

Algorithms with Performance Guarantee for Some Quadratic Euclidean Problems of 2-Partitioning a Set and a Sequence

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

Vladimir Khandeev
Sobolev Institute of Mathematics
Acad. Koptug avenue 4,
Novosibirsk State University
Pirogova str. 1,
630090 Novosibirsk, Russia
khandeev@math.nsc.ru

Abstract

We consider the problems of 2-partitioning a finite set and a finite sequence of points in Euclidean space into clusters (subsets or subsequences) minimizing the sum of squared distances between cluster elements and the corresponding cluster centers. It is assumed that the center of one of the desired clusters is the origin, while the center of the other cluster is unknown and determined as the mean value over cluster elements. In this work, we present a short survey on some results for these problems and a new result: a randomized algorithm for the sequence partitioning problem.

1 Introduction

In this work, we consider two strongly NP-hard discrete optimization problems. The goal of our work is to present a short survey on some recent results for these problems.

The problems under consideration are closely related to the well-known strongly NP-hard 2-MSSC (Minimum Sum-of-Squares Clustering) problem [Aloise et al., 2009]. This problem (also known as 2-Means) is to find a partition of a finite set \mathcal{Y} of points from a Euclidean space into two subsets (clusters) \mathcal{C} and $\mathcal{Y} \setminus \mathcal{C}$ such that

$$\sum_{y \in \mathcal{C}} \|y - \bar{y}(\mathcal{C})\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y - \bar{y}(\mathcal{Y} \setminus \mathcal{C})\|^2 \rightarrow \min,$$

where $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ and $\bar{y}(\mathcal{Y} \setminus \mathcal{C}) = \frac{1}{|\mathcal{Y}| - |\mathcal{C}|} \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} y$ are the centroids (the geometric centers) of sets \mathcal{C} and $\mathcal{Y} \setminus \mathcal{C}$, respectively.

The example of an input set of points on a plane is presented at Fig. 1. One needs to find a partition of this set into two clusters and find their centroids. It is assumed that the points are scattered around these centroids (the points of different clusters are colored differently).

Copyright © by the paper's authors. Copying permitted for private and academic purposes.

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>

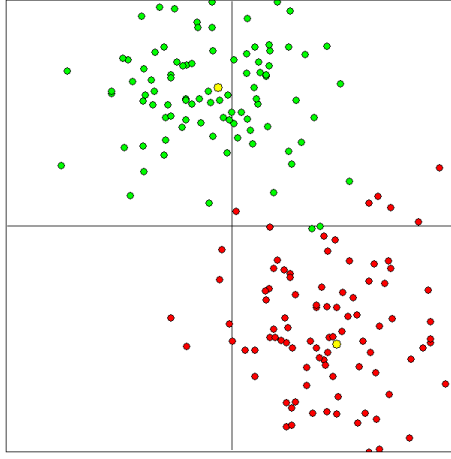


Figure 1: Example of the input set

The main difference of the considered problems from the 2-MSSC problem is that the center of one the clusters is fixed at a certain point in the Euclidean space (see the following sections); without loss of generality we can assume that this point is the origin. Additionally, in one of the problems under consideration the input is a sequence rather than a set, and there are constraints on the indices of elements of subsequences that are included in one cluster or another. Despite the similarities, the considered problems and the 2-MSSC problem are not equivalent. Therefore, the both considered problems need an individual study. Also, the problems under consideration are important for applications, e.g. for approximation theory, statistics, data mining, the remote sensing, geophysics, biometrics, medicine and technical diagnostics, processing the speech signals, electronic intelligence, radiolocation, hydroacoustics, etc.

2 Partitioning a Set

Problem 1.

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

Find: a subset $\mathcal{C} \subseteq \mathcal{Y}$ that minimizes the objective function

$$S(\mathcal{C}) = \sum_{y \in \mathcal{C}} \|y - \bar{y}(\mathcal{C})\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{C}} \|y\|^2,$$

where $\bar{y}(\mathcal{C}) = \frac{1}{|\mathcal{C}|} \sum_{y \in \mathcal{C}} y$ is the centroid (the geometric center) of set \mathcal{C} , under constrain $|\mathcal{C}| = M$.

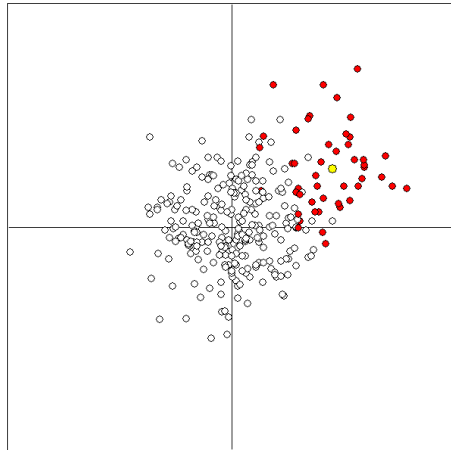


Figure 2: Example of the input set

The example of an input set of points on a plane is presented at Fig. 2. One needs to find a partition of this set into two clusters. Unlike 2-MSSC problem, in Problem 1 one needs to find the centroid of only one cluster. It is assumed that the points of this cluster (shown by dark color) are scattered around its unknown centroid and that the points of the second one (shown by white color) are scattered around the origin.

The strong NP-hardness of Problem 1 is shown in [Kel'manov & Pyatkin, 2009]. In the case of fixed space dimension the problem is solvable in polynomial time [Gimadi et al., 2010, Shenmaier, 2016].

A 2-approximation algorithm for Problem 1 is presented in [Dolgushev & Kel'manov, 2011]. The running time of the algorithm is $O(qN^2)$. A polynomial-time approximation scheme for Problem 1 with a $O(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$ -time complexity, where ε is an arbitrary relative error, is substantiated in [Dolgushev et al., 2015].

In [Kel'manov & Khandeev, 2015a], a randomized algorithm for Problem 1 is presented. For an established parameter value, the algorithm finds an approximate solution of the problem in polynomial time for given values of the relative error and failure probability; the conditions are established under which the algorithm is asymptotically exact and runs in polynomial time.

Algorithm \mathcal{A}_1 (randomized algorithm).

Input: \mathcal{Y} , M , and a positive integer parameter k .

Step 1. Generate a multiset \mathcal{T} by independently and randomly choosing k elements one after another (with replacement) from \mathcal{Y} .

Step 2. For each nonempty $\mathcal{H} \subseteq \mathcal{T}$, compute the centroid $\bar{y}(\mathcal{H})$ and form a subset $\mathcal{B}_M(\bar{y}(\mathcal{H}))$ consisting of the M points from \mathcal{Y} , having the largest projections onto the direction specified by this centroid. Compute $S(\mathcal{B}_M(\bar{y}(\mathcal{H})))$.

Step 3. In the family of solutions found at Step 2, choose the subset $\mathcal{C}_A = \mathcal{B}_M(\bar{y}(\mathcal{H}))$ for which $S(\mathcal{B}_M(\bar{y}(\mathcal{H})))$ is minimal. If there are several solutions with the same minimal value, take an arbitrary of them.

Output: \mathcal{C}_A .

Theorem 1. For arbitrary $\delta \in (0, 1)$ and positive integer $t \leq k$, algorithm \mathcal{A}_1 finds a $(1 + \frac{1}{\delta t})$ -approximate solution of Problem 1 in $\mathcal{O}(2^k q(k + N))$ time with probability not less than $1 - (\delta + \alpha)$, where $\alpha = \sum_{i=0}^{t-1} \binom{k}{i} \left(\frac{M}{N}\right)^i \left(1 - \frac{M}{N}\right)^{k-i}$.

Corollary 1. Assume that $M \geq \beta N$, where $\beta \in (0, 1)$ is a constant. Then, given $\varepsilon > 0$ and $\gamma \in (0, 1)$ for the fixed parameter $k = \max(\lceil \frac{2}{\beta} \lceil \frac{2}{\gamma \varepsilon} \rceil \rceil, \lceil \frac{8}{\beta} \ln \frac{2}{\gamma} \rceil)$ algorithm \mathcal{A}_1 finds a $(1 + \varepsilon)$ -approximate solution of Problem 1 with probability not less than $1 - \gamma$ in $\mathcal{O}(qN)$ time.

Below are the conditions for algorithm \mathcal{A}_1 to be asymptotically exact.

Theorem 2. Under the conditions of Theorem 1, let $k = \lceil \log_2 N \rceil$, $\delta = (\log_2 N)^{-1/2}$, $t = \lceil \frac{kM}{2N} \rceil$. Assume that $M \geq \beta N$, where $\beta \in (0, 1)$ is a constant. Then algorithm \mathcal{A}_1 finds a $(1 + \varepsilon_N)$ -approximate solution of Problem 1 with probability $1 - \gamma_N$ in $\mathcal{O}(qN^2)$ time, where $\varepsilon_N \leq \frac{2}{\beta} (\log_2 N)^{-1/2}$, $\gamma_N \leq (\log_2 N)^{-1/2} + N^{-\frac{\beta}{8 \ln 2}}$.

A pseudopolynomial algorithm which finds an optimal solution in the case of integer components of the points in the input set and fixed space dimension is proposed in [Kel'manov & Khandeev, 2015b].

Algorithm \mathcal{A}_2 (exact pseudopolynomial algorithm).

Input: \mathcal{Y} , M .

Step 1. Given $x \in \mathcal{G}$, where \mathcal{G} is a multidimensional cubic uniform in each coordinate grid of size $2D$ with the distance $\frac{1}{M}$ between the nodes and the center at the origin, and D is the maximal modulus of the coordinates of input points, construct the subset $\mathcal{B}_M(x)$ consisting of the M points from \mathcal{Y} , having the largest projections onto the direction specified by x . Calculate $Q(\mathcal{B}_M(x), x) = \sum_{y \in \mathcal{B}_M(x)} \|y - x\|^2 + \sum_{y \in \mathcal{Y} \setminus \mathcal{B}_M(x)} \|y\|^2$.

Step 2. Find the point $x_A = \arg \min_{x \in \mathcal{G}} Q(\mathcal{B}_M(x), x)$ (if there are several points with the same minimal value, take an arbitrary of them) and the corresponding subset $\mathcal{B}_M(x_A)$. As a solution to the problem, take $\mathcal{C}_A = \mathcal{B}_M(x_A)$.

Output: \mathcal{C}_A .

Theorem 3. Suppose that, under the conditions of Problem 1, the points of \mathcal{Y} have integer components lying in $[-D, D]$. Then algorithm \mathcal{A}_2 finds an optimal solution in $\mathcal{O}(qN(2MD + 1)^q)$ time.

In the case of fixed space dimension q the algorithm is pseudopolynomial, since under this condition, the running time of the algorithm is estimated by $\mathcal{O}(N(MD)^q)$.

In [Kel'manov & Khandeev, 2016], it was proved that, unless $P=NP$, in the general case of Problem 1 there is no fully polynomial-time approximation scheme (FPTAS). In addition, such a scheme was presented for the case of fixed space dimension.

Algorithm \mathcal{A}_3 (approximation scheme).

Input: \mathcal{Y} , M and $\varepsilon > 0$.

For each point $y \in \mathcal{Y}$ Steps 1–5 are executed.

Step 1. Construct the set $\mathcal{B}_M(y)$ of the M points from \mathcal{Y} , having the largest projections onto the direction specified by y .

Step 2. Compute $S(\mathcal{B}_M(y))$, $h(y, \varepsilon) = \sqrt{\frac{2\varepsilon}{qM} S(\mathcal{B}_M(y))}$ and $H(y) = \sqrt{\frac{1}{M} S(\mathcal{B}_M(y))}$.

Step 3. If $S(\mathcal{B}_M(y)) = 0$, then return the set $\mathcal{B}_M(y)$ as a result \mathcal{C}_A produced by the algorithm and exit; otherwise, go to the next step.

Step 4. Construct a multidimensional cubic uniform in each coordinate grid $\mathcal{G}(y, h, H + h/2)$ of size $2H + h$ with the distance h between the nodes and the center at the point y .

Step 5. For each node x of grid $\mathcal{G}(y, h, H + h/2)$, construct the set $\mathcal{B}_M(x)$ of the M points from \mathcal{Y} , having the largest projections onto the direction specified by x , and compute $S(\mathcal{B}_M(x))$.

Step 6. In the family $\{\mathcal{B}_M(x) \mid x \in \mathcal{G}(y, h, H + h/2), y \in \mathcal{Y}\}$ of sets, choose, as a solution \mathcal{C}_A , the set $\mathcal{B}_M(x)$ for which $S(\mathcal{B}_M(x))$ is minimal. If there are several solutions with the same minimal value, take an arbitrary of them.

Output: \mathcal{C}_A .

Theorem 4. For any fixed $\varepsilon > 0$ algorithm \mathcal{A}_3 finds a $(1 + \varepsilon)$ -approximate solution of Problem 1 in $\mathcal{O}(qN^2(\sqrt{\frac{2q}{\varepsilon}} + 2)^q)$ time.

In the case of fixed space dimension q the algorithm implements an FPTAS, since under this condition, the running time of the algorithm is estimated by $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$.

3 Partitioning a Sequence

Problem 2.

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 subset $\mathcal{M} = \{n_1, \dots, n_M\} \subseteq \mathcal{N} = \{1, \dots, N\}$ such that

$$F(\mathcal{M}) = \sum_{j \in \mathcal{M}} \|y_j - \bar{y}(\mathcal{M})\|^2 + \sum_{i \in \mathcal{N} \setminus \mathcal{M}} \|y_i\|^2 \rightarrow \min,$$

where $\bar{y}(\mathcal{M}) = \frac{1}{|\mathcal{M}|} \sum_{i \in \mathcal{M}} y_i$, under constraints

$$1 \leq T_{\min} \leq n_m - n_{m-1} \leq T_{\max} \leq N, \quad m = 2, \dots, M, \quad (1)$$

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

In this problem, (1) define the constraints on the elements of the cluster $\{y_i \mid i \in \mathcal{M}\}$ with unknown centroid.

In [Kel'manov & Pyatkin, 2013], the variant of Problem 2 was under study with T_{\min} and T_{\max} as parameters, and it was established that Problem 2 is strongly NP-hard for every $T_{\min} < T_{\max}$. In the trivial case when $T_{\min} = T_{\max}$, this problem is solvable in polynomial time.

A 2-approximation polynomial algorithm with complexity $\mathcal{O}(N^2(MN + q))$ was proposed in [Kel'manov & Khamidullin, 2014].

For the case where the components of the points are integers and the dimension q is fixed, the exact pseudopolynomial algorithm was introduced in [Kel'manov et al., 2017a].

Algorithm \mathcal{A}_4 (exact pseudopolynomial algorithm).

Input: \mathcal{Y} , T_{\min} , T_{\max} , M .

Step 1. For each point $x \in \mathcal{G}$, where \mathcal{G} is a multidimensional cubic uniform in each coordinate grid of size $2D$ with the distance $\frac{1}{M}$ between the nodes and the center at the origin, and D is the maximal modulus of the coordinates of input points, find the optimal solution \mathcal{M}^x of the problem

$$G^x(\mathcal{M}) = \sum_{n \in \mathcal{M}} (2\langle y_n, x \rangle - \|x\|^2) \rightarrow \max \quad (2)$$

under constraints (1) on the elements of \mathcal{M} . Compute $G^x(\mathcal{M}^x)$.

Step 2. Find the point $x_A = \arg \max_{x \in \mathcal{G}} G^x(\mathcal{M}^x)$ (if there are several points with the same maximal value, take an arbitrary of them) and the corresponding subset $\mathcal{M}_A = \mathcal{M}^{x_A}$.

Output: \mathcal{M}_A .

At Step 1 the optimal solution of the problem (2) is found in $\mathcal{O}(N(M(T_{\max} - T_{\min} + 1) + q))$ time using the dynamic programming scheme [Kel'manov & Khamidullin, 2014].

Theorem 5. Suppose that, under the conditions of Problem 2, the points of \mathcal{Y} have integer components lying in $[-D, D]$. Then algorithm \mathcal{A}_4 finds an optimal solution in $\mathcal{O}(N(M(T_{\max} - T_{\min} + 1) + q)(2MD + 1)^q)$ time.

In the case of fixed space dimension q the algorithm is pseudopolynomial, since under this condition, the running time of the algorithm is estimated by $\mathcal{O}(N^3(MD)^q)$.

In [Kel'manov et al., 2016], an approximation algorithm was proposed which for the fixed dimension q of the space implements the FPTAS scheme.

Algorithm \mathcal{A}_5 (approximation scheme).

Input: \mathcal{Y} , T_{\min} , T_{\max} , M and $\varepsilon > 0$.

For each point $y \in \mathcal{Y}$ Steps 1–5 are executed.

Step 1. Find [Kel'manov & Khamidullin, 2014] an optimal solution \mathcal{M}^y of problem (2) for $x = y$.

Step 2. Compute $F(\mathcal{M}^y)$, $h = \sqrt{\frac{2\varepsilon}{qM}F(\mathcal{M}^y)}$ and $H = \sqrt{\frac{1}{M}F(\mathcal{M}^y)}$.

Step 3. If $F(\mathcal{M}^y) = 0$, then return the set \mathcal{M}^y as a result \mathcal{M}_A produced by the algorithm and exit; otherwise, go to the next step.

Step 4. Construct a multidimensional cubic uniform in each coordinate grid $\mathcal{G}(y, h, H + h/2)$ of size $2H + h$ with the distance h between the nodes and the center at the point y .

Step 5. For each node x of grid $\mathcal{G}(y, h, H + h/2)$ construct [Kel'manov & Khamidullin, 2014] an optimal solution \mathcal{M}^x of problem (2) and compute $F(\mathcal{M}^x)$.

Step 6. In the family $\{\mathcal{M}^x \mid x \in \mathcal{G}(y, h, H + h/2), y \in \mathcal{Y}\}$ of sets, choose, as a solution \mathcal{M}_A , the set \mathcal{M}^x , for which $F(\mathcal{M}^x)$ is minimal. If there are several solutions with the same minimal value, take an arbitrary of them.

Output: \mathcal{M}_A .

Theorem 6. For any fixed $\varepsilon > 0$ algorithm \mathcal{A}_5 finds a $(1 + \varepsilon)$ -approximate solution of Problem 2 in $\mathcal{O}(N^2(M(T_{\max} - T_{\min} + 1) + q)(\sqrt{\frac{2q}{\varepsilon}} + 2)^q)$ time.

In the case of fixed space dimension q the algorithm implements an FPTAS, since under this condition, the running time of the algorithm is estimated by $\mathcal{O}(MN^3(1/\varepsilon)^{q/2})$.

In [Kel'manov et al., 2017b], a randomized algorithm for Problem 2 is presented.

Algorithm \mathcal{A}_6 (randomized algorithm).

Input: \mathcal{Y} , T_{\min} , T_{\max} , M , and a positive integer parameter k .

Step 1. Generate a multiset \mathcal{T} by independently and randomly choosing k elements one after another (with replacement) from \mathcal{Y} .

Step 2. For each nonempty $\mathcal{H} \subseteq \mathcal{T}$, compute the centroid $\bar{y}(\mathcal{H})$ and construct [Kel'manov & Khamidullin, 2014] an optimal solution \mathcal{M}^x of problem (2) for $x = \bar{y}(\mathcal{H})$.

Step 3. In the family of solutions found at Step 2, of sets, choose a set $\mathcal{M}_A = \mathcal{M}^x$ for which $F(\mathcal{M}^x)$ is minimal. If there are several solutions with the same minimal value, take an arbitrary of them.

Output: \mathcal{M}_A .

Theorem 7. For arbitrary $\delta \in (0, 1)$ and positive integer $t \leq k$, algorithm \mathcal{A}_6 finds a $(1 + \frac{1}{\delta t})$ -approximate solution of Problem 2 in $\mathcal{O}(2^k(qk + N(M(T_{\max} - T_{\min} + 1) + q)))$ time with probability not less than $1 - (\delta + \alpha)$, where $\alpha = \sum_{i=0}^{t-1} \binom{k}{i} \left(\frac{M}{N}\right)^i \left(1 - \frac{M}{N}\right)^{k-i}$.

Corollary 2. Assume that $M \geq \beta N$, where $\beta \in (0, 1)$ is a constant. Then, given $\varepsilon > 0$ and $\gamma \in (0, 1)$ for the fixed parameter $k = \max(\lceil \frac{2}{\beta} \lceil \frac{2}{\gamma\varepsilon} \rceil \rceil, \lceil \frac{8}{\beta} \ln \frac{2}{\gamma} \rceil)$ algorithm \mathcal{A}_6 finds a $(1 + \varepsilon)$ -approximate solution of Problem 2 with probability not less than $1 - \gamma$ in $\mathcal{O}(qMN^2)$ time.

Theorem 8. Under the conditions of Theorem 7, let $k = \lceil \log_2 N \rceil$, $\delta = (\log_2 N)^{-1/2}$, $t = \lceil \frac{kM}{2N} \rceil$. Assume that $M \geq \beta N$, where $\beta \in (0, 1)$ is a constant. Then algorithm \mathcal{A}_6 finds a $(1 + \varepsilon_N)$ -approximate solution of Problem 2 with probability $1 - \gamma_N$ in $\mathcal{O}(qMN^2(T_{\max} - T_{\min} + 1))$ time, where $\varepsilon_N \leq \frac{2}{\beta}(\log_2 N)^{-1/2}$, $\gamma_N \leq (\log_2 N)^{-1/2} + N^{-\frac{\beta}{8 \ln 2}}$.

Since $T_{\max} - T_{\min} + 1 \leq N$, under the specified conditions the running time of the algorithm is estimated by $\mathcal{O}(qMN^3)$.

4 Conclusion

In the paper we present a short survey on some results for two problems of 2-partitioning a finite set and a finite sequence of points in Euclidean space into clusters. Also we present a new randomized algorithm for the

sequence partitioning problem. For an established parameter value, the algorithm finds an approximate solution in polynomial time for fixed values of the relative error and failure probability. The conditions are found under which the algorithm is asymptotically exact and has polynomial-time complexity.

The issue of great interest is the substantiation of faster randomized algorithms for the problems and the search for subclasses of the problems for which linear time and sub-linear time randomized algorithms can be designed.

Acknowledgements

This work was supported by Russian Foundation for Basic Research, Projects 15-01-00462, 16-31-00186, 16-07-00168.

References

- [Aloise et al., 2009] Aloise, D., Deshpande, A., Hansen, P., & Popat, P. (2009). NP-hardness of Euclidean sum-of-squares clustering. *Mach. Learn.*, 75(2), 245–248.
- [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. and Math. Phys.*, 49(11), 1966–1971.
- [Gimadi et al., 2010] Gimadi, E.Kh., Pyatkin, A.V., & Rykov, I.A. (2010). On polynomial solvability of some problems of a vector subset choice in a Euclidean space of fixed dimension. *J. Appl. Ind. Math.*, 4(1), 48–53.
- [Shenmaier, 2016] Shenmaier, V.V. (2016). Solving some vector subset problems by Voronoi diagrams. *J. Appl. Ind. Math.*, 10(4), 560–566.
- [Doligushev & Kel'manov, 2011] Doligushev, A.V., & Kel'manov, A.V. (2011). An approximation algorithm for solving a problem of cluster analysis. *J. Appl. Ind. Math.*, 5(4), 551–558.
- [Doligushev et al., 2015] Doligushev, A.V., Kel'manov, A.V., & Shenmaier, V.V. (2015). Polynomial-time approximation scheme for a problem of partitioning a finite set into two clusters (in Russian). *Tr. Inst. Mat. Mekh.*, 21(3), 100–109.
- [Kel'manov & Khandeev, 2015a] Kel'manov, A.V., & Khandeev, V.I. (2015). A randomized algorithm for two-cluster partition of a set of vectors. *Comput. Math. Math. Phys.*, 55(2), 330–339.
- [Kel'manov & Khandeev, 2015b] Kel'manov, A.V., & Khandeev, V.I. (2015). An exact pseudopolynomial algorithm for a problem of the two-cluster partitioning of a set of vectors. *J. Appl. Ind. Math.*, 9(4), 497–502.
- [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. *J. Appl. Ind. Math.*, 56(2), 334–341.
- [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.
- [Kel'manov & Khamidullin, 2014] Kel'manov A.V., & Khamidullin S.A. (2014). An Approximating Polynomial Algorithm for a Sequence Partitioning Problem. *J. Appl. Indust. Math.*, 8(2), 236–244.
- [Kel'manov et al., 2017a] Kel'manov A.V., Khamidullin S.A., & Khandeev V.I. (2017). Exact Pseudopolynomial Algorithm for one Sequence Partitioning Problem. *Automat. Remote Control*, 78(1), 66–73.
- [Kel'manov et al., 2016] Kel'manov A.V., Khamidullin S.A., & Khandeev V.I. (2016). A Fully Polynomial-Time Approximation Scheme for a Sequence 2-Cluster Partitioning Problem. *J. Appl. Ind. Math.*, 10(2), 209–219.
- [Kel'manov et al., 2017b] Kel'manov A.V., Khamidullin S.A., & Khandeev V.I. (2017). A randomized algorithm for a sequence 2-clustering problem (in Russian). *Comput. Math. and Math. Phys.* (accepted).