ТРУДЫ ИНСТИТУТА МАТЕМАТИКИ И МЕХАНИКИ УрО РАН

Том 23 № 3 2017

УДК 519.16+519.85

ПРИБЛИЖЕННАЯ СХЕМА ДЛЯ ЗАДАЧИ ВЗВЕШЕННОЙ 2-КЛАСТЕРИЗАЦИИ С ФИКСИРОВАННЫМ ЦЕНТРОМ ОДНОГО КЛАСТЕРА[1]

А. В. Кельманов, А. В. Моткова, В. В. Шенмайер

Рассматривается одна из труднорешаемых задач разбиения конечного множества точек евклидова пространства на два кластера. Критерием решения является минимум суммы по обоим кластерам взвешенных сумм квадратов внутрикластерных расстояний от элементов кластеров до их центров. Центр одного из кластеров неизвестен и определяется как точка пространства, равная среднему значению элементов этого кластера (т. е. равная центроиду этого кластера). Центр другого кластера фиксирован в начале координат. Весовые множители для обеих внутрикластерных сумм заданы на входе. Предложен алгоритм приближенного решения задачи, основанный на адаптивном сеточном подходе поиска центра оптимального кластера. Показано, что алгоритм является полностью полиномиальной приближенной схемой (FPTAS) в случае фиксированной размерности пространства. В случае, когда размерность пространства не фиксирована, но ограничена медленно растущей функцией от мощности входного множества, алгоритм реализует полиномиальную аппроксимационную схему (PTAS).

Ключевые слова: евклидово пространство, кластеризация, NP-трудность, FPTAS, PTAS.

A. V. Kel'manov, A. V. Motkova, V. V. Shenmaier. Approximation scheme for the problem of weighted 2-partitioning with a fixed center of one cluster.

We consider the intractable problem of partitioning a finite set of points in Euclidean space into two clusters with minimum sum over the clusters of weighted sums of squared distances between the elements of the clusters and their centers. The center of one cluster is unknown and is defined as the mean value of its elements (i.e., it is the centroid of the cluster). The center of the other cluster is fixed at the origin. The weight factors for the intracluster sums are given as input. We present an approximation algorithm for this problem, which is based on the adaptive grid approach to finding the center of the optimal cluster. We show that the algorithm implements a fully polynomial-time approximation scheme (FPTAS) in the case of fixed space dimension. If the dimension is not fixed but is bounded by a slowly growing function of the number of input points, the algorithm realizes a polynomial-time approximation scheme (PTAS).

Keywords: Euclidean space, partitioning, NP-hardness, FPTAS, PTAS.

MSC: 68W25, 68Q25

DOI: 10.21538/0134-4889-2017-23-3-159-170

Введение

Предметом настоящего исследования является задача взвешенного разбиения конечного множества точек евклидова пространства на два кластера с фиксированным центром одного из кластеров. Цель исследования — обоснование аппроксимационной схемы.

Исследование мотивировано слабой изученностью рассматриваемой задачи в алгоритмическом плане и актуальностью ее приложений, таких как статистические проблемы совместного оценивания и проверки гипотез по неоднородным выборкам, проблемы кластерного анализа данных, проблемы интерпретации данных и др.

Статья развивает результаты из [1–3] по построению аппроксимационных схем и имеет следующую структуру. В разд. 1 приведены формулировка задачи, ее интерпретации и примеры приложений. Там же приведены основные известные результаты и анонсированы полученные алгоритмические результаты. В разд. 2 сформулированы и доказаны геометрические утверждения, обеспечивающие установление оценок качества (точности и временной сложности)

[1]Работа выполнена при поддержке РНФ (проект 16-11-10041).

предложенного алгоритма. В разд. 3 приведено описание алгоритма и показано, что при фиксированной размерности пространства он реализует полностью полиномиальную приближенную схему (FPTAS). Наконец, в разд. 4. предложена ускоренная модификация алгоритма и показано, что ускоренный алгоритм реализует полиномиальную приближенную схему (PTAS) в случае, когда размерность пространства является медленно растущей функцией от мощности входного множества.

1. Формулировка, интерпретация и приложения задачи, известные и полученные результаты

Всюду далее $\mathbb{R}$ — множество вещественных чисел, $\mathbb{R}_+$ — множество положительных вещественных чисел, $\mathbb{Z}$ — множество целых чисел, $\|\cdot\|$ — евклидова норма, $\langle\cdot,\cdot\rangle$ — скалярное произведение.

Рассматривается следующая задача.

З а д а ч а Weighted variance-based 2-clustering with given center. Дано: N-элементное множество $\mathcal{Y}$ точек из $\mathbb{R}^q$, натуральное число $M \leq N$ и два вещественных числа $w_1 > 0$ и $w_2 \geq 0$. Найти: разбиение множества $\mathcal{Y}$ на два кластера $\mathcal{C}$ и $\mathcal{Y}\setminus\mathcal{C}$ такое, что

$$F(\mathcal{C}) = w_1 \sum_{y\in\mathcal{C}} \|y - \overline{y}(\mathcal{C})\|^2 + w_2 \sum_{y\in\mathcal{Y}\setminus\mathcal{C}} \|y\|^2 \longrightarrow \min, \tag{1.1}$$

где $\overline{y}(\mathcal{C}) = \dfrac{1}{|\mathcal{C}|}\sum\limits_{y\in\mathcal{C}} y$ — геометрический центр (центроид) кластера $\mathcal{C}$ при ограничении $|\mathcal{C}| = M$.

В последнее десятилетие задача (1.1) привлекает живой интерес исследователей в области вычислительной сложности и аппроксимируемости труднорешаемых задач. По просьбе редакции журнала приводится краткий обзор известных нам результатов, непосредственно связанных с результатами, представленными в разд. 2–4 данной статьи.

Сформулированную задачу, очевидно, можно интерпретировать как задачу о взвешенном (весами w_1 и w_2) разбиении (кластеризации). Ранее исследовались следующие варианты этой задачи: (1) $\omega_1 = 1$ и $\omega_2 = 0$, (2) $\omega_1 = \omega_2 = 1$, (3) $\omega_1 = |\mathcal{C}|$ и $\omega_2 = N - |\mathcal{C}|$. Для указанных вариантов задачи к настоящему времени получен целый ряд результатов.

Первый вариант задачи ($\omega_1 = 1, \omega_2 = 0$) известен под названием M-Variance [4]. Он моделирует одну из простейших проблем анализа данных и распознавания образов — поиск (выбор) во множестве объектов совокупности похожих (близких) элементов. Сильная NP-трудность этого варианта задачи установлена в [5].

Точные алгоритмы с трудоемкостью $\mathcal{O}(qN^{q+1})$ предложены в [4;6]. Для случая, в котором размерность q пространства фиксирована, а координаты точек входного множества имеют целочисленные значения построен [7] точный псевдополиномиальный алгоритм. Его трудоемкость есть величина $\mathcal{O}(qN(2MB+1)^q)$, где B — максимальное абсолютное значение координаты точек входного множества. В [8] построен 2-приближенный полиномиальный алгоритм, имеющий временную сложность $\mathcal{O}(qN^2)$. Схема PTAS, позволяющая находить приближенное решение с относительной погрешностью $\varepsilon > 0$ за время $\mathcal{O}(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$, предложена в [9]. Факт несуществования схемы FPTAS установлен в [1], и там же такая схема предложена для случая, в котором размерность q пространства фиксирована. Эта схема позволяет находить $(1+\varepsilon)$-приближенное решение задачи для заданного $\varepsilon \in (0,1)$ за время $\mathcal{O}(N^2(M/\varepsilon)^q)$.

Во втором варианте, когда $\omega_1 = \omega_2 = 1$, требуется найти разбиение входного множества на два кластера, минимизирующее сумму двух внутрикластерных сумм. Первая из них — суммарный квадратичный разброс точек кластера $\mathcal{C}$ относительно неизвестного центроида. Вторая — суммарный квадратичный разброс точек кластера $\mathcal{Y}\setminus\mathcal{C}$ относительно начала координат. Этот вариант задачи индуцируется (см. [10–12]), в частности, одной из проблем помехоустойчивого кластерного анализа данных, когда известно, что в одном из кластеров $(\mathcal{Y} \setminus \mathcal{C})$ шумящие данные разбросаны (распределены) относительно заданной точки (известного среднего), которую без ограничения общности можно считать началом координат.

Сильная NP-трудность этого варианта задачи следует из результатов [10–13]. Напомним, что в этих работах была установлена NP-трудность в сильном смысле полиномиально эквивалентной задачи на максимум.

Для второго варианта задачи к настоящему времени предложены: точные алгоритмы [6; 14] с трудоемкостью $\mathcal{O}(qN^{q+1})$, $\mathcal{O}(q^2N^{2q})$; точные алгоритмы [13; 15; 16] для случая целочисленных входов с трудоемкостью $\mathcal{O}(Nq^{q+1}(MB)^{q-1})$, $\mathcal{O}(qMN(2MB)^{q-1})$ и $\mathcal{O}(qN(2MB + 1)^q)$, соответственно, где B — максимальное абсолютное значение координат входных точек; 2-приближенный полиномиальный алгоритм [17] с трудоемкостью $\mathcal{O}(qN^2)$; схема PTAS [18] с трудоемкостью $\mathcal{O}(qN^{2/\varepsilon+1}(9/\varepsilon)^{3/\varepsilon})$, где ε — относительная погрешность.

Кроме того, в [2] установлено, что для этого варианта задачи схема FPTAS не существует, и в этой же работе такая схема предложена для случая, в котором размерность q пространства фиксирована. Эта схема позволяет находить приближенное решение задачи за время $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$.

Наконец, в [19] предложен рандомизированный алгоритм, который для заданных $\varepsilon > 0$ и $\gamma \in (0, 1)$ при установленном значении параметра находит $(1 + \varepsilon)$-приближенное решение с вероятностью не менее $1-\gamma$ за время $\mathcal{O}(qN)$. Там же найдены условия, при которых алгоритм асимптотически точен и имеет трудоемкость $\mathcal{O}(qN^2)$.

Третий вариант задачи, когда $\omega_1 = |\mathcal{C}|$ и $\omega_2 = N - |\mathcal{C}|$, можно трактовать как взвешенное мощностями кластеров 2-разбиение точек входного множества. Этот вариант задачи, как и второй вариант, возникает в задачах помехоустойчивого кластерного анализа данных в ситуации, когда известен центр одного из кластеров, относительно которого распределены (разбросаны) неизвестные точки, а центроид другого кластера требуется найти (оценить) (см., например, [3; 20–22]). Сильная NP-трудность и факт несуществования схемы FPTAS доказаны в [21; 22], соответственно.

Для этого варианта в [20] был предложен точный алгоритм, ориентированный на случай целочисленных входов задачи; трудоемкость алгоритма есть величина $\mathcal{O}(qN(2MB + 1)^q)$, где B, как и выше, — максимальное абсолютное значение координат входных точек. Кроме того, в [3] обоснован приближенный алгоритм, который в случае фиксированной размерности пространства реализует схему FPTAS. Эта схема позволяет находить приближенное решение задачи за время $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$.

Общая задача взвешенного разбиения, рассматриваемая в настоящей статье, как и ее отмеченные варианты, относится к числу типичных проблем, возникающих, в частности, в анализе данных (Data analysis), распознавании образов (Pattern recognition), машинном обучении (Machine learning) и интерпретации данных (Data mining). В этих проблемах модели разбиения множеств на кластеры, как известно, играют ключевую роль (см., например, [23–27]).

В настоящей работе построен приближенный алгоритм, который применим к сформулированной обобщенной задаче с произвольными весами $w_1 > 0$ и $w_2 \geq 0$, а не только для вариантов задачи с весами, указанными выше. Для заданной относительной погрешности ε этот алгоритм позволяет находить $(1+\varepsilon)$-приближенное решение задачи за время $\mathcal{O}\big(qN^2(\sqrt{2q/\varepsilon}+2)^q\big)$. Кроме того, предложена модификация алгоритма с улучшенной трудоемкостью, оцениваемой величиной $\mathcal{O}\big(\sqrt{q}N^2(\pi e/2)^{q/2}(\sqrt{2/\varepsilon} + 2)^q\big)$. В случае, когда размерность пространства ограничена константой, обе версии алгоритма реализуют схему FPTAS с трудоемкостью $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$. В случае, когда размерность пространства ограничена величиной $C \log N$, где C — некоторая положительная константа, модифицированная версия алгоритма остается полиномиальной и реализует схему PTAS с трудоемкостью $\mathcal{O}\big(N^{C\,(1.05+\log(2+\sqrt{2/\varepsilon}))}\big)$.

2. Геометрические основы алгоритма

Для обоснования алгоритма нам потребуется несколько базовых утверждений.

Следующие две леммы относятся к числу хорошо известных (см., например, [1; 8]).

Лемма 1. *Пусть $\mathcal{Z}$ — конечное множество в $\mathbb{R}^q$, $\overline{z}$ — его центроид и x — произвольная точка в $\mathbb{R}^q$. Тогда*

$$\sum_{z\in\mathcal{Z}} \|z-x\|^2 = \sum_{z\in\mathcal{Z}} \|z-\overline{z}\|^2 + |\mathcal{Z}|\,\|x-\overline{z}\|^2.$$

Лемма 2. *Пусть $\mathcal{Z}$ — конечное множество в $\mathbb{R}^q$, $\overline{z}$ — его центроид и точка $u \in \mathbb{R}^q$ находится ближе к $\overline{z}$, чем любая точка из $\mathcal{Z}$. Тогда*

$$\sum_{z\in\mathcal{Z}} \|z-u\|^2 \le 2\sum_{z\in\mathcal{Z}} \|z-\overline{z}\|^2\ .$$

Лемма 3. *Пусть*

$$S(\mathcal{C},x) = w_1\sum_{y\in\mathcal{C}} \|y-x\|^2 + w_2\sum_{y\in\mathcal{Y}\setminus\mathcal{C}} \|y\|^2,$$

где $\mathcal{C} \subseteq \mathcal{Y}$ и $x \in \mathbb{R}^q$. Тогда справедливы следующие утверждения:

(1) *при любом непустом фиксированном подмножестве $\mathcal{C} \subseteq \mathcal{Y}$ минимум функции $S(\mathcal{C},x)$ по всем точкам $x \in \mathbb{R}^q$ достигается в точке $\overline{y}(\mathcal{C}) = 1/|\mathcal{C}| \sum\limits_{y\in\mathcal{C}} y$;*

(2) *при любой фиксированной точке $x \in \mathbb{R}^q$ минимум функции $S(\mathcal{C},x)$ по всем M-элементным подмножествам $\mathcal{C} \subseteq \mathcal{Y}$ достигается на подмножестве $\mathcal{B}^x$, состоящем из M точек множества $\mathcal{Y}$, в которых функция*

$$g^x(y) = (w_1-w_2)\,\|y\|^2 - 2w_1\,\langle y,x\rangle\,, \qquad y \in \mathcal{Y} \tag{2.1}$$

принимает наименьшие значения.

Д о к а з а т е л ь с т в о. Утверждение (1) является прямым следствием леммы 1 и определений функций S и F. Утверждение (2) основано на цепочке равенств

$$\begin{aligned} S(\mathcal{C},x) &= w_1\sum_{y\in\mathcal{C}} \|y-x\|^2 + w_2\sum_{y\in\mathcal{Y}\setminus\mathcal{C}} \|y\|^2 \\ &= w_1\sum_{y\in\mathcal{C}} \|y\|^2 - 2w_1\sum_{y\in\mathcal{C}} \langle y,x\rangle + Mw_1\|x\|^2 + w_2\Big(\sum_{y\in\mathcal{Y}} \|y\|^2 - \sum_{y\in\mathcal{C}} \|y\|^2\Big) \\ &= \sum_{y\in\mathcal{C}} \{(w_1-w_2)\|y\|^2 - 2w_1\langle y,x\rangle\} + Mw_1\|x\|^2 + w_2\sum_{y\in\mathcal{Y}} \|y\|^2. \end{aligned}$$

Остается заметить, что последние два слагаемых в последнем равенстве не зависят от $\mathcal{C}$.

Лемма 3 доказана.

Лемма 4. *Пусть $\mathcal{C}^*$ — оптимальное решение задачи и t — точка из множества $\mathcal{C}^*$, ближайшая к его центроиду. Тогда*

$$\|t-\overline{y}(\mathcal{C}^*)\|^2 \le \frac{1}{Mw_1}F(\mathcal{B}^t). \tag{2.2}$$

Д о к а з а т е л ь с т в о. В силу выбора точки t имеем $\|t-\overline{y}(\mathcal{C}^*)\|^2 \le \|y-\overline{y}(\mathcal{C}^*)\|^2$, где $y \in \mathcal{C}^*$. Суммируя это неравенство по всем $y \in \mathcal{C}^*$, получаем

$$M\|t-\overline{y}(\mathcal{C}^*)\|^2 \le \sum_{y\in\mathcal{C}^*} \|y-\overline{y}(\mathcal{C}^*)\|^2.$$

С другой стороны, из определения (1.1) и оптимальности $\mathcal{C}^*$ имеем

$$w_1 \sum_{y\in\mathcal{C}^*} \|y-\overline{y}(\mathcal{C}^*)\|^2 \le F(\mathcal{C}^*) \le F(\mathcal{B}^t).$$

Отсюда получаем утверждение леммы.

Лемма 4 доказана.

Лемма 5. *Если для произвольного $\varepsilon > 0$ и для некоторой точки $x \in \mathbb{R}^q$ справедливо неравенство*

$$\|x-\overline{y}(\mathcal{C}^*)\|^2 \le \frac{\varepsilon}{2Mw_1} F(\mathcal{B}^t), \tag{2.3}$$

то подмножество $\mathcal{B}^x$, определенное в лемме 3, является $(1+\varepsilon)$-приближенным решением задачи.

Доказательство. Согласно лемме 3 справедливы соотношения

$$F(\mathcal{B}^t) = S(\mathcal{B}^t, \overline{y}(\mathcal{B}^t)) \le S(\mathcal{B}^t, t) \le S(\mathcal{C}^*, t). \tag{2.4}$$

С другой стороны, в силу леммы 2 имеем $\sum_{y\in\mathcal{C}^*} \|y-t\|^2 \le 2 \sum_{y\in\mathcal{C}^*} \|y-\overline{y}(\mathcal{C}^*)\|^2$, а значит,

$$S(\mathcal{C}^*, t) \le 2F(\mathcal{C}^*). \tag{2.5}$$

Объединяя (2.3)–(2.5), получим

$$\|x-\overline{y}(\mathcal{C}^*)\|^2 \le \frac{\varepsilon}{2Mw_1} F(\mathcal{B}^t) \le \frac{\varepsilon}{2Mw_1} S(\mathcal{C}^*, t) \le \frac{\varepsilon}{Mw_1} F(\mathcal{C}^*).$$

Отсюда и из лемм 3 и 1 вытекает цепочка неравенств

$$F(\mathcal{B}^x) = S(\mathcal{B}^x, \overline{y}(\mathcal{B}^x)) \le S(\mathcal{B}^x, x) \le S(\mathcal{C}^*, x) \le F(\mathcal{C}^*) + Mw_1 \|x-\overline{y}(\mathcal{C}^*)\|^2 \le (1+\varepsilon)F(\mathcal{C}^*).$$

Лемма 5 доказана.

3. Приближенный алгоритм

Суть предлагаемого подхода к поиску приближенного решения задачи состоит в следующем. Для каждой точки входного множества строится область (кубической формы) так, чтобы одна из построенных областей гарантированно включала центроид искомого подмножества. По заданной на входе желаемой относительной погрешности решения строится решетка, дискретизирующая указанную область с равномерным по всем координатам шагом. Размер и шаг решетки вычисляются адаптивно (см. далее) для каждой из входных точек. Для каждого узла решетки формируется подмножество из M точек входного множества, имеющих наименьшие значения функции (2.1). Сформированное подмножество объявляется претендентом на решение. В качестве окончательного решения выбирается тот претендент (подмножество), для которого значение целевой функции задачи минимально.

Напомним, что этот по своей сути адаптивный сеточный подход ранее уже применялся в [1–3] для решения трех упомянутых в разд. 1 вариантов задачи. Настоящая работа демонстрирует результативность адаптивного сеточного подхода к решению обобщенной задачи.

Для произвольной точки $x \in \mathbb{R}^q$ и положительных чисел h и H определим множество точек

$$\mathcal{G}(x,h,H) = \{d \in \mathbb{R}^q \mid d = x + h(i_1,\dots,i_q),\ i_k \in \mathbb{Z},\ |hi_k| \le H,\ k \in \{1,\dots,q\}\} \tag{3.1}$$

— многомерную решетку кубической формы размера $2H$ с центром в точке x и покоординатным шагом h между узлами.

З а м е ч а н и е 1. Если для произвольных точек x и $z \in \mathbb{R}^q$ верно, что $\|z - x\| \leq H$, то расстояние от z до ближайшего узла решетки $\mathcal{G}(x, h, H + h/2)$ не превосходит $h\sqrt{q}/2$.

Для построения алгоритмического решения нам потребуется для каждой точки y входного множества $\mathcal{Y}$ адаптивно определять полуразмер H решетки и ее шаг h таким образом, чтобы область решетки гарантированно включала центроид искомого подмножества, а шаг решетки определялся заданной относительной погрешностью ε. В связи с этим положим

$$H(y) = \sqrt{\frac{1}{Mw_1} F(\mathcal{B}^y)}, \qquad y \in \mathcal{Y}, \tag{3.2}$$

$$h(y, \varepsilon) = \sqrt{\frac{2\varepsilon}{qMw_1} F(\mathcal{B}^y)}, \qquad y \in \mathcal{Y}, \qquad \varepsilon \in \mathbb{R}_+, \tag{3.3}$$

где $\mathcal{B}^y$ — множество, определенное в лемме 3.

З а м е ч а н и е 2. Для произвольной точки $y \in \mathcal{Y}$ мощность решетки $\mathcal{G}(y, h, H + h/2)$ не превосходит значения

$$L = \Big(2\Big\lfloor \frac{H + h/2}{h} \Big\rfloor + 1\Big)^q \leq \Big(2\frac{H}{h} + 2\Big)^q = \Big(\sqrt{\frac{2q}{\varepsilon}} + 2\Big)^q$$

в силу (3.2) и (3.3).

Приведем пошаговую запись алгоритма.

А л г о р и т м $\mathcal{A}$

Вход алгоритма: N-элементное множество $\mathcal{Y} \subset \mathbb{R}^q$, натуральное число $M \leq N$ и вещественные числа $w_1 > 0$, $w_2 \geq 0$ и $\varepsilon \in (0, 1)$.

Для каждой точки $y \in \mathcal{Y}$ выполним шаги 1–5.

Ш а г 1. Вычислим значения $g^y(z)$, $z \in \mathcal{Y}$, по формуле (2.1); найдем M-элементное подмножество $\mathcal{B}^y \subseteq \mathcal{Y}$ с наименьшими значениями $g^y(z)$, вычислим значение $F(\mathcal{B}^y)$ по формуле (1.1).

Ш а г 2. Если $F(\mathcal{B}^y) = 0$, то положим $\mathcal{C}_{\mathcal{A}} = \mathcal{B}^y$; выход.

Ш а г 3. Вычислим значения H и h по формулам (3.2) и (3.3) соответственно.

Ш а г 4. Построим решетку $\mathcal{G}(y, h, H + h/2)$ по формуле (3.1).

Ш а г 5. Для каждого узла x решетки $\mathcal{G}(y, h, H + h/2)$ вычислим значения $g^x(y)$, $y \in \mathcal{Y}$, по формуле (2.1) и найдем M-элементное подмножество $\mathcal{B}^x \subseteq \mathcal{Y}$ с наименьшими значениями $g^x(y)$. Вычислим значение $F(\mathcal{B}^x)$ по формуле (1.1); запомним это значение и множество $\mathcal{B}^x$.

Ш а г 6. В семействе $\{\mathcal{B}^x \mid x \in \mathcal{G}(y, h, H + h/2),\ y \in \mathcal{Y}\}$ допустимых множеств, построенных на шагах 1–5, выберем в качестве решения $\mathcal{C}_{\mathcal{A}}$ то множество $\mathcal{B}^x$, для которого значение $F(\mathcal{B}^x)$ минимально.

Выход алгоритма: множество $\mathcal{C}_{\mathcal{A}}$.

Теорема 1. *Для любого* $\varepsilon \in (0, 1)$ *алгоритм* $\mathcal{A}$ *находит* $(1 + \varepsilon)$-*приближенное решение задачи за время* $\mathcal{O}\big(qN^2(\sqrt{2q/\varepsilon} + 2)^q\big)$.

Д о к а з а т е л ь с т в о. Оценим точность алгоритма. Очевидно, что один из циклов алгоритма будет выполнен для точки $t \in \mathcal{Y}$, ближайшей к центроиду оптимального множества $\mathcal{C}^*$. По лемме 4 для этой точки будет выполнено неравенство (2.2), которое в соответствии с (3.2) означает, что $\|t - \overline{y}(\mathcal{C}^*)\| \leq H(t)$. Следовательно, центроид оптимального подмножества лежит в кубе с ребром $2H(t)$ и центром в t.

Пусть x^* — узел решетки $\mathcal{G}(t, h, H + h/2)$, ближайший к центроиду $\mathcal{C}^*$. В соответствии с замечанием 1 квадрат расстояния от оптимального центроида $\overline{y}(\mathcal{C}^*)$ до ближайшего узла x^* решетки не превосходит $h^2q/4$. Поэтому

$$\|x^* - \overline{y}(\mathcal{C}^*)\|^2 \leq \frac{h^2 q}{4} = \frac{\varepsilon}{2Mw_1} F(\mathcal{B}^t). \tag{3.4}$$

Следовательно, точка x^* удовлетворяет условиям леммы 5, а соответствующее ей подмножество $\mathcal{B}^{x^*}$ является $(1+\varepsilon)$-приближенным решением задачи.

Оценим временную сложность алгоритма. На шаге 1 для вычисления значений $g^y(z)$ требуется не более $\mathcal{O}(qN)$ операций. Для поиска M наименьших элементов во множестве из N элементов потребуется $\mathcal{O}(N)$ операций (например, с помощью алгоритма отыскания n-го наименьшего значения в неупорядоченном массиве [28]). Вычисление значения $F(\mathcal{B}^y)$ выполняется за время $\mathcal{O}(qN)$. Шаги 2 и 3 выполняются за время $\mathcal{O}(q)$. Для построения решетки на шаге 4 требуется $\mathcal{O}(qL)$ операций (по замечанию 2). На шаге 5 вычисление элементов множества $\mathcal{B}^x$ для каждого узла x решетки выполняется за время $\mathcal{O}(qN)$, как и вычисление значения $F(\mathcal{B}^x)$ (по аналогии с вычислениями на шаге 1). Поэтому суммарное время вычислений для всех узлов решетки на этом шаге равно $\mathcal{O}(qNL)$. Наконец, поскольку шаги 1–5 выполняются N раз, трудоемкость выполнения этих шагов равна $\mathcal{O}(qN^2L)$. Трудоемкость шага 6 оценивается величиной $\mathcal{O}(NL)$. В итоге суммарные затраты на всех шагах равны $\mathcal{O}(qN^2L) = \mathcal{O}\big(qN^2(\sqrt{2q/\varepsilon}+2)^q\big)$.

Теорема 1 доказана.

Замечание 3. Легко видеть, что в случае фиксированной размерности пространства трудоемкость алгоритма $\mathcal{A}$ оценивается величиной $\mathcal{O}(N^2(1/\varepsilon)^{q/2})$ и он реализует схему FPTAS.

4. Ускоренный алгоритм

Предложенный выше алгоритм можно ускорить путем исключения из процесса вычислений значительной части узлов адаптивно строящихся решеток кубической формы. Действительно, поскольку центроид оптимального кластера лежит на расстоянии, не превосходящем некоторый порог H, от одной из точек входного множества, достаточно рассматривать только те узлы построенных решеток, что расположены на расстоянии, не превышающем H от их центров, с небольшим запасом, который определен ниже в лемме 6.

Для каждого $y \in \mathcal{Y}$ положим $R = H + h\sqrt{q}/2$, где $H = H(y)$, $h = h(y,\varepsilon)$ — параметры решетки, определенные по формулам (3.2) и (3.3) соответственно. Построим "сокращенную" решетку сферической формы: $\mathcal{G}_R(y,h,H+h/2) = \mathcal{G}(y,h,H+h/2)\cap B(y,R)$, где $B(y,R) = \{x \in \mathbb{R}^q \mid \|x-y\| \le R\}$ — шар радиуса R с центром в y.

Справедливы следующие утверждения.

Лемма 6. *Пусть x — произвольная точка шара $B(y,H)$, $y \in \mathcal{Y}$. Тогда ближайший к x узел решетки $\mathcal{G}_R(y,h,H+h/2)$ расположен на расстоянии от x, не превышающем $h\sqrt{q}/2$.*

Доказательство. Пусть z — ближайший к x узел решетки $\mathcal{G}(y,h,H+h/2)$. Тогда в соответствии с замечанием 1 справедливо неравенство $\|z-x\| \le h\sqrt{q}/2$. С другой стороны, $\|x-y\| \le H$, отсюда и из неравенства треугольника получаем $\|z-y\| \le H + h\sqrt{q}/2 = R$. Следовательно, $z \in \mathcal{G}_R(y,h,H+h/2)$.

Лемма 6 доказана.

Лемма 7. *Для любого $y \in \mathcal{Y}$ мощность множества $\mathcal{G}_R(y,h,H+h/2)$ не превосходит величины*

$$\frac{1}{\sqrt{\pi q}}\Big(\frac{2\pi e}{q}\Big)^{q/2}\Big(\frac{H}{h}+\sqrt{q}\Big)^q.$$

Доказательство. Поскольку каждый узел z решетки $\mathcal{G}_R(y,h,H+h/2)$ лежит внутри шара $B(y,R)$, то согласно неравенству треугольника все точки q-мерного куба со стороной h с центром в z лежат внутри шара $B(y,R+h\sqrt{q}/2)$. Следовательно, суммарный объем всех таких кубов не превышает объема q-мерного шара радиуса $R+h\sqrt{q}/2 = H+h\sqrt{q}$. Отсюда $L_R h^q \le V_q(H+h\sqrt{q})^q$, где L_R — мощность множества $\mathcal{G}_R(y,h,H+h/2)$, а V_q — объем q-мерного единичного шара, оцениваемый по общеизвестной формуле $V_q \le 1/\sqrt{\pi q}\,\Big(\dfrac{2\pi e}{q}\Big)^{q/2}$

(см. например, [29]). Сопоставляя два последних неравенства, получаем заявленную в лемме оценку.

Лемма 7 доказана.

Обозначим через $\mathcal{A}_R$ алгоритм, отличающийся от алгоритма $\mathcal{A}$ тем, что в описании шагов 4, 5 вместо решеток $\mathcal{G}(y, h, H + h/2)$ рассматриваются решетки $\mathcal{G}_R(y, h, H + h/2)$. Тогда справедлива следующая теорема.

Теорема 2. *Для любого* $\varepsilon \in (0, 1)$ *алгоритм* $\mathcal{A}_R$ *находит* $(1 + \varepsilon)$*-приближенное решение задачи за время*

$$\mathcal{O}\Big(\sqrt{q}N^2\Big(\frac{\pi e}{2}\Big)^{q/2}\Big(\sqrt{\frac{2}{\varepsilon}} + 2\Big)^q\Big). \tag{4.1}$$

Д о к а з а т е л ь с т в о. Согласно равенству $H/h = (1/\sqrt{2})\sqrt{q/\varepsilon}$ и лемме 7 получаем

$$L_R \le \frac{1}{\sqrt{\pi q}}\Big(\frac{2\pi e}{q}\Big)^{q/2}\Big(\frac{1}{\sqrt{2}}\sqrt{\frac{q}{\varepsilon}} + \sqrt{q}\Big)^q = \frac{1}{\sqrt{\pi q}}\Big(\frac{\pi e}{2}\Big)^{q/2}\Big(\sqrt{\frac{2}{\varepsilon}} + 2\Big)^q.$$

Отсюда с учетом оценок трудоемкости шагов 1–6 алгоритма $\mathcal{A}$, установленных в доказательстве теоремы 1, получаем оценку (4.1) трудоемкости алгоритма $\mathcal{A}_R$.

Остается заметить, что оценки точности обоих алгоритмов совпадают. Действительно, для некоторого $y \in \mathcal{Y}$ центроид оптимального кластера лежит в шаре $B(y, H)$, а значит, согласно лемме 6 находится на расстоянии, не превышающем $h\sqrt{q}/2$, от одного из узлов "сокращенной" решетки $\mathcal{G}_R(y, h, H + h/2)$. Тем самым выполняется соотношение (3.4), откуда следует, что алгоритм находит $(1 + \varepsilon)$-приближенное решение задачи.

Теорема 2 доказана.

З а м е ч а н и е 4. В связи с тем что $\mathcal{G}_R(y, h, H + h/2) \subset \mathcal{G}(y, h, H + h/2)$, время работы алгоритма $\mathcal{A}_R$ при любых N, q и ε строго меньше времени работы алгоритма $\mathcal{A}$.

З а м е ч а н и е 5. Алгоритм $\mathcal{A}_R$ полиномиален не только в случае фиксированной размерности q пространства, но и в случае, когда размерность пространства ограничена величиной $C \log N$, где C — некоторая положительная константа. В этом случае согласно теореме 2 алгоритм находит $(1 + \varepsilon)$-приближенное решение задачи за время $O(N^d \log N)$, где

$$d = \frac{C}{2}\log\frac{\pi e}{2} + C\log\Big(2 + \sqrt{\frac{2}{\varepsilon}}\Big) < C\,\Big(1.05 + \log\Big(2 + \sqrt{\frac{2}{\varepsilon}}\Big)\Big).$$

Таким образом, в указанном случае алгоритм реализует схему PTAS.

Заключение

В работе построен приближенный алгоритм сеточного типа для решения квадратичной евклидовой задачи взвешенной 2-кластеризации конечного множества точек при заданном центре одного из кластеров. Показано, что алгоритм является полностью полиномиальной аппроксимационной схемой, если размерность пространства ограничена константой. Алгоритм остается полиномиальным, даже если размерность пространства не фиксирована, но ограничена величиной $\mathcal{O}(\log N)$, медленно растущей с ростом мощности входного множества точек. Актуальность этого случая объясняется тем, что размерность пространства $\mathcal{O}(\log N)$ является минимальной; при данной размерности возможно существование N-элементного множества точек с координатами из фиксированного конечного набора значений.

Рассмотренная задача кластеризации относится к числу слабоизученных проблем дискретной оптимизации. Продолжение исследований этой задачи представляется важным делом ближайшей перспективы.

СПИСОК ЛИТЕРАТУРЫ

1. **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.
2. **Кельманов А. В., Хандеев В. И.** Полностью полиномиальная аппроксимационная схема для специального случая одной квадратичной евклидовой задачи 2-кластеризации // Журн. вычисл. математики и мат. физики. 2016. Т. 56, № 2. С. 332–340. doi: 10.7868/S0044466916020113.
3. **Kel'manov A. V., Motkova A. V.** A fully polynomial-time approximation scheme for a special case of a balanced 2-clustering problem // Proc. Internat. Conf. on Discrete Optimization and Operations Research (DOOR 2016). 2016. P. 182–192. (Lecture Notes Comp. Sci.; vol. 9869.) doi: 10.1007/978-3-319-44914-2-15.
4. Finding k points with minimum diameter and related problems / A. Aggarwal, H. Imai, N. Katoh, S. Suri // J. Algorithms. 1991. Vol. 12, № 1. P. 38–56. doi: 10.1016/0196-6774(91)90022-Q.
5. **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.
6. **Шенмайер В. В.** Решение некоторых задач поиска подмножества векторов с использованием диаграмм Вороного // Дискретный анализ и исследование операций. 2016. Т. 23, № 4. С. 102–115. doi: 10.17377/daio.2016.23.526.
7. **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.
8. **Кельманов А. В., Романченко С. М.** Приближенный алгоритм для решения одной задачи поиска подмножества векторов // Дискретный анализ и исследование операций. 2011. Т. 18, № 1. С. 61–69.
9. **Шенмайер В. В.** Аппроксимационная схема для одной задачи поиска подмножества векторов // Дискретный анализ и исследование операций. 2012. Т. 19, № 2. С. 92–100.
10. Апостериорное обнаружение в числовой последовательности квазипериодического фрагмента при заданном числе повторов / Э. Х. Гимади, А. В. Кельманов, М. А. Кельманова, С. А. Хамидуллин // Сиб. журн. индустр. математики. 2006. Т. 9, № 1(25). С. 55–74.
11. A posteriori detecting a quasiperiodic fragment in a numerical sequence / E. Kh. Gimadi, A. V. Kel'manov, M. A. Kel'manova, S. A. Khamidullin // Pattern Recognition and Image Analysis. 2008. Vol. 18, № 1. P. 30–42. doi:10.1134/S1054661808010057.
12. **Кельманов А. В., Пяткин А. В.** О сложности некоторых задач поиска подмножеств векторов и кластерного анализа // Журн. вычисл. математики и мат. физики. 2009. Т. 49, № 11. С. 2059–2065.
13. Задача отыскания подмножества векторов с максимальным суммарным весом / А. Е. Бабурин, Э. Х. Гимади, Н. И. Глебов, А. В. Пяткин // Дискретный анализ и исследование операций. 2007. Т. 14, № 1. С. 32–42.
14. **Гимади Э. Х., Пяткин А. В., Рыков И. А.** О полиномиальной разрешимости некоторых задач выбора подмножества векторов в евклидовом пространстве фиксированной размерности // Дискретный анализ и исследование операций. 2008. Т. 15, № 6. С. 11–19.
15. **Гимади Э. Х., Глазков Ю. В., Рыков И. А.** О двух задачах выбора подмножества векторов с целочисленными координатами в евклидовом пространстве с максимальной нормой суммы // Дискретный анализ и исследование операций. 2008. Т. 15, № 4. С. 30–43.
16. **Кельманов А. В., Хандеев В. И.** Точный псевдополиномиальный алгоритм для одной задачи двухкластерного разбиения множества векторов // Дискретный анализ и исследование операций. 2015. Т. 22, № 4. С. 50–62. doi: 10.17377/daio.2015.22.463.
17. **Долгушев А. В., Кельманов А. В.** Приближенный алгоритм решения одной задачи кластерного анализа // Дискретный анализ и исследование операций. 2011. Т. 18, № 2. С. 29–40.
18. **Долгушев А. В., Кельманов А. В., Шенмайер В. В.** Полиномиальная аппроксимационная схема для одной задачи разбиения конечного множества на два кластера // Тр. Ин-та математики и механики УрО РАН. 2015. Т. 21, № 3. С 100–109.
19. **Кельманов А. В., Хандеев В. И.** Рандомизированный алгоритм для одной задачи двухкластерного разбиения множества векторов // Журн. вычисл. математики и мат. физики. 2015. Т. 55, № 2. С. 335–344. doi: 10.7868/S0044466915020131.

20. **Кельманов А. В., Моткова А. В.** Точные псевдополиномиальные алгоритмы для задачи сбалансированной 2-кластеризации // Дискретный анализ и исследование операций. 2016. Т. 23, № 3. С. 21–34. doi: 10.17377/daio.2016.23.520
21. **Кельманов А. В., Пяткин А. В.** NP-трудность некоторых квадратичных евклидовых задач 2-кластеризации // Докл. РАН. 2015. Т. 464, № 5. С. 535–538. doi: 10.7868/S0869565215290058
22. **Кельманов А. В., Пяткин А. В.** О сложности некоторых квадратичных евклидовых задач 2-кластеризации // Журн. вычисл. математики и мат. физики. 2016. Т. 56, № 3. С. 498–504. doi: 0.7868/S0044466916030091 .
23. **Aggarwal C. C.** Data Mining: The Textbook. Cham: Springer International Publishing, 2015. 734 p. ISBN: 978-3319141411 .
24. **Bishop C. M.** Pattern Recognition and Machine Learning. New York: Springer Science+Business Media, LLC, 2006. 738 p. ISBN: 978-0-387-31073-2 .
25. **Hastie T., Tibshirani R., Friedman J.** The Elements of statistical learning: Data mining, inference, and prediction. New York: Springer-Verlag, 2009. 763 p. doi: 10.1007/978-0-387-84858-7 .
26. An introduction to statistical learning / G. James, D. Witten, T. Hastie, R. Tibshirani. New York: Springer Science+Business Media, LLC, 2013. 426 p. doi: 10.1007/978-1-4614-7138-7 .
27. **Jain A. K.** Data clustering: 50 years beyond k-means // Pattern Recognition Lett. 2010. Vol. 31. P. 651–666. doi: 10.1016/j.patrec.2009.09.011 .
28. **Wirth N.** Algorithms + data structures = programs. New Jersey: Prentice Hall, 1976. 366 p. ISBN: 0130224189 .
29. **Ball K.** An elementary introduction to modern convex geometry. Flavors of geometry // MSRI Publications / ed. S. Levi. 1997. Vol. 31. P. 1–58. ISBN: 0-521-62048-1 .

Кельманов Александр Васильевич
д-р физ.-мат. наук
зав. лабораторией
Институт математики им. С. Л. Соболева СО РАН,
Новосибирский государственный университет,
г. Новосибирск
e-mail: kelm@math.nsc.ru

Поступила 24.05.2017

Моткова Анна Владимировна
студент
Новосибирский государственный университет
г. Новосибирск,
e-mail: anitamo@mail.ru

Шенмайер Владимир Владимирович
канд. физ.-мат. наук
старший науч. сотрудник
Институт математики им. С. Л. Соболева СО РАН,
г. Новосибирск
e-mail: shenmaier@mail.ru

REFERENCES

1. Kel'manov A. V., Romanchenko S. M. An FPTAS for a vector subset search problem. *J. Appl. Ind. Math.*, 2014, vol. 8, no. 3, pp. 329–336. doi: 10.1134/S1990478914030041 .
2. Kel'manov A. V., Khandeev V. I. Fully polynomial-time approximation scheme for a special case of a quadratic euclidean 2-clustering problem. *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 2, pp. 334–341. doi: 10.1134/S0965542516020111 .
3. Kel'manov A. V., Motkova A. V. A fully polynomial-time approximation scheme for a special case of a balanced 2-clustering problem. Proc. Internat. Conf. on Discrete Optimization and Operations Research (DOOR 2016), 2016, *Ser. Lecture Notes Comp. Sci.*, vol. 9869, pp. 182–192. doi: 10.1007/978-3-319-44914-2−15 .

4. Aggarwal A., Imai H., Katoh N., Suri S. Finding k points with minimum diameter and related problems. *J. Algorithms*, 1991, vol. 12, no. 1, pp. 38–56. doi: 10.1016/0196-6774(91)90022-Q .
5. Kel'manov A. V., Pyatkin A. V. NP-completeness of some problems of choosing a vector subset. *J. Appl. Ind. Math.*, 2011, vol. 5, no. 3, pp. 352–357. doi: 10.1134/S1990478911030069 .
6. Shenmaier V. V. Solving some vector subset problems by Voronoi diagrams. *J. Appl. Industr. Math.*, 2016, vol. 10, no. 4, pp. 560–566. doi: 10.1134/S199047891604013X .
7. Kel'manov A. V., Romanchenko S. M. Pseudopolynomial algorithms for certain computationally hard vector subset and cluster analysis problems. *Automation and Remote Control*, 2012, vol. 73, no. 2, pp. 349–354. doi: 10.1134/S0005117912020129 .
8. Kel'manov A. V., Romanchenko S. M. An approximation algorithm for solving a problem of search for a vector subset. *J. Appl. Ind. Math.*, 2012, vol. 6, no. 1, pp. 90–96. doi: 10.1134/S1990478912010097 .
9. Shenmaier V. V. An approximation scheme for a problem of search for a vector subset. *J. Appl. Ind. Math.*, 2012, vol. 6, no. 3, pp. 381–386. doi: 10.1134/S0081543816090066 .
10. Gimadi E.Kh., Kel'manov A.V., Kelmanova M.A., Khamidullin S.A. A posteriori detection of a quasiperiodic fragment with a given number of repetitions in a numerical sequence. *Sib. Zh. Ind. Mat.*, 2006, vol. 9, no. 1, pp. 55–74 (in Russian).
11. Gimadi E. Kh., Kel'manov A. V., Kel'manova M. A., Khamidullin S. A. A posteriori detecting a quasiperiodic fragment in a numerical sequence. *Pattern Recognition and Image Analysis*, 2008, vol. 18, no. 1, pp. 30–42. doi:10.1134/S1054661808010057 .
12. Kel'manov A. V., Pyatkin A. V. Complexity of certain problems of searching for subsets of vectors and cluster analysis. *Comput. Math. Math. Phys.*, 2009, vol. 49, no. 11, pp. 1966–1971. doi: 10.1134/S0965542509110128 .
13. Baburin A.E., Gimadi E.Kh., Glebov, N.I., Pyatkin, A.V. The problem of finding a subset of vectors with the maximum total weight. *J. Appl. Industr. Math.*, 2008, vol. 2, no. 1, pp. 32–38. doi: 10.1007/s11754-008-1004-3 .
14. Gimadi E. Kh., Pyatkin A. V., Rykov I. A. On polynomial solvability of some problems of a vector subset choice in a Euclidean space of fixed dimension. *J. Appl. Indust. Math.*, 2010, vol. 4, no. 1, pp. 48–53. doi: 10.1134/S1990478910010084 .
15. Gimadi E.K., Glazkov Y.V., Rykov I.A. On two problems of choosing some subset of vectors with integer coordinates that has maximum norm of the sum of elements in Euclidean space. *J. Appl. Ind. Math.*, 2009, vol. 3, no. 3, pp. 343–352. doi: 10.1134/S1990478909030041 .
16. Kel'manov A. V., Khandeev V. I. An exact pseudopolynomial algorithm for a problem of the two-cluster partitioning of a set of vectors. *J. Appl. Ind. Math.*, 2015, vol. 9, no. 4, pp. 497–502. doi: 10.1134/S1990478915040067 .
17. Dolgushev A. V., Kel'manov A. V. An approximation algorithm for solving a problem of cluster analysis. *J. Appl. Indust. Math.*, 2011, vol. 5, no. 4, pp. 551–558. doi: 10.1134/S1990478911040107 .
18. A.V. Dolgushev, A.V. Kel'manov, V.V. Shenmaier. Polynomial-time approximation scheme for a problem of partitioning a finite set into two clusters. *Proc. Steklov Inst. Math.*, 2016, vol. 295, suppl. 1, pp. 47–56. doi: 10.1134/S0081543816090066 .
19. Kel'manov A. V., Khandeev V. I. A Randomized algorithm for two-cluster partition of a set of vectors. *Comput. Math. Math. Phys.*, 2015, vol. 55, no. 2, pp. 330–339. doi: 10.1134/S096554251502013X .
20. Kel'manov A. V., Motkova A. V. *J. Appl. Ind. Math.*, 2016, vol. 10, no. 3, pp. 349–355. doi: 10.1134/S1990478916030054 .
21. Kel'manov A. V., Pyatkin A. V. NP-hardness of some quadratic euclidean 2-clustering problems. *Dokl. Math.*, 2015, vol. 92, no. 2, pp. 634–637. doi: 10.1134/S1064562415050233 .
22. Kel'manov A. V., Pyatkin A. V. On the complexity of some quadratic euclidean 2-clustering problems. *Comput. Math. Math. Phys.*, 2016, vol. 56, no. 3, pp. 491–497. doi: 10.1134/S096554251603009X .
23. Aggarwal C. C. *Data Mining: The Textbook.* Cham, Springer International Publishing, 2015, 734 p. ISBN: 978-3319141411 .
24. Bishop C. M. *Pattern Recognition and Machine Learning.* New York: Springer Science+Business Media, LLC, 2006. 738 p. ISBN: 978-0-387-31073-2 .
25. Hastie T., Tibshirani R., Friedman J. *The Elements of statistical learning: Data mining, inference, and prediction.* New York: Springer-Verlag. 2009. 763 p. doi: 10.1007/978-0-387-84858-7 .
26. James G., Witten D., Hastie T., Tibshirani R. *An introduction to statistical learning.* New York: Springer Science+Business Media, LLC, 2013, 426 p. doi: 10.1007/978-1-4614-7138-7 .
27. Jain A. K. Data clustering: 50 years beyond k-means. *Pattern Recognition Lett.*, 2010, vol. 31, no. 8, pp. 651–666. doi: 10.1016/j.patrec.2009.09.011 .

28. Wirth N. *Algorithms + data structures = programs.* New Jersey: Prentice Hall, 1976, 366 p. ISBN: 0130224189 .
29. Ball K. An elementary introduction to modern convex geometry. Flavors of geometry. *MSRI Publications*, S. Levy editor, 1997, vol. 31, pp. 1–58. ISBN: 0-521-62048-1 .

The paper was received by the Editorial Office on May 24, 2017.

Aleksander Vasil'evich Kel'manov, Dr. Phys.-Math. Sci., Sobolev Institute of Mathematics; Novosibirsk State University, Novosibirsk, 630990 Russia, e-mail: kelm@math.nsc.ru .

Anna Vladimirovna Motkova, undergraduate student, Novosibirsk State University, Novosibirsk, 630990 Russia, e-mail: anitamo@mail.ru .

Vladimir Vladimirovich Shenmaier, Cand. Sci. (Phys.-Math.), Sobolev Institute of Mathematics, Novosibirsk, 630990 Russia, e-mail: shenmaier@mail.ru .