathcal{C}_J^*))$ belong to the set $\mathcal{D}$, according to Lemma 1, from the definitions (8), (12) and (1), the following is implied

$$G^{d^A} \le \sum\nolimits_{j=1}^{J} \sum_{y \in \mathcal{C}_j^*} \|y - \overline{y}(\mathcal{C}_j^*)\|^2 = F(\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*) \ . \tag{13}$$

Then, using Lemma 2, definitions (1), (8), (11), we obtain

$$\begin{aligned} F(\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A) &= \sum\nolimits_{j=1}^{J} \sum_{y \in \mathcal{C}_j^A} \|y - \overline{y}(\mathcal{C}_j^A)\|^2 \\ &\le \sum\nolimits_{j=1}^{J} \sum_{y \in \mathcal{C}_j^A} \|y - d_j^A\|^2 = \sum\nolimits_{j=1}^{J} \sum_{y \in \mathcal{B}_j(d^A)} \|y - d_j^A\|^2 = G^{d^A} \ . \end{aligned} \tag{14}$$

Combining (13) and (14), we find the estimation

$$F(\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A) \le F(\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*) \ .$$

On the other hand, the sets $\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A$ are the feasible solution of Problem 1. Thus, we have the estimation

$$F(\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*) \le F(\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A) \ .$$

From the obtained estimations, we have the equiality

$$F(\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A) = F(\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*) \ .$$

Let us estimate the time complexity of the algorithm using its stepwise notation.

By Lemma 3, computational costs of computing the least common multiple for the numbers $M_1, \ldots, M_J$ on step 1 are $\mathcal{O}(J \sum_{k=1}^{J} \log^2 M_k)$, which is not greater than $\mathcal{O}(J^2 \log^2 N)$. On this step, computational costs of calculation of the values $B$ and $D$ are bounded by $\mathcal{O}(qN)$ and $\mathcal{O}(1)$ respectively.

The time complexity required to find an exact solution for the auxiliary Problem 2 (see Section 4) is equal to $\mathcal{O}(N(N^2+qJ))$. On step 2, this problem is solved $\mathcal{O}(|\mathcal{D}|^J)$ times. Therefore, according to (7), the complexity of step 2 is equal to $\mathcal{O}(N(N^2+qJ)(2MB+1)^{qJ})$.

Summing up all the costs on all the steps, we find that the time complexity of the algorithm is bounded by $\mathcal{O}(N(N^2+qJ)(2MB+1)^{qJ}+J^2\,log^2N)$. □

*Remark 1.* If the dimension $q$ of the space and the number $J$ of desired subsets are fixed, the coordinates of the input points from $\mathcal{Y}$ are integer and belong to an interval $[-B,B]$, then algorithm $\mathcal{A}$ is an exact pseudopolynomial algorithm solving Problem 1 in $\mathcal{O}(N^3(MB)^{qJ})$ time.

## 5 Conclusion

In the paper, we have considered the strongly NP-hard problem of finding a family of disjoint subsets in a finite set of points from Euclidean space.

For the special case of the problem when the input points coordinates are integer, we have constructed an exact algorithm. This algorithm is a pseudopolynomial one when the input space dimension and the number of required subsets are fixed.

Of considerable interest is justification of faster approximation polynomial-time algorithms with guaranteed accuracy for general problem.

**Acknowledgments.** This work was supported by the Russian Foundation for Basic Research (project nos. 15-01-00462, and 16-07-00168) and by the grant of Presidium of RAS (program 5, project 227).

## References

1. Hastie T., Tibshirani R., Friedman J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Springer-Verlag, New York (2009)
2. James G., Witten D., Hastie T., Tibshirani R.: An Introduction to Statistical Learning. Springer Science+Business Media, LLC, New York (2013)
3. Bishop C.M.: Pattern Recognition and Machine Learning. Springer Science+Business Media, LLC, New York (2006)
4. Flach P.: Machine Learning: The Art and Science of Algorithms that Make Sense of Data. Cambridge University Press, New York (2012)
5. Galashov A.E., Kelmanov A.V.: A 2-Approximate Algorithm to Solve One Problem of the Family of Disjoint Vector Subsets. J. Automation and Remote Control. 75(4), 595–606 (2014)
6. Aloise D., Deshpande A., Hansen P., Popat P.: NP-hardness of Euclidean sum-of-squares clustering. J. Machine Learning. 75(2), 245–248 (2009)
7. Jain A.K.: Data Clustering: 50 Years Beyond $k$-Means. J. Pattern Recognition Lett. 31, 651–666 (2010)
8. Edwards A.W.F., Cavalli-Sforza L.L.: A Method for Cluster Analysis. J. Biometrics. 21, 362–375, (1965)

9. MacQueen J.B.: Some Methods for Classification and Analysis of Multivariate Observations. In: Fith Berkley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 281 – 297. University of California Press, Berkley, California (1967)
10. Bagdonavicius, V.,Kruopis, J., Nikulin, M.S.: Non-parametric Tests for Censored Data. ISTE/WILEY, London, (2011)
11. Kel'manov A.V., Pyatkin A.V.: NP-Completeness of Some Problems of Choosing a Vector Subset. J. Applied and Industrial Mathematics. 5(3), 352–357, (2011)
12. Kel'manov A.V., Romanchenko S.M.: An Approximation Algorithm for Solving a Problem of Search for a Vector Subset. J. Applied and Industrial Mathematics. 6(1), 90–96, (2012)
13. Kel'manov A.V., Romanchenko S.M.: Pseudopolynomial Algorithms for Certain Computationally Hard Vector Subset and Cluster Analysis Problems. J. Automation and Remote Control. 73(2), 349–354, (2012)
14. Kel'manov A.V., Romanchenko S.M.: An FPTAS for a Vector Subset Search Problem. J. Applied and Industrial Mathematics. 8(3), 329–336, (2014)
15. Schenmaier V.V.: An approximation scheme for a problem of search for a vector subset. J. Applied and Industrial Mathematics. 6(3), 381–386, (2012)
16. Papadimitriou C. H., Steiglitz K.: Combinatorial Optimization: Algorithms and Complexity. Prentice-Hall, Englewood Cliffs, New Jersey (1982)