On Some Euclidean Clustering Problems: NP-Hardness and Efficient Approximation Algorithms

Alexander Kel'manov
Sobolev Institute of Mathematics
Acad. Koptyug avenue, 4,
630090 Novosibirsk, Russia
Novosibirsk State University
Pirogova str. 1,
630090 Novosibirsk, Russia.
kelm@math.nsc.ru

Abstract

We consider some poorly studied clustering problems. The paper purpose is to present a short survey on some new results on the computational complexity of these problems, and on efficient algorithms with performance guarantees for their solutions.

1 Introduction

The subject of this study is several related discrete optimization problems of partitioning (clustering) a finite set and a finite sequence of Euclidean points into a family of clusters. Our goal is to review some recent (2016-2017) results of the author together with his colleagues and students (from Sobolev Institute of Mathematics and Novosibirsk State University) on these problems. We present the results on the computational complexity of the problems under consideration and on the efficient approximation algorithms for their solution. Our research is motivated by the insufficient studies of the problems and by their importance, in particular, for computer geometry, statistics, physics, mathematical problems of clustering, pattern recognition, machine learning, data mining.

2 Problems Formulation

Throughout this paper, $\mathbb{R}$ denotes the set of real numbers and $\|\cdot\|$ is the Euclidean norm. The following problems are considered.

Problem 1 (*Subset of points with the largest cardinality under constraint on the total quadratic variation*). *Given*: a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of points from $\mathbb{R}^q$ and number $\alpha \in (0, 1)$.
Find: a subset $\mathcal{C} \subset \mathcal{Y}$ with the largest cardinality such that

$$\sum_{y \in \mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 \leq \alpha \sum_{y \in \mathcal{Y}} \|y - \overline{y}(\mathcal{Y})\|^2,$$

In: Yu. G. Evtushenko, M. Yu. Khachay, O. V. Khamisov, Yu. A. Kochetov, V.U. Malkova, M.A. Posypkin (eds.): Proceedings of the OPTIMA-2017 Conference, Petrovac, Montenegro, 02-Oct-2017, published at http://ceur-ws.org

where $\overline{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|}\sum_{y\in\mathcal{C}} y$ is the centroid (the geometrical center) of subset $\mathcal{C}$, and $\overline{y}(\mathcal{Y}) = \frac{1}{|\mathcal{Y}|}\sum_{y\in\mathcal{Y}} y$ is the centroid of the input set.

The strong NP-hardness of Problem 1 was proved in [Ageev et al., 2017] and in the same paper a polynomial-time 1/2-approximation algorithm with $\mathcal{O}(N^2(q+\log N))$ running time was proposed.

These results supplement and develop the results obtained in [Aggarwal et al., 1991], [Kel'manov & Pyatkin, 2011], [Shenmaier, 2016], [Kel'manov & Romanchenko, 2011], [Shenmaier, 2012], [Kel'manov & Romanchenko, 2012], [Kel'manov & Romanchenko, 2014] for M-Variance problem. In this problem we need to find a subset $\mathcal{C}$ with given cardinality in N-element set $\mathcal{Y}$ of points minimizing the value of $\sum_{y\in\mathcal{C}}\|y-\overline{y}(\mathcal{C})\|^2$.

Problem 2 (*Subset of vectors with the largest cardinality under constraint on normalized length of vectors sum*).

Given: a set $\mathcal{Y} = \{y_1,\ldots,y_N\}$ of points (vectors) from $\mathbb{R}^q$ and a number $\alpha\in(0,1)$.

Find: a subset $\mathcal{C}\subset\mathcal{Y}$ with the largest cardinality such that

$$\frac{1}{|\mathcal{C}|}\Big\|\sum_{y\in\mathcal{C}} y\Big\|^2 \le \alpha\frac{1}{|\mathcal{Y}|}\Big\|\sum_{y\in\mathcal{Y}} y\Big\|^2.$$

In [Eremeev et al., 2017], it was shown that Problem 2 is NP-hard even in terms of finding a feasible solution. An exact algorithm for the case of integer components of the input vectors is proposed in the same paper. The algorithm has a pseudo-polynomial time complexity if the dimension q of the space is bounded from above by a constant.

These results supplement and develop the results obtained in [Eremeev et al, 2016] for another subset searching problem. In this problem we have to find a subset $\mathcal{C}\subseteq\mathcal{Y}$ minimizing the value of $\frac{1}{|\mathcal{C}|}\|\sum_{y\in\mathcal{C}} y\|^2$.

Problem 3 (*Cardinality-weighted variance-based 2-clustering with given center*).

Given: a set $\mathcal{Y} = \{y_1,\ldots,y_N\}$ of points from $\mathbb{R}^q$ and a positive integer M.

Find: a partition of $\mathcal{Y}$ into two clusters $\mathcal{C}$ and $\mathcal{Y}\setminus\mathcal{C}$ such that

$$|\mathcal{C}|\sum_{y\in\mathcal{C}}\|y-\overline{y}(\mathcal{C})\|^2 + |\mathcal{Y}\setminus\mathcal{C}|\sum_{y\in\mathcal{Y}\setminus\mathcal{C}}\|y\|^2 \longrightarrow \min$$

subject to constraint $|\mathcal{C}| = M$.

The strong NP-hardness of Problem 3 was proved in [Kel'manov & Pyatkin, 2015], [Kel'manov & Pyatkin, 2016]. In the same papers, the nonexistence of an FPTAS (Fully Polynomial-Time Approximation Scheme) was shown (unless P=NP) for this problem.

In [Kel'manov & Motkova, 2016], an exact algorithm for the case of integer components of the input points was presented. If the dimension q of the space is bounded by a constant, then this algorithm has a pseudopolynomial complexity and the running time of the algorithm is $\mathcal{O}(N(MB)^q)$, where B is the maximum absolute coordinate value of the points.

An approximation algorithm that allows to find a $(1+\varepsilon)$-approximate solution in $\mathcal{O}(qN^2(\sqrt{\frac{2q}{\varepsilon}}+2)^q)$ time for a given relative error ε was proposed in [Kel'manov & Motkova, 2016]. If the space dimension is bounded by a constant, then this algorithm implements an FPTAS with $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$ running time.

Following Problem 4 is a generalization of Problem 3.

Problem 4 (*Weighted variance-based 2-clustering with given center*).

Given: an N-element set $\mathcal{Y}$ of points from $\mathbb{R}^q$, a positive integer $M\le N$, and real numbers (weights) $\omega_1>0$ and $\omega_2\ge 0$.

Find: a partition of $\mathcal{Y}$ into two clusters $\mathcal{C}$ and $\mathcal{Y}\backslash\mathcal{C}$ minimizing the value of

$$w_1\sum_{y\in\mathcal{C}}\|y-\overline{y}(\mathcal{C})\|^2 + w_2\sum_{y\in\mathcal{Y}\backslash\mathcal{C}}\|y\|^2$$

subject to constraint $|\mathcal{C}| = M$.

In [Kel'manov et al., 2017], an approximation algorithm is constructed. It allows to find a $(1+\varepsilon)$-approximate solution of the problem for an arbitrary $\varepsilon\in(0,1)$ in $\mathcal{O}\left(qN^2(\sqrt{\frac{2q}{\varepsilon}}+2)^q\right)$ time. Moreover, in the same paper, the modification of this algorithm with improved running time of $\mathcal{O}\Big(\sqrt{q}N^2\big(\frac{\pi e}{2}\big)^{q/2}\big(\frac{1}{\sqrt{\varepsilon}}+2\big)^q\Big)$ was proposed. The algorithm implements an FPTAS for the fixed space dimension. If the dimension of space is bounded by $C\log n$, where C is a positive constant, the algorithm remains polynomial and implements a PTAS (Polynomial-Time Approximation Scheme) with $\mathcal{O}\left(N^{C\,(1.05+\log(2+\frac{1}{\sqrt{\varepsilon}}))}\right)$ running time.

These results supplement and develop the results obtained in [Kel'manov & Pyatkin, 2015], [Kel'manov & Pyatkin, 2016], [Kel'manov & Motkova, 2016], [Kel'manov & Motkova, 2016] for Problem 3.

The complexity status of following Problems 5–9 seemed to be unclear up to now.

Problem 5 (*Normalized K-means clustering*).

Given: a set $\mathcal{Y}$ of N points in $\mathbb{R}^q$ and a positive ineger $K \geq 2$.

Find: a partition of $\mathcal{Y}$ into clusters $\mathcal{C}_1, \ldots, \mathcal{C}_K$ minimizing the value of

$$\sum_{k=1}^{K} \frac{1}{|\mathcal{C}_k| - 1} \sum_{y \in \mathcal{C}_k} \|y - \overline{y}(\mathcal{C}_k)\|^2,$$

where $\overline{y}(\mathcal{C}_k)$ is a centroid of cluster $\mathcal{C}_k$.

Problem 6 (*Normalized K-Means clustering with a given center*).

Given: a set $\mathcal{Y}$ of N points in $\mathbb{R}^d$ and a positive ineger $K \geq 2$.

Find: a partition of $\mathcal{Y}$ into clusters $\mathcal{C}_1, \ldots, \mathcal{C}_K$ minimizing the value of

$$\sum_{k=1}^{K-1} \frac{1}{|\mathcal{C}_k| - 1} \sum_{y \in \mathcal{C}_k} \|y - \overline{y}(\mathcal{C}_k)\|^2 + \frac{1}{|\mathcal{C}_K| - 1} \sum_{y \in \mathcal{C}_K} \|y\|^2,$$

where $\overline{y}(\mathcal{C}_k)$ is a centroid of cluster $\mathcal{C}_k$.

In [Ageev, 2017], it was proved that Problem 5 is strongly NP-hard for each fixed $K \geq 3$ and Problem 6 is strongly NP-hard for each fixed $K \geq 4$.

Problem 7 (*Minimum sum of normalized squares of norms clustering*).

Given: a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of points from $\mathbb{R}^q$ and a positive integer $J > 1$.

Find: a partition of $\mathcal{Y}$ into nonempty clusters $\mathcal{C}_1, \ldots, \mathcal{C}_J$ such that

$$\sum_{j=1}^{J} \frac{1}{|\mathcal{C}_j|} \Big\| \sum_{y \in \mathcal{C}_j} y \Big\|^2 \longrightarrow \min.$$

In this problem the criterion is minimizing the sum over all clusters of normalized by the cardinality squared norms of the sum of cluster elements.

Problem 8 (*Minimum sum of squared norms clustering*).

Given: a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of points from $\mathbb{R}^q$ and a positive integer $J > 1$.

Find: a partition of $\mathcal{Y}$ into nonempty clusters $\mathcal{C}_1, \ldots, \mathcal{C}_J$ such that

$$\sum_{j=1}^{J} \Big\| \sum_{y \in \mathcal{C}_j} y \Big\|^2 \longrightarrow \min.$$

In this problem the criterion is minimizing the sum over all clusters of squared norms of the sum of cluster elements.

Problem 9 (*Minimum sum-of-norms clustering*).

Given: a set $\mathcal{Y} = \{y_1, \ldots, y_N\}$ of points from $\mathbb{R}^q$ and a positive integer $J > 1$.

Find: a partition of $\mathcal{Y}$ into nonempty clusters $\mathcal{C}_1, \ldots, \mathcal{C}_J$ such that

$$\sum_{j=1}^{J} \Big\| \sum_{y \in \mathcal{C}_j} y \Big\| \longrightarrow \min.$$

In this problem the criterion is minimizing the sum over all clusters of norms of the sum of cluster elements.

It is proved [Kel'manov & Pyatkin, 2016] that problems 7–9 are strongly NP-hard if the number of clusters is a part of the input, and NP-hard in the ordinary sense if the number of clusters is not a part of the input (is fixed). Moreover, the problems are NP-hard even in the case of dimension 1 (on a line).

These results supplement and develop the results obtained in [Kel'manov & Pyatkin, 2016], [Eremeev et al, 2016], [Kel'manov & Pyatkin, 2009].

Problem 10 (*Finding a subsequence in a sequence*)

Given: a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of points from $\mathbb{R}^q$ and positive integer numbers $T_{\min}$, $T_{\max}$ and $M > 1$.

Find: a tuple $\mathcal{M} = (n_1, \dots, n_M)$, where $n_m \in \mathcal{N} = \{1, \dots, N\}, m = 1, \dots, N$, such that

$$\sum_{j \in \mathcal{M}} \|y_j - \overline{y}(\mathcal{M})\|^2 \to \min,$$

where $\overline{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$ is a geometric center (centroid) of the subsequence $\{y_i \in \mathcal{Y} \,|\, i \in \mathcal{M}\}$ subject to constraints

$$T_{\min} \le n_m - n_{m-1} \le T_{\max} \le N, \; m = 2, \dots, M, \tag{1}$$

on the elements of the tuple $(n_1, \dots, n_M)$.

Problem 10 is among poorly studied strongly NP-hard discrete optimization problems. By this time for Problem 10 the following results were obtained. First, we should note that there are neither exact polynomial-time, nor pseudo-polynomial-time algorithms or FPTAS schemes, unless P=NP.

The case of Problem 10 when $T_{\min}$ and $T_{\max}$ are parameters is analyzed in [Kel'manov & Pyatkin, 2013]. In this work the authors showed that this problem is strongly NP-hard for any $T_{\min} < T_{\max}$. In the trivial case when $T_{\min} = T_{\max}$ this problem can be solved through a polynomial time.

In [Kel'manov et al., 2012] a 2-approximation polynomial-time algorithm is proposed; the running time of the algorithm is $\mathcal{O}(N^2(MN + q))$.

In the case of Problem 10 with integer components of the sequence elements and fixed dimension q of the space in [Kel'manov et al., 2013] an exact pseudo-polynomial-time algorithm is substantiated. This algorithm finds an optimal solution of Problem 10 in $\mathcal{O}(N^3(MD)^q)$ time.

The main result of [Kel'manov et al., 2016] is an approximation algorithm which allows to find a $(1 + \varepsilon)$–approximate solution for any $\varepsilon > 0$ in $\mathcal{O}(N^2(M(T_{\max} - T_{\min} + 1) + q)(\sqrt{\frac{2q}{\varepsilon}} + 2)^q)$ time. If the dimension q of the space is fixed then the time complexity of this algorithm is equal to $\mathcal{O}(MN^3(1/\varepsilon)^{q/2})$, and it implements an FPTAS.

Problem 11 (*Minimum sum-of-squares* 2*-clustering problem on sequence with given center of one cluster*).

Given: a sequence $\mathcal{Y} = (y_1, \dots, y_N)$ of points from $\mathbb{R}^q$, and some positive integer numbers $T_{\min}$, $T_{\max}$, and M.

Find: a tuple $\mathcal{M} = (n_1, \dots, n_M)$, where $n_m \in \mathcal{N} = \{1, \dots, N\}, m = 1, \dots, N$, such that

$$\sum_{j \in \mathcal{M}} \|y_j - \overline{y}(\mathcal{M})\|^2 + \sum_{j \in \mathcal{N} \setminus \mathcal{M}} \|y_j\|^2 \to \min,$$

where $\overline{y}(\mathcal{M}) = \frac{1}{M} \sum_{i \in \mathcal{M}} y_i$, under constraints (1) on the elements of $(n_1, \dots, n_M)$.

A special case of Problem 11 where $T_{\min} = 1$ and $T_{\max} = N$ is equivalent [Kel'manov & Pyatkin, 2013] to the strongly NP-hard problem of partitioning a set [Kel'manov & Pyatkin, 2009], which does not admit [Kel'manov & Khandeev, 2016] FPTAS unless P = NP. In other words, Problem 11 of partitioning a sequence is the generalization of the strongly NP-hard problem of partitioning a set. Therefore, according to [Garey & Johnson, 1979], Problem 11 also admits neither exact polynomial, nor exact pseudopolynomial algorithms, nor FPTAS unless P = NP.

In [Kel'manov & Pyatkin, 2013], the variant of Problem 11 in which $T_{\min}$ and $T_{\max}$ are the parameters was analyzed. In the cited work it was shown that Problem 11 is strongly NP-hard for any $T_{\min} < T_{\max}$. In the trivial case when $T_{\min} = T_{\max}$, the problem is solvable in polynomial time.

A 2-approximation algorithm for Problem 11 with $\mathcal{O}(N^2(MN + q))$ running time was presented in [Kel'manov & Khamidullin, 2014].

Special cases of the problem were studied in [Kel'manov et al., 2017a], [Kel'manov et al., 2016]. In [Kel'manov et al., 2017a], for the case of integer inputs and fixed space dimension q, an exact pseudopolynomial algorithm was proposed. The running time of the algorithm is $\mathcal{O}(N^3(MD)^q)$, where D is the maximal absolute value of the coordinates of input points. For the case with fixed space dimension in [Kel'manov et al., 2016] an FPTAS was constructed which, given a relative error ε, finds a $(1 + \varepsilon)$-approximate solution of Problem 11 in $\mathcal{O}(MN^3(1/\varepsilon)^{q/2})$ time.

The new result [Kel'manov et al., 2017b] is a randomized algorithm for Problem 11. For an established parameter value, given $\varepsilon > 0$ and fixed $\gamma \in (0, 1)$, this algorithm allows to find a $(1 + \varepsilon)$-approximate solution of the problem with a probability of at least $1-\gamma$ in $\mathcal{O}(qMN^2)$ time. The conditions are established under which the algorithm is asymptotically exact and its time complexity is $\mathcal{O}(qMN^3)$.

Problem 12 (*Minimum sum-of-squares clustering problem on sequence with given center of one cluster and cluster cardinalities*).

Given: a sequence $\mathcal{Y} = (y_1, \ldots, y_N)$ of points from $\mathbb{R}^q$ and some positive integers $T_{\min}$, $T_{\max}$, L, and M.

Find: nonempty disjoint subsets $\mathcal{M}_1, \ldots, \mathcal{M}_L$ of $\mathcal{N} = \{1, \ldots, N\}$ such that

$$\sum_{l=1}^{L} \sum_{j \in \mathcal{M}_l} \|y_j - \overline{y}(\mathcal{M}_l)\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|y_i\|^2 \to \min, \tag{2}$$

where $\mathcal{M} = \cup_{l=1}^{L} \mathcal{M}_l$, and $\overline{y}(\mathcal{M}_l) = \frac{1}{|\mathcal{M}_l|} \sum_{j \in \mathcal{M}_l} y_j$ is the centroid of subset $\{y_j | \, j \in \mathcal{M}_l\}$, under the following constraints:

(1) the cardinality of $\mathcal{M}$ is equal to M,

(2) concatenation of elements of subsets $\mathcal{M}_1, \ldots, \mathcal{M}_L$ is an increasing sequence, provided that the elements of each subset are in ascending order,

(3) the inequalities (1) for the elements of $\mathcal{M} = \{n_1, \ldots, n_M\}$ are satisfied.

Problem 13 (*Minimum sum-of-squares clustering problem on sequence with given center of one cluster*).

Given: a sequence $\mathcal{Y} = (y_1, \ldots, y_N)$ of points from $\mathbb{R}^q$ and some positive integers $T_{\min}$, $T_{\max}$, and L.

Find: nonempty disjoint subsets $\mathcal{M}_1, \ldots, \mathcal{M}_L$ of $\mathcal{N} = \{1, \ldots, N\}$ minimizing (2) under the following constraints:

(1) concatenation of elements of subsets $\mathcal{M}_1, \ldots, \mathcal{M}_L$ is an increasing sequence, provided that the elements of each subset are in ascending order,

(2) the inequalities (1) for the elements of $\mathcal{M} = \{n_1, \ldots, n_M\}$ are satisfied (the cardinality of $\mathcal{M}$ assumed to be unknown).

In [Kel'manov & Pyatkin, 2013], the variants of Problems 12, 13 in which $T_{\min}$ and $T_{\max}$ are the parameters was analyzed. In the cited work it was shown that both parameterized variants of the problems are strongly NP-hard for any $T_{\min} < T_{\max}$. In the trivial case when $T_{\min} = T_{\max}$, the problems are solvable in polynomial time.

Except for the case with $L = 1$ in equation (2), no algorithms with guaranteed approximation factor were known up to 2016 for Problem 11. This case is equivalent to Problem 12. For this problem, the existing results are listed above.

The new result of [Kel'manov et al., 2016] is an algorithm that allows to find a 2-approximate solution of the problem in $\mathcal{O}(LN^{L+1}(MN+q))$ time, which is polynomial if the number L of clusters with unknown centroids is fixed.

For Problem 13, except for the case with $L = 1$ in equation (2), no algorithms with guaranteed approximation factor were known up to 2017. For this special case, in [Kel'manov & Khamidullin, 2015], a 2-approximation polynomial-time algorithm with $\mathcal{O}(N^2(N+q))$ running time was presented.

The new result of [Kel'manov et al., 2017c] is an algorithm that allows to find a 2-approximate solution of the problem in $\mathcal{O}(LN^{L+1}(N+q))$ time, which is polynomial if the number L of clusters with unknown centroids is fixed.

Acknowledgements

The study of problems 1–3, 5, 6, 10, 11 and 12 was supported by the the Russian Foundation for Basic Research (projects 15-01-00462, 16-31-00186, 16-07-00168), and by the Ministry of Science and Education of the Russian Federation under the 5-100 Excellence Programme. The study of problems 4, 7–9 and 13 was supported by the Russian Science Foundation (project 16-11-10041).

References

[Ageev et al., 2017] Ageev A.A., Kel'manov A.V., Pyatkin A.V., Khamidullin S.A., and Shenmaier V.V. (2017). Approximation Polynomial Algorithm for the Data Editing and Data Cleaning Problem. *Pattern Recognition and Image Analysis, 17(3)*, (accepted).

[Aggarwal et al., 1991] Aggarwal A., Imai H., Katoh N., Suri S. (1991). Finding k Points With Minimum Diameter and Related Problems. *J. Algorithms. 12(1)*, 38–56. doi.org/10.1016/0196-6774(91)90022-Q

[Kel'manov & Pyatkin, 2011] Kel'manov A. V., Pyatkin A. V. (2011). NP-Completeness of Some Problems of Choosing a Vector Subset. *J. Appl. Indust. Math. 5(3)*, 352–357. doi: 10.1134/S1990478911030069

[Shenmaier, 2016] Shenmaier V. V. (2016). Solving Some Vector Subset Problems by Voronoi Diagrams. *J. Appl. Indust. Math. 10(2)*, 550–566. doi: 10.1134/S199047891604013X

[Kel'manov & Romanchenko, 2011] Kel'manov A. V., Romanchenko S. M. (2012). An Approximation Algorithm for Solving a Problem of Search for a Vector Subset. *J. Appl. Indust. Math. 6(1)*, 90–96. doi: 10.1134/S1990478912010097

[Shenmaier, 2012] Shenmaier V. V. (2012). An Approximation Scheme for a Problem of Search for a Vector Subset. *J. Appl. Indust. Math. 6(3)*, 381–386. doi: 10.1134/S1990478912030131

[Kel'manov & Romanchenko, 2012] Kel'manov A.V., Romanchenko S. M. (2012). Pseudopolynomial Algorithms for Certain Computationally Hard Vector Subset and Cluster Analysis Problems. *Automation and Remote Control. 73(2)*, 349–354. doi: 10.1134/S0005117912020129

[Kel'manov & Romanchenko, 2014] Kel'manov A.V., Romanchenko S. M. (2014). An FPTAS for a Vector Subset Search Problem. *J. Appl. Indust. Math. 8(3)*, 329–336. doi: 10.1134/S1990478914030041

[Eremeev et al., 2017] Eremeev A.V., Kel'manov A.V., Pyatkin A.V., Ziegler I.A. (2017). Maximum Cardinality Subset of Vectors with a Constraint on Normalized Squared Length of Vectors Sum. Proc. of the 6th International Conference on Analysis of Images, Social Networks, and Texts (AIST 2017), Moscow, Russia, July 27-29, 2017. *Lecture Notes in Computer Science.*(accepted).

[Eremeev et al, 2016] Eremeev A.V., Kel'manov A.V., Pyatkin A.V. (2016). On the Complexity of Some Euclidean Optimal Summing Problems. *Doklady Mathematics, 93(3)*, 286–288. doi: 10.1134/S1064562416030157

[Kel'manov & Pyatkin, 2015] Kel'manov A.V., Pyatkin A.V. (2015). NP-Hardness of Some Quadratic Euclidean 2-clustering Problems. *Doklady Mathematics. 92(2)*, 634–637. doi: 10.1134/S1064562415050233

[Kel'manov & Pyatkin, 2016] Kel'manov A.V., Pyatkin A.V. (2016). On the Complexity of Some Quadratic Euclidean 2-Clustering Problems. *Comput. Math. Math. Phys. 56(3)*, 491–497. doi: 10.1134/S096554251603009X

[Kel'manov & Motkova, 2016] Kel'manov A.V., Motkova A.V. (2016). Exact Pseudopolynomial Algorithm for a Balanced 2-Clustering Problem. *J. Appl. Indust. Math. 10(3)*, 349–355. doi: 10.1134/S1990478916030054

[Kel'manov & Motkova, 2016] Kel'manov A.V., Motkova A.V. (2016). A Fully Polynomial-Time Approximation Scheme for a Special Case of a Balanced 2-Clustering Problem. *Lecture Notes in Computer Science. 9869.* 182–192. doi: 10.1007/978-3-319-44914-2-15

[Kel'manov et al., 2017] Kelmanov A.V., Motkova A.V., Shenmaier V.V. (2017). Polynomial-Time Approximation Scheme for a Problem of Partitioning a Finite Set into Two Clusters. Proceedings of the Steklov Institute of Mathematics (accepted).

[Ageev, 2017] Ageev A.A. (2017). Complexity of Normalized K-Means Clustering Problems. Proc. of the 17th Baikal International Triannual School-Seminar Methods of Optimization and Their Applications (MOPT-2017), Maksimikha, Buryatia, Russia, July 26 – Aug 6, 2017. (accepted).

[Kel'manov & Pyatkin, 2016] Kel'manov A.V., Pyatkin A.V. (2016). On the Complexity of Some Euclidean Problems of Partitioning a Finite Set of Points. *Doklady Mathematics, 94(3)*, 635–638. doi: 10.1134/S1064562416060089

[Kel'manov & Pyatkin, 2016] Kel'manov A.V., Pyatkin A.V. (2016). On the Complexity of Some Quadratic Euclidean 2-Clustering Problems. *Comput. Math. Math. Phys. 56(3)*, 491–497. doi: 10.1134/S096554251603009X

[Kel'manov & Pyatkin, 2013] Kel'manov, A.V., Pyatkin, A.V. (2013). On Complexity of Some Problems of Cluster Analysis of Vector Sequences. *J. Appl. Indust. Math. 7(3)*, 363–369. doi: 10.1134/S1990478913030095

[Kel'manov et al., 2012] Kel'manov A.V., Romanchenko S.M., Khamidullin S.A. (2012). Approximation Algorithms for Some Intractable Problems of Choosing a Vector Subsequence. *J. Appl. Indust. Math. 6(4)*, 443–450. doi: 10.1134/S1990478912040059

[Kel'manov et al., 2013] Kel'manov A.V., Romanchenko S.M., Khamidullin S.A. (2013). Exact Pseudopolynomial Algorithms for Some NP-Hard Problems of Searching a Vectors Subsequence. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki (in Russian). 53(1)*, 143–153. doi: 10.7868/S0044466913010055

[Kel'manov et al., 2016] Kel'manov A.V., Romanchenko S.M., Khamidullin S.A. (2016). Fully Polynomial-time Approximation Scheme for a Problem of Finding a Subsequence. 9th International Conference on Discrete Optimization and Operations Research, DOOR 2016; Vladivostok; Russian Federation; September 19-23, *CEUR Workshop Proceedings. Vol. 1623*, 516–525.

[Kel'manov & Pyatkin, 2009] Kel'manov A.V., Pyatkin A.V. (2009). Complexity of Certain Problems of Searching for Subsets of Vectors and Cluster Analysis. *Comput. Math. Math. Phys. 49(11)*, 1966–1971. doi: 10.1134/S0965542509110128

[Kel'manov & Khandeev, 2016] Kel'manov A.V., Khandeev V.I. (2016). Fully Polynomial-time Approximation Scheme for a Special Case of a Quadratic Euclidean 2-Clustering Problem. *Comput. Math. Math. Phys. 56(2)*, 334–341. doi: 10.1134/S0965542516020111

[Garey & Johnson, 1979] Garey M.R., Johnson D.S. (1979). Computers and Intractability: A Guide to the Theory of NP-Completeness. San Francisco: Freeman.

[Kel'manov & Khamidullin, 2014] Kelmanov, A.V., Khamidullin, S.A. (2014). An Approximating Polynomial Algorithm for a Sequence Partitioning Problem. *J. Appl. Indust. Math. 8(2)*, 236–244. doi: 10.1134/S1990478914020100

[Kel'manov et al., 2017a] Kelmanov, A.V., Khamidullin, S.A., Khandeev, V.I. (2017). Exact Pseudopolynomial Algorithm for One Sequence Partitioning Problem. *Automation and Remote Control. 78(1)*, 67–74. doi: 10.1134/S0005117917010052

[Kel'manov et al., 2016] Kelmanov, A.V., Khamidullin, S.A., Khandeev, V.I. (2016). A Fully Polynomial-time Approximation Scheme for a Sequence 2-Cluster Partitioning Problem. *J. Appl. Indust. Math. 10(2)*, 209–219. doi: 10.1134/S199047891602006X

[Kel'manov et al., 2017b] Kelmanov, A.V., Khamidullin, S.A., Khandeev, V.I. (2017). A Randomized Algorithm for 2-Partition of a Sequence. *Lecture Notes in Computer Science*. Proc. of the 6th International Conference on Analysis of Images, Social Networks, and Texts (AIST 2017), Moscow, Russia, July 27–29. *Lecture Notes in Computer Science.* (accepted).

[Kel'manov et al., 2016] Kel'manov A.V., Mikhailova L.V., Khamidullin S.A., Khandeev V.I. (2016). An Approximation Algorithm for a Problem of Partitioning a Sequence into Clusters with Restrictions on Their Cardinalities. *Lecture Notes in Computer Science. 9869.* 182–192. doi: 10.1007/978-3-319-44914-2-14

[Kel'manov et al., 2017c] Kel'manov A.V., Mikhailova L.V., Khamidullin S.A., Khandeev V.I. (2017). An Approximation Algorithm for a Problem of Partitioning a Sequence into Clusters. *Zhurnal Vychislitelnoi Matematiki i Matematicheskoi Fiziki. 57(8)*, 149–157. (in Russian) doi: 10.7868/S0044466917080087.

[Kel'manov & Khamidullin, 2015] Kel'manov A.V., Khamidullin S.A. (2015). An Approximation Polynomial-Time Algorithm for a Sequence Bi-Clustering Problem. *Comput. Math. Math. Phys. 55(6)*, 1068–1076. doi: 10.1134/S0965542515060068