УДК 519.2+621.391

О псевдополиномиальной разрешимости квадратичной евклидовой задачи поиска семейства непересекающихся подмножеств*

А.Е. Галашов[1], А.В. Кельманов[1,2]

[1]Новосибирский национальный исследовательский государственный университет (НГУ), ул. Пирогова, 2, Новосибирск, 630090

[2]Институт математики им. С.Л. Соболева Сибирского отделения Российской академии наук, просп. Акад. В.А. Коптюга, 4, Новосибирск, 630090

E-mails: galashov.alexandr@gmail.com (Галашов А.Е.), kelm@math.nsc.ru (Кельманов А.В.)

Галашов А.Е., Кельманов А.В. О псевдополиномиальной разрешимости квадратичной евклидовой задачи поиска семейства непересекающихся подмножеств // Сиб. журн. вычисл. математики / РАН. Сиб. отд-ние. — Новосибирск, 2017. — Т. 20, № 1. — С. 15–22.

Рассматривается NP-трудная в сильном смысле задача поиска в конечном множестве точек евклидова пространства семейства непересекающихся подмножеств, имеющих заданные мощности. Критерием решения задачи является минимум суммы по всем подмножествам сумм квадратов расстояний от элементов подмножеств до их геометрических центров. Доказано, что задача разрешима за псевдополиномиальное время, если координаты входных точек целочисленны, а размерность пространства и число искомых подмножеств фиксированы (ограничены константами).

DOI: 10.15372/SJNM20170102

Ключевые слова: *поиск подмножеств, кластерный анализ, евклидово пространство, минимум суммы квадратов расстояний,* NP*-трудная задача, точный псевдополинимиальный алгоритм.*

Galashov A.E., Kel'manov A.V. On pseudopolynomial-time solvability of a quadratic Euclidean problem of finding a family of disjoint subsets // Siberian J. Num. Math. / Sib. Branch of Russ. Acad. of Sci. — Novosibirsk, 2017. — Vol. 20, № 1. — P. 15–22.

We consider a strongly NP-hard problem of finding a family of disjoint subsets with given cardinalities in a finite set of points from the Euclidean space. A minimum of the sum over all required subsets of the sum of the squared distances from the elements of these subsets to their geometric centers is used as a search criterion. We have proved that if the coordinates of the input points are integer, and the space dimension and the number of required subsets are fixed (i.e. bounded by some constants), then the problem is a pseudopolynomial-time solvable one.

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

Введение

Предметом исследования настоящей работы является одна из NP-трудных в сильном смысле задач поиска семейства непересекающихся подмножеств в конечном множестве точек евклидова пространства. Цель работы — выяснение вопроса о псевдополиномиальной разрешимости задачи.

*Работа выполнена при поддержке РФФИ (проекты № 15-01-00462, № 16-07-00168).

Исследование мотивировано слабой изученностью задачи в алгоритмическом плане. В частности, до настоящего времени оставался открытым вопрос о существовании алгоритма, который гарантирует отыскание точного решения за псевдополиномиальное время для какого-либо подкласса (случая) задачи. Поиск подклассов NP-трудных в сильном смысле задач, для которых точные псевдополиномиальные алгоритмы существуют, является традиционным и важным направлением исследований как для дискретной оптимизации, так и для вычислительной математики.

Статья развивает и дополняет результаты, полученные в [1, 2], и имеет следующую структуру. В пункте 1 приведена формулировка задачи, дан обзор известных результатов и анонсирован полученный результат. В п. 2 сформулированы базовые утверждения, обеспечивающие доказательство основного результата. Наконец, в п. 3 конструктивно (алгоритмически) показана псевдополиномиальная разрешимость задачи в случае, когда размерность пространства и число искомых подмножеств не являются частью входа (т. е. фиксированы или ограничены константами), а координаты входных точек целочисленны.

1. Формулировка задачи, известные и полученный результаты

Всюду далее $\mathbb{R}$ — множество вещественных чисел, $\|\cdot\|$ — евклидова норма, а $\mathbb{Z}$ — множество целых чисел.

Рассматриваемая задача имеет следующую формулировку [2].

Задача 1. *Дано*: множество $\mathcal{Y} = \{y_1, \ldots, y_N\}$ точек из $\mathbb{R}^q$ и натуральные числа $M_1, \ldots, M_J$. *Найти*: семейство $\{\mathcal{C}_1, \ldots, \mathcal{C}_J\}$ непересекающихся подмножеств множества $\mathcal{Y}$ такое, что

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

где $\overline{y}(\mathcal{C}_j) = \frac{1}{|\mathcal{C}_j|}\sum_{y \in \mathcal{C}_j} y$ — центроид (геометрический центр) подмножества $\mathcal{C}_j$ при ограничениях $|\mathcal{C}_j| = M_j$, $j = 1, \ldots, J$, на мощности искомых подмножеств.

Задача 1 имеет очевидную геометрическую интерпретацию: поиск в конечном множестве точек евклидова пространства дизъюнктных структур с экстремальным свойством, описываемым целевой функцией (1). Истоки (приложения) и содержательная трактовка задачи приведены в [2]. Там же представлены статистическая и аппроксимационная задачи, которые индуцируют задачу 1.

Напомним (см. [2]), что задача 1 в постановочном плане близка к хорошо известной NP-трудной в сильном смысле [3] задаче MSSC (*Minimum Sum-of-Squares Clustering*) кластеризации, которая известна также под названием *k-means* (*k-средних*) [4–6]. В этой задаче целевая функция такая же, как и в задаче 1. Отличия задачи MSSC состоят в том, что в ней мощности кластеров являются функциями от оптимизируемых переменных (искомых подмножеств), а вместо поиска семейства подмножеств, объединение которых может не покрывать входное множество, требуется найти разбиение этого множества. Поэтому задачу 1 можно трактовать как обобщение задачи MSSC с дополнительным ограничением на мощности искомых подмножеств (кластеров).

Указанные отличия, фактически, показывают, что задача 1 не эквивалентна задаче MSSC. К тому же она не является частным случаем какой-либо другой известной задачи. По этой причине алгоритмические результаты для известных задач не переносятся на задачу 1. Она требует самостоятельного исследования.

Поскольку задача 1 NP-трудна в сильном смысле (см. далее), для нее, согласно [7], не существует ни точного полиномиального, ни точного псевдополиномиального алгоритмов, если гипотеза P≠NP верна. В связи с этим с математической точки зрения представляет интерес выяснение вопросов аппроксимируемости задачи 1. В частности, традиционным и актуальным для дискретной оптимизации и вычислительной математики является вопрос поиска подклассов (случаев) какой-либо сильно NP-трудной задачи, для которой точные псевдополиномиальные алгоритмы существуют. К числу таких задач относится рассматриваемая задача.

Сильная NP-трудность задачи 1 следует из результатов, полученных в [1], так как в цитируемой работе установлено, что частный случай задачи 1, в котором $J = 1$, является NP-трудной в сильном смысле задачей.

В алгоритмическом плане задача 1 относится к числу слабоизученных проблем. В [2] для этой задачи предложен 2-приближенный алгоритм, который находит решение за время $\mathcal{O}\big(N^2(N^{J+1} + q)\big)$. Алгоритм полиномиален, если число J искомых подмножеств фиксировано. Другие алгоритмические результаты для задачи 1 в настоящее время отсутствуют, а известные результаты [8–11] получены лишь для ее частного случая, в котором $J = 1$. Эти результаты приведены ниже.

В [8] предложен 2-приближенный полиномиальный алгоритм, трудоемкость которого есть величина $\mathcal{O}(qN^2)$. Точный псевдополиномиальный алгоритм для случая задачи, в котором размерность q пространства фиксирована, а компоненты векторов имеют целочисленные значения, обоснован в [9]. Временная сложность этого алгоритма есть величина $\mathcal{O}\big(N(MB)^q\big)$, где B — максимальное абсолютное значение координат точек входного множества, а M — мощность искомого подмножества. Кроме того, для случая фиксированной размерности пространства в [10] обоснована полностью полиномиальная приближенная схема (FPTAS), которая для заданной относительной погрешности ε гарантирует отыскание $(1+\varepsilon)$-приближенного решения задачи 1 за время $\mathcal{O}\big(N^2(M/\varepsilon)^q\big)$. В [11] построена приближенная полиномиальная схема (PTAS), обеспечивающая решение задачи с произвольной относительной погрешностью ε за время $\mathcal{O}\big(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon}\big)$.

В настоящей работе для задачи 1 с целочисленными входами построен алгоритм, который, при фиксированных размерности q пространства и числе $J > 1$ искомых подмножеств, гарантирует отыскание точного решения задачи за псевдополиномиальное время $\mathcal{O}\big(\mathcal{N}^3(MB)^{qJ}\big)$, где B — максимальное абсолютное значение координат входных точек, а M — наименьшее общее кратное чисел $M_1, \ldots, M_J$. Полученный результат устанавливает условия, при которых задача 1 разрешима за псевдополиномиальное время, а именно: целочисленность входов и ограниченность константами размерности пространства и числа искомых подмножеств.

Идея алгоритма весьма проста. Она базируется на традиционном для вычислительной математики сеточном подходе и на полиномиальной разрешимости одной из немногих задач комбинаторной оптимизации — задачи о *назначениях* (*Assignment* problem).

2. Основы алгоритма

Далее мы используем $(x)^i$ для обозначения i-й координаты точки $x \in \mathbb{R}^q$.

Для обоснования алгоритма нам потребуется вспомогательная задача, точный полиномиальный алгоритм ее решения, а также следующая лемма.

Лемма 1. *Пусть выполнены условия задачи 1,* $\mathcal{Y} \subset \mathbb{Z}^q$ *и*

$$B = \max_{y \in \mathcal{Y}} \max_{i \in \{1,\ldots,q\}} |(y)^i|. \tag{2}$$

Тогда центроид $\overline{y}(\mathcal{C}_j)$, $j = 1, \dots, J$, *лежит во множестве*

$$\mathcal{D} = \Big\{ z \in \mathbb{R}^q \;\Big|\; (z)^k = \frac{1}{M}(v)^k,\ (v)^k \in \mathbb{Z},\ |(v)^k| \le MB,\ k = 1, \dots, q \Big\}, \tag{3}$$

где M *— наименьшее общее кратное для чисел* $M_1, \dots, M_J$.

Доказательство. По условию леммы координаты точек из $\mathcal{Y}$ целочисленны, их абсолютные значения не превосходят B, а $|\mathcal{C}_j| = M_j$, $j = 1, \dots, J$. Поэтому k-я координата ($k = 1, \dots, q$) центроида $\overline{y}(\mathcal{C}_j)$, $j = 1, \dots, J$, равная $\frac{1}{M_j}\sum_{y \in \mathcal{C}_j}(y)^k$, является рациональным числом таким, что

$$\big|(\overline{y}(\mathcal{C}_j))^k\big| = \frac{1}{|\mathcal{C}_j|}\Big|\sum_{y \in \mathcal{C}_j}(y)^k\Big| = \frac{1}{M_j}\Big|\sum_{y \in \mathcal{C}_j}(y)^k\Big| \le B \tag{4}$$

в соответствии с определением геометрического центра (как среднего значения на отрезке $[-B, B]$).

По определению наименьшего общего кратного M чисел $M_1, \dots, M_J$ для каждого $j = 1, \dots, J$ существует целое число $s(j)$ такое, что $0 < s(j) \le M$, причем $M = s(j)M_j$. Поэтому для k-й координаты j-го центроида имеем эквивалентное представление в форме

$$\big(\overline{y}(\mathcal{C}_j)\big)^k = \frac{s(j)}{M}\sum_{y \in \mathcal{C}_j}(y)^k$$

рационального числа, знаменатель которого равен наименьшему общему кратному M. Это рациональное число лежит в интервале $[-B, B]$. Действительно,

$$\Big|\frac{s(j)}{M}\sum_{y \in \mathcal{C}_j}(y)^k\Big| = \frac{s(j)}{M}\Big|\sum_{y \in \mathcal{C}_j}(y)^k\Big| \le \frac{s(j)M_jB}{M} = B$$

в соответствии со свойствами чисел $s(j)$ и M и в соответствии с неравенством (4). Отсюда следует справедливость леммы. □

Множество $\mathcal{D}$ является многомерной решеткой с рациональным шагом $1/M$ и центром в начале координат. Для ее мощности (числа узлов) справедлива очевидная формула

$$|\mathcal{D}| = (2MB + 1)^q, \tag{5}$$

которая нам потребуется далее при оценке трудоемкости алгоритма.

Сформулированная ниже вспомогательная задача 2 и точный полиномиальный алгоритм ее решения являются вычислительной базой предлагаемого алгоритма.

Задача 2. *Дано*: множество $\mathcal{Y} = \{y_1, \dots, y_N\}$ и набор $b = (b_1, \dots, b_J)$ точек из $\mathbb{R}^q$, натуральные числа $M_1, \dots, M_J$. *Найти*: семейство $\{\mathcal{B}_1, \dots, \mathcal{B}_J\}$ непересекающихся подмножеств множества $\mathcal{Y}$ такое, что

$$G^b(\mathcal{B}_1, \dots, \mathcal{B}_J) = \sum_{j=1}^{J}\sum_{y \in \mathcal{B}_j}\|y - b_j\|^2 \to \min, \tag{6}$$

при ограничениях $|\mathcal{B}_j| = M_j$, $j = 1, \dots, J$, на мощности искомых подмножеств.

В [2] показано, что задача 2 полиномиально сводится к известной (см., например, [12]) полиномиально разрешимой задаче о *назначениях* (*Assignment* problem), а точное алгоритмическое решение задачи 2 можно найти за время $\mathcal{O}\big(N(N^2 + qJ)\big)$. Обозначим через $\mathcal{A}_1$ алгоритм решения вспомогательной задачи 2, предложенный в [2].

3. Точный псевдополиномиальный алгоритм

Суть предлагаемого алгоритмического решения состоит в следующем. В области пространства, определяемой максимальным абсолютным значением B координат входных точек множества $\mathcal{Y}$, строится многомерная равномерная по каждой координате решетка $\mathcal{D}$ с рациональным шагом. Шаг решетки выбирается по наименьшему общему кратному M для чисел $M_1, \ldots, M_J$, т. е. так, чтобы множество узлов решетки содержало узлы (точки), совпадающие с центроидами всех оптимальных подмножеств. Для каждого набора из J различных q-мерных точек (узлов) построенной решетки решается вспомогательная задача 2. Ее решение — семейство непересекающихся подмножеств — включается в совокупность претендентов на решение исходной задачи 1. Решением задачи 1 объявляется семейство подмножеств из построенной совокупности, для которого целевая функция вспомогательной задачи 2 имеет наименьшее значение.

А л г о р и т м $\mathcal{A}$. *Вход алгоритма*: множество $\mathcal{Y}$ и натуральные числа $M_1, \ldots, M_J$.

Шаг 1. Найдем значение M наименьшего общего кратного чисел $M_1, \ldots, M_J$. Вычислим значение B по формуле (2). Построим решетку $\mathcal{D}$ по формуле (3).

Шаг 2. Для каждого набора $d = (d_1, \ldots, d_J) \in \mathcal{D}^J$ различных точек с помощью алгоритма $\mathcal{A}_1$ найдем и запомним точное решение $\{\mathcal{B}_1(d), \ldots, \mathcal{B}_J(d)\}$ вспомогательной задачи 2 (при $b = d$ в формуле (6)) и значение целевой функции G^d.

Шаг 3. В совокупности $\{\{\mathcal{B}_1(d), \ldots, \mathcal{B}_J(d)\},\ d \in \mathcal{D}^J\}$ решений (семейств подмножеств), полученных на шаге 2, найдем набор $d_A = (d_1^A, \ldots, d_J^A)$, доставляющий минимум целевой функции G^d, $d \in \mathcal{D}^J$. В качестве решения $\{\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\}$ задачи 1 возьмем семейство $\{\mathcal{B}_1(d_A), \ldots, \mathcal{B}_J(d_A)\}$, соответствующее набору d_A.

Выход алгоритма: набор $\{\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\}$.

Свойства этого алгоритма устанавливает

Теорема 1. *Пусть в условиях задачи 1 компоненты всех точек из множества* $\mathcal{Y}$ *имеют целочисленные значения в интервале* $[-B, B]$. *Тогда при* $J > 1$ *алгоритм* $\mathcal{A}$ *находит оптимальное решение этой задачи за время*

$$\mathcal{O}(\mathcal{N}(\mathcal{N}^2 + qJ)(2MB + 1)^{qJ} + (J-1)\lg N).$$

Доказательство. Пусть $\{\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*\}$ — оптимальное решение задачи 1, а $\{\overline{y}(\mathcal{C}_1^*), \ldots, \overline{y}(\mathcal{C}_j^*)\}$ — набор центроидов оптимальных подмножеств. В соответствии с определением шага 3 алгоритм $\mathcal{A}$ находит решение задачи 1 в виде:

$$\big\{\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\big\} = \big\{\mathcal{B}_1(d_A), \ldots, \mathcal{B}_J(d_A)\big\}, \tag{7}$$

где

$$d_A = \arg\min_{d \in \mathcal{D}^J} G^d\big(\mathcal{B}_1(d), \ldots, \mathcal{B}_J(d)\big). \tag{8}$$

Поэтому, заметив, что в соответствии с леммой 1 центроиды оптимальных подмножеств лежат во множестве $\mathcal{D}$, из формул (6), (8) и (1) имеем

$$G^{d_A}\big(\mathcal{C}_1^A, \ldots, \mathcal{C}_J^A\big) = \sum_{j=1}^{J} \sum_{y \in \mathcal{C}_j^A} \big\|y - d_j^A\big\|^2 \le \sum_{j=1}^{J} \sum_{y \in \mathcal{C}_j^*} \big\|y - \overline{y}(\mathcal{C}_j^*)\big\|^2 = F\big(\mathcal{C}_1^*, \ldots, \mathcal{C}_J^*\big). \tag{9}$$

Легко проверить с помощью дифференцирования, что для любого непустого конечного множества $\mathcal{Z}$ точек из $\mathbb{R}^q$ минимум суммы $\sum_{z\in\mathcal{Z}}\|z-x\|^2$ по x достигается в точке $x=\overline{z}=\frac{1}{|\mathcal{Z}|}\sum_{z\in\mathcal{Z}} z$, т. е. в точке, равной центроиду этого множества. Поэтому из (1) и (6)–(8) получим

$$F\big(\mathcal{C}_1^A,\dots,\mathcal{C}_J^A\big)=\sum_{j=1}^J\sum_{y\in\mathcal{C}_j^A}\big\|y-\overline{y}(\mathcal{C}_j^A)\big\|^2\le\sum_{j=1}^J\sum_{y\in\mathcal{C}_j^A}\big\|y-d_j^A\big\|^2=G^{d_A}\big(\mathcal{C}_1^A,\dots,\mathcal{C}_J^A\big).\qquad(10)$$

Объединив (9) и (10), найдем оценку $F\big(\mathcal{C}_1^A,\dots,\mathcal{C}_J^A\big)\le F\big(\mathcal{C}_1^*,\dots,\mathcal{C}_J^*\big)$. С другой стороны, множества $\mathcal{C}_1^A,\dots,\mathcal{C}_J^A$ являются допустимым решением задачи 1. Поэтому справедлива оценка $F\big(\mathcal{C}_1^*,\dots,\mathcal{C}_J^*\big)\le F\big(\mathcal{C}_1^A,\dots,\mathcal{C}_J^A\big)$. Из полученных оценок следует равенство $F\big(\mathcal{C}_1^A,\dots,\mathcal{C}_J^A\big)=F\big(\mathcal{C}_1^*,\dots,\mathcal{C}_J^*\big)$.

Оценим временную сложность.

Как известно, наименьшее общее кратное lcm (least common multiple) пары натуральных чисел a и b и наибольший общий делитель gcd (greatest common divisor) этих чисел связаны равенством

$$\operatorname{lcm}(a,b)=\frac{ab}{\gcd(a,b)}.\qquad(11)$$

При этом наибольший общий делитель $\gcd(a,b)$ можно найти с помощью алгоритма Евклида за время $\mathcal{O}\big(\lg(\min(a,b))\big)$ (см., например, [13]). Кроме того, опираясь на известное свойство наименьшего общего кратного для чисел $M_1,\dots,M_J$, имеем равенство

$$\operatorname{lcm}(M_1,\dots,M_J)=\operatorname{lcm}\big(\operatorname{lcm}(M_1,\dots,M_{J-1}),M_J\big),\qquad(12)$$

позволяющее по рекуррентной схеме из $J-1$ последовательных вычислений (каждое по формулам (11) и (12)) найти наименьшее общее кратное этих чисел. По условию задачи 1 каждое из чисел $M_1,\dots,M_J$ не превосходит N. Поэтому из (11), (12) и трудоемкости алгоритма Евклида следует, что в этой последовательной схеме время вычисления каждого из $J-1$ наибольших общих делителей не превосходит $\mathcal{O}(\lg N)$, а время вычисления наименьшего общего кратного чисел $M_1,\dots,M_J$ на шаге 1 не превосходит $\mathcal{O}\big((J-1)\lg N\big)$. На этом же шаге затраты на вычисление значения B и построение решетки $\mathcal{D}$ не превосходят $\mathcal{O}(qN)$ и $\mathcal{O}\big(q|\mathcal{D}|\big)$ соответственно.

Время решения вспомогательной задачи (см п. 2) равно $\mathcal{O}\big(N(N^2+qJ)\big)$. На шаге 2 эта задача решается $|\mathcal{D}|^J$ раз. Поэтому в соответствии с (5) трудоемкость шага 2 есть величина $\mathcal{O}\big(N(N^2+qJ)(2MB+1)^{qJ}\big)$.

Просуммировав затраты на всех шагах, получим оценку, приведенную в теореме. □

Заметим, что $(2MB+1)^q\le(3MB)^q=3^q(MB)^q$. Поэтому, если размерность q пространства и число искомых подмножеств ограничены константами, то время работы алгоритма есть величина $\mathcal{O}\big(\mathcal{N}^3(MB)^{qJ}\big)$. Поскольку B — числовое значение, при указанных условиях алгоритм $\mathcal{A}$ решения задачи 1 псевдополиномиален.

Таким образом, если координаты точек входного множества целочисленны, а размерность пространства и число искомых подмножеств не являются частью входа задачи 1, то она разрешима за псевдополиномиальное время.

Заключение

Для целочисленного случая одной из NP-трудных в сильном смысле задач поиска семейства непересекающихся подмножеств в конечном множестве точек евклидова пространства построен точный алгоритм. Показано, что алгоритм псевдополиномиален, если число искомых подмножеств и размерность пространства ограничены константами. Таким образом, установлены условия, при которых рассматриваемая труднорешаемая задача разрешима за псевдополиномиальное время.

Полученный результат важен не только в теоретическом плане. Ясно, что предложенный алгоритм является одним из инструментов при решении практических целочисленных задач небольших размерностей.

Делом ближайшей перспективы представляется построение полиномиальных приближенных схем, обоснование приближенных более быстрых алгоритмов, в том числе алгоритмов рандомизированного типа.

Литература

1. **Кельманов А.В., Пяткин А.В.** NP-полнота некоторых задач выбора подмножества векторов // Дискретный анализ и исследование операций. — 2010. — Т. 17, № 5. — С. 37–45.

2. **Галашов А.Е., Кельманов А.В.** 2-приближенный алгоритм для одной задачи поиска семейства непересекающихся подмножеств векторов // Автоматика и телемеханика. — 2014. — № 4. — С. 5–19.

3. **Aloise D., Deshpande A., Hansen P., and Popat P.** NP-hardness of Euclidean sum-of-squares clustering // Machine Learning. — 2009. — Vol. 75, № 2. — P. 245–248.

4. **Jain A.K.** Data clustering: 50 years beyond K-means // Pattern Recognition Lett. — 2010. — Vol. 31. — P. 651–666.

5. **Edwards A.W.F., Cavalli-Sforza L.L.** A method for cluster analysis // Biometrics. — 1965. — Vol. 21. — P. 362–375.

6. **MacQueen J.B.** Some methods for classification and analysis of multivariate observations // Proc. 5th Berkeley Symp. Math. Statist. Probability. — 1967. — Vol. 1. — P. 281–297.

7. **Garey M.R., Johnson D.S.** Computers and Intractability: A Guide to the Theory of NP-Completeness. — San Francisco: Freeman, 1979.

8. **Кельманов А.В., Романченко С.М.** Приближенный алгоритм решения одной задачи поиска подмножества векторов // Дискретный анализ и исследование операций. — 2011. — Т. 18, № 1. — С. 61–69.

9. **Кельманов А.В., Романченко С.М.** Псевдополиномиальные алгоритмы для некоторых труднорешаемых задач поиска подмножества векторов и кластерного анализа // Автоматика и телемеханика. — 2012. — № 2. — С. 156–162.

10. **Кельманов А.В., Романченко С.М.** FPTAS для одной задачи поиска подмножества векторов // Дискретный анализ и исследование операций. — 2014. — Т. 21, № 3. — С. 41–52.

11. **Шенмайер В.В.** Аппроксимационная схема для одной задачи поиска подмножества векторов // Дискретный анализ и исследование операций. — 2012. — Т. 19, № 2. — С. 92–100.

12. **Papadimitriou C.H., Steiglitz K.** Combinatorial Optimization: Algorithms and Complexity.— Englewood Cliffs, N.J.: Prentice-Hall, 1982.

13. **Mollin R.A.** Fundamental Number Theory with Applications. Second Edition. — CRC Press, 2008.

Поступила в редакцию 26 июля 2016 г.,
в окончательном варианте 29 августа 2016 г.

Литература в транслитерации

1. **Kel'manov A.V., Pyatkin A.V.** NP-polnota nekotorykh zadach vybora podmnozhestva vektorov // Diskretnyy analiz i issledovanie operatsiy. — 2010. — T. 17, № 5. — S. 37–45.
2. **Galashov A.E., Kel'manov A.V.** 2-priblizhennyy algoritm dlya odnoy zadachi poiska semeystva neperesekayushchikhsya podmnozhestv vektorov // Avtomatika i telemekhanika. — 2014. — № 4. — S. 5–19.
3. **Aloise D., Deshpande A., Hansen P., and Popat P.** NP-hardness of Euclidean sum-of-squares clustering // Machine Learning. — 2009. — Vol. 75, № 2. — P. 245–248.
4. **Jain A.K.** Data clustering: 50 years beyond K-means // Pattern Recognition Lett. — 2010. — Vol. 31. — P. 651–666.
5. **Edwards A.W.F., Cavalli-Sforza L.L.** A method for cluster analysis // Biometrics. — 1965. — Vol. 21. — P. 362–375.
6. **MacQueen J.B.** Some methods for classification and analysis of multivariate observations // Proc. 5th Berkeley Symp. Math. Statist. Probability. — 1967. — Vol. 1. — P. 281–297.
7. **Garey M.R., Johnson D.S.** Computers and Intractability: A Guide to the Theory of NP-Completeness. — San Francisco: Freeman, 1979.
8. **Kel'manov A.V., Romanchenko S.M.** Priblizhennyy algoritm resheniya odnoy zadachi poiska podmnozhestva vektorov // Diskretnyy analiz i issledovanie operatsiy. — 2011. — T. 18, № 1. — S. 61–69.
9. **Kel'manov A.V., Romanchenko S.M.** Psevdopolinomial'nye algoritmy dlya nekotorykh trudnoreshaemykh zadach poiska podmnozhestva vektorov i klasternogo analiza // Avtomatika i telemekhanika. — 2012. — № 2. — S. 156–162.
10. **Kel'manov A.V., Romanchenko S.M.** FPTAS dlya odnoy zadachi poiska podmnozhestva vektorov // Diskretnyy analiz i issledovanie operatsiy. — 2014. — T. 21, № 3. — S. 41–52.
11. **Shenmayer V.V.** Approksimatsionnaya skhema dlya odnoy zadachi poiska podmnozhestva vektorov // Diskretnyy analiz i issledovanie operatsiy. — 2012. — T. 19, № 2. — S. 92–100.
12. **Papadimitriou C.H., Steiglitz K.** Combinatorial Optimization: Algorithms and Complexity.— Englewood Cliffs, N.J.: Prentice-Hall, 1982.
13. **Mollin R.A.** Fundamental Number Theory with Applications. Second Edition. — CRC Press, 2008.