ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Октябрь—декабрь 2016. Том 23, № 4. С. 102–115

УДК 519.176 DOI: 10.17377/daio.2016.23.526

РЕШЕНИЕ НЕКОТОРЫХ ЗАДАЧ ПОИСКА ПОДМНОЖЕСТВА ВЕКТОРОВ С ИСПОЛЬЗОВАНИЕМ ДИАГРАММ ВОРОНОГО *)

В. В. Шенмайер[1]
[1]Институт математики им. С. Л. Соболева СО РАН,
пр. Акад. Коптюга, 4, 630090 Новосибирск, Россия
e-mail: shenmaier@mail.ru

Аннотация. Предлагается общий подход к решению некоторых задач поиска подмножества векторов в евклидовом пространстве, основанный на использовании диаграмм Вороного высших порядков. В случае фиксированной размерности пространства данный подход позволяет находить оптимальные решения этих задач за полиномиальное время, меньшее, чем время работы известных ранее алгоритмов. Ил. 1, библиогр. 16.

Ключевые слова: вычислительная геометрия, поиск подмножества векторов, евклидово пространство, диаграмма Вороного, полиномиальный алгоритм.

Введение

Рассматриваются задачи поиска подмножества векторов, принадлежащих заданному конечному множеству в пространстве $\mathbb{R}^d$ и обладающих свойством максимальной совокупной геометрической «близости» (в том или ином смысле). Известно, что в случае фиксированной размерности пространства многие подобные задачи могут быть эффективно решены с помощью диаграмм Вороного высших порядков [11, 13, 16]. В статье доказана возможность применения этого инструмента к некоторым задачам, для которых он ранее не использовался. Предложенный подход позволяет улучшить алгоритмические результаты для этих задач.

Постановки задач. В общем виде задачу поиска подмножества векторов можно сформулировать следующим образом.

*)Исследование выполнено при финансовой поддержке Российского научного фонда (проект 16–11–10041).

Задача Vector Subset. Дано: n-элементное множество X в евклидовом пространстве $\mathbb{R}^d$, целевая вещественная функция f, определённая для любого подмножества $S \subseteq X$, и предикат Γ, определяющий, является ли подмножество S допустимым решением. Найти: допустимое подмножество $S^* \subseteq X$, на котором достигается оптимум (минимум или максимум) функции f:

$$f(S^*) = \mathrm{opt}\{f(S) \mid S \subseteq X,\, \Gamma(S)\}.$$

Более подробно остановимся на следующих частных случаях этой задачи. Для первых трёх из них входом задачи являются конечное множество X в евклидовом пространстве $\mathbb{R}^d$ и целое число k.

Задача 1. Найти: k-элементное подмножество $S \subseteq X$, на котором достигается минимум функции $f_1(S) = \sum\limits_{x \in S} \|x - c(S)\|^2$, где $c(S) = \frac{1}{k} \sum\limits_{x \in S} x$ — среднее множества S.

Задача 2. Найти: k-элементное подмножество $S \subseteq X$, на котором достигается минимум функции $f_2(S) = \sum\limits_{x \in S} \|x - c(S)\|^2 + \sum\limits_{x \in X \setminus S} \|x\|^2$.

Задача 3. Найти: k-элементное подмножество $S \subseteq X$, на котором достигается максимум функции $f_3(S) = \|\sum\limits_{x \in S} x\|$.

Задача 4. Дано: конечное множество X в евклидовом пространстве $\mathbb{R}^d$. Найти: непустое подмножество $S \subseteq X$, на котором достигается максимум функции $f_4(S) = \|\sum\limits_{x \in S} x\|^2 \,/\, |S|$.

Задача 5. Дано: два непересекающихся конечных множества X и Y в евклидовом пространстве $\mathbb{R}^d$. Найти: максимальное по мощности подмножество $S \subseteq X$, покрываемое шаром, не пересекающимся с Y.

Все перечисленные задачи моделируют актуальные прикладные проблемы анализа данных [1, 3, 4, 6–10, 12]. Отметим важную общую черту этих задач. Каждая из них обладает одним из свойств «локальности», сформулированных ниже.

Свойство L1. *Для каждого входа задачи существуют точка $x^* \in \mathbb{R}^d$ и целое число $k \in [1, n]$ такие, что всякое k-элементное множество векторов из X, ближайших к x^*, является оптимальным решением задачи.*

Свойство L2. *Для каждого входа задачи существуют точка $y^* \in \mathbb{R}^d$ и целое число $k \in [1, n]$ такие, что всякое k-элементное множество векторов из X, имеющих максимальные скалярные произведения с y^*, является оптимальным решением задачи.*

В дальнейшем точки x^* и y^*, о которых идёт речь в этих свойствах, будем называть соответственно *оптимальным центром* и *оптимальным направлением* рассматриваемой задачи. Легко показать, что задачи 1 и 5 обладают свойством L1: оптимальным центром в задаче 1 является среднее арифметическое векторов, входящих в оптимальное подмножество, а в задаче 5 — центр шара, покрывающего оптимальное подмножество и не пересекающегося с Y (см. разд. 2). Задачи 2–4 обладают свойством L2: оптимальным направлением в этих задачах является сумма векторов оптимального подмножества (см. разд. 3).

Известные результаты. В общем случае задачи 1–4 NP-трудны в сильном смысле [1,3,7,8]. Задача 5 NP-трудна, даже если множество Y состоит из одного элемента [15].

В случае, когда размерность пространства фиксирована, задачи 1–4 полиномиально разрешимы. Так, оптимальное решение задачи 1 может быть найдено за время $O(dn^{d+1})$ [11] с использованием диаграмм Вороного k-го порядка. Оптимальное решение задач 2–4 можно найти за время $O(n^{2d+2})$ [4, 5][1)] с использованием идей из [2]. В [12] утверждается (без доказательства), что задача 5 разрешима за время $n^{O(d)}$, где $n = |X \cup Y|$.

Результаты статьи. Предложен простой универсальный алгоритм решения задач, обладающих одним из свойств L1 либо L2. Алгоритм основан на стандартной идее перебора ячеек диаграмм Вороного, подобной той, что используется в [11, 16]. Обоснование алгоритма происходит в два этапа: вначале устанавливается, что он находит оптимальное решение задач со свойством L1 (теорема 1), после чего доказывается, что свойство L2 является частным случаем свойства L1 (лемма 3). В случае задач 1–5 трудоёмкость предложенного алгоритма оценивается величиной $O(dn^{d+1})$, причём решения задач 1–3 определяются за это время сразу для всех k. Отметим, что тем самым существенно улучшено время решения задач 2–4 и обоснована полиномиальная разрешимость при фиксированном d задачи 5.

1. Диаграммы Вороного k-го порядка

Приведём основные определения, а также некоторые геометрические и алгоритмические факты, связанные с диаграммами Вороного высших порядков [13, 14, 16].

[1)] При аккуратном вычислении трудоёмкости формула получается именно такая, а не $O(d^2n^{2d})$, как указано в [4, 5].

Ячейкой Вороного для произвольного подмножества $S \subseteq X$ называется множество

$$V(S,X) = \{x \in \mathbb{R}^d \mid \|x-y\| < \|x-z\| \text{ для всех } y \in S,\ z \in X \setminus S\},$$

состоящее из точек пространства, для которых точки из множества S ближе, чем точки из $X \setminus S$. Очевидно, что $V(S,X)$ является пересечением полупространств вида $\{x \in \mathbb{R}^d \mid \|x-y\| < \|x-z\|\}$, где $y \in S,\ z \in X \setminus S$. Каждое из этих полупространств лежит по одну сторону от гиперплоскости $\|x-y\| = \|x-z\|$, называемой *бисектором* соответствующих точек y и z. Отсюда следует, что ячейка $V(S,X)$ представляет собой (возможно, пустой либо неограниченный) выпуклый многогранник в пространстве $\mathbb{R}^d$.

Очевидно, что замыканием непустой ячейки Вороного $V(S,X)$ является множество

$$\overline{V(S,X)} = \{x \in \mathbb{R}^d \mid \|x-y\| \leqslant \|x-z\| \text{ для всех } y \in S,\ z \in X \setminus S\}. \quad (1)$$

Непосредственным следствием (1) является

Факт 1. *Если $S \subseteq X$, $|S| = k$ и $x \in \overline{V(S,X)}$, то множество S состоит из k векторов множества X, ближайших к x.*

Диагаммой Вороного k-го порядка для множества X называется множество $V_k(X)$ всех непустых ячеек Вороного вида $V(S,X)$, где $S \subseteq X$ и $|S| = k$. Легко заметить, что ячейки диаграммы $V_k(X)$ являются непересекающимися множествами и что их замыкания покрывают всё пространство $\mathbb{R}^d$.

Для произвольной ячейки Вороного $C \in V_k(X)$ легко восстановить k-элементное множество $S = S(C)$ такое, что $C = V(S,X)$. Очевидно, что множество $S(C)$ состоит из k векторов из X, ближайших к точкам ячейки C. Из определения ячейки Вороного следует, что такое множество единственно и одинаково для всех точек из C.

Факт 2 [14]. *Общее количество ячеек всех диаграмм Вороного порядка от* 1 *до n равно $O(n^{d+1})$.*

В [14] предложен алгоритм построения диаграмм Вороного всех порядков, который в дальнейшем будем называть *алгоритмом* $\mathcal{V}$. Выходом алгоритма является структура данных, описывающая ячейки всех диаграмм Вороного $V_k(X)$, $k = 1,\dots,n$. В частности, для каждой ячейки $C \in V_k(X)$, $k = 2,\dots,n$, указаны ячейка $C' \in V_{k-1}(X)$ и точка $s(C) \in S(C) \setminus S(C')$, для которых $S(C) = S(C') \cup \{s(C)\}$. Для ячеек диаграммы $V_1(X)$ указана точка $s(C)$ такая, что $S(C) = \{s(C)\}$.

Факт 3 [14]. *Трудоёмкость алгоритма $\mathcal{V}$ равна $O(dn^{d+1})$.*

2. Решение задач, обладающих свойством L1

Покажем, как, используя данные факты, можно решать задачи, обладающие свойством L1. Рассмотрим следующий «универсальный» алгоритм решения этих задач.

Алгоритм $\mathcal{A}$

Шаг 1. Строим диаграммы Вороного $V_k(X)$ для всех $k = 1, \ldots, n$ с помощью алгоритма $\mathcal{V}$.

Шаг 2. Для каждой ячейки $C \in V_1(X)$ определяем множество $S(C) = \{s(C)\}$; для каждого $k = 2, \ldots, n$ и каждой ячейки $C \in V_k(X)$ определяем множество $S(C)$ путём добавления к множеству $S(C')$ точки $s(C)$. Проверяем множество $S(C)$ на допустимость и вычисляем $f(S(C))$.

Шаг 3. Среди всех допустимых множеств, построенных на предыдущем шаге, выбираем наилучшее с точки зрения целевой функции.

Обозначим через $T = T(X, f, \Gamma)$ максимальное время проверки допустимости и вычисления значения целевой функции для произвольного подмножества $S \subseteq X$.

Теорема 1. *Если задача Vector Subset обладает свойством* L1, *то её оптимальное решение может быть найдено с помощью алгоритма* $\mathcal{A}$ *за время* $O(n^{d+1} \max\{d, T\})$.

Доказательство. Согласно свойству L1 существуют точка $x^* \in \mathbb{R}^d$ и число $k \in [1, n]$ такие, что всякое k-элементное множество векторов из X, ближайших к x^*, является оптимальным решением задачи. Поскольку замыкания ячеек диаграммы Вороного $V_k(X)$ покрывают всё пространство $\mathbb{R}^d$, существует ячейка $C \in V_k(X)$ такая, что $x^* \in \overline{C}$. Согласно факту 1 множество $S(C)$ составляют k векторов, ближайших к x^*. Отсюда по выбору x^* и k получаем, что $S(C)$ — оптимальное решение задачи. При этом множество $S(C)$ будет рассмотрено на шаге 2 и учтено на шаге 3, поскольку всякое оптимальное множество допустимо. Следовательно, значение целевой функции на подмножестве, найденном алгоритмом $\mathcal{A}$, будет оптимальным.

Остаётся оценить трудоёмкость алгоритма. Построение диаграмм Вороного всех порядков занимает время $O(dn^{d+1})$ согласно факту 3. Количество ячеек Вороного ограничено величиной $O(n^{d+1})$ согласно факту 2. Определение допустимости множества $S(C)$ и вычисление значения целевой функции $f(S(C))$ на шаге 2 можно выполнить за время $O(T)$. Таким образом, общая трудоёмкость алгоритма равна $O(n^{d+1} \max\{d, T\})$. Теорема 1 доказана.

Замечание 1. Если предикат допустимости Γ определяет мощность искомого подмножества, т. е. $\Gamma(S) \Leftrightarrow |S| = k$, то алгоритм $\mathcal{A}$ за указанное в теореме 1 время позволяет найти оптимальные решения сразу для всех k от 1 до n, поскольку на шаге 2 рассматриваются ячейки диаграмм Вороного всех порядков и соответственно все k-элементные подмножества вида $S(C)$, $C \in V_k(X)$, для каждого k.

Задача 1. Примером задачи Vector Subset, обладающей свойством L1, является задача 1: легко показать, что оптимальным центром в ней является среднее арифметическое векторов оптимального подмножества. Действительно, пусть $x^* = c(S^*)$, где S^* — оптимальное решение этой задачи, и $k = |S^*|$. Тогда для произвольного множества S имеем

$$f_1(S) = \sum_{x \in S} \|x - c(S)\|^2 \leqslant \sum_{x \in S} \|x - x^*\|^2,$$

поскольку минимум суммы квадратов расстояний до векторов из S достигается именно в точке $c(S)$ (см., например, [10]). Тем самым если множество S состоит из k векторов, ближайших к x^*, то

$$f_1(S) \leqslant \sum_{x \in S^*} \|x - x^*\|^2 = f_1(S^*),$$

а значит, S — оптимальное решение.

Покажем, что вычисление целевой функции задачи 1 на шаге 2 может быть выполнено за время $O(d)$. Пусть $S = S(C)$, $S' = S(C')$ и $s = s(C)$. Тогда $s \in S \setminus S'$ и $S = S' \cup \{s\}$, а значит,

$$f_1(S) = f_1(S, c(S)) = f_1(S', c(S)) + \|s - c(S)\|^2,$$

где $f_1(S, y) = \sum\limits_{x \in S} \|x - y\|^2$. Отсюда

$$f_1(S) = f_1(S') + |S'\,|\; \|c(S) - c(S')\|^2 + \|s - c(S)\|^2$$

согласно равенству $f_1(S, y) = f_1(S, c(S)) + |S\,|\; \|y - c(S)\|^2$ [10]. При этом вектор $c(S)$ можно определить равенством $c(S) = c(S')(1 - 1/|S|) + s/|S|$. Следовательно, если запоминать вектор $c(S(C))$ и значение $f_1(S(C))$ для всех ячеек диаграммы $V_k(X)$, то вычисление соответствующего вектора и значения f для ячеек диаграммы $V_{k+1}(X)$ займёт время $O(d)$.

Проверка допустимости подмножества $S(C)$ на шаге 2 в случае задачи 1 тривиальна, поскольку мощность $S(C)$ равна текущему значению индекса k, для которого выполняется данный шаг. Таким образом, с учётом замечания 1 получаем следующее утверждение.

Теорема 2. *Оптимальные решения задачи* 1 *для всех* k *могут быть найдены за время* $O(dn^{d+1})$.

Отметим, что процесс вычисления целевой функции задачи 1 в алгоритме из [11] организован иначе. Данный алгоритм рассматривает только ячейки диаграммы $V_k(X)$, где k — заданная мощность искомого подмножества. При этом предполагается, что структура данных, описывающая $V_k(X)$, обладает следующим свойством: ячейки диаграммы упорядочены в ней так, что каждая следующая ячейка граничит с предыдущей. В этом случае при переходе от ячейки C_1 к ячейке C_2 множество $S(C_2)$ отличается от множества $S(C_1)$ ровно на один элемент, за счёт чего время вычисления функции $f_1(S)$ может быть также сведено к величине $O(d)$. К сожалению, возможность указанного упорядочения ячеек Вороного требует обоснования, которое отсутствует в [11].

Задача 5. Другим примером задачи, обладающей свойством L1, является задача 5, рассматриваемая как задача Vector Subset на множестве $X \cup Y$. Действительно, пусть S^* — оптимальное решение этой задачи, $k = |S^*|$ и B^* — шар, покрывающий S^* и не пересекающийся с Y. Обозначим через x^* и r^* центр и радиус этого шара. Тогда если множество S состоит из k ближайших к x^* точек множества $X \cup Y$, то S также покрывается шаром B^*, а значит, S — допустимое решение задачи 5, причём из максимальности k следует, что это решение оптимально.

Отсюда согласно теореме 1 задача 5 разрешима алгоритмом $\mathcal{A}$. Чтобы уточнить время его работы, определим время вычисления предиката допустимости и целевой функции на шаге 2 этого алгоритма. Заметим, что произвольное подмножество вида $S(C)$, где $C \in V_k(X \cup Y)$, допустимо тогда и только тогда, когда $S(C) \cap Y = \varnothing$. Это следует из того, что любая точка ячейки C является центром некоторого шара, покрывающего $S(C)$ и не пересекающегося с другими элементами $X \cup Y$. Таким образом, проверка допустимости множества $S(C)$ может быть выполнена за время $O(1)$ с использованием формулы $\Gamma(S(C)) = \Gamma(S(C')) \wedge (s(C) \notin Y)$. Столько же занимает и вычисление целевой функции на множестве $S(C)$, поскольку мощность этого множества равна текущему значению индекса k, для которого выполняется шаг 2. Отсюда и из теоремы 1 получается

Теорема 3. *Оптимальное решение задачи* 5 *может быть найдено за время* $O(dn^{d+1})$, *где* $n = |X \cup Y|$.

Замечание 2. В качестве центра шара, покрывающего оптимальное подмножество X и не пересекающегося с Y, можно взять любую точку, принадлежащую соответствующей ячейке Вороного. Такую точ-

ку для каждой ячейки содержит структура данных, возвращаемая алгоритмом $\mathcal{V}$ при построении диаграмм Вороного [14]. Таким образом, за указанное в теореме 3 время можно найти не только оптимальное решение задачи 5, но и сам шар, о котором идёт речь в этой задаче.

Замечание 3. Легко показать, что аналогичный результат справедлив и для более общей задачи, в которой для заданного k требуется найти шар, покрывающий максимальное количество точек из X и содержащий не более k точек из Y. То же самое относится и к двойственной задаче — о шаре, покрывающем не менее k точек из X и содержащем минимальное количество точек из Y.

3. Решение задач, обладающих свойством L2

Теорема 1 обосновывает применение алгоритма 1 для задач, обладающих свойством L1. Покажем, что данный алгоритм находит оптимальное решение задач со свойством L2. Для этого достаточно доказать, что свойство L2 является частным случаем свойства L1.

Обозначим через $\langle \cdot\, , \cdot \rangle$ скалярное произведение в пространстве $\mathbb{R}^d$.

Лемма 1. *Пусть $\langle x, v\rangle > \langle y, v\rangle$ для некоторых векторов $x, y, v \in \mathbb{R}^d$. Тогда существует число $\alpha(x, y, v)$ такое, что $\|x-\alpha v\| < \|y-\alpha v\|$ для всех $\alpha > \alpha(x, y, v)$.*

Доказательство. Очевидно, что вектор v ненулевой и x не совпадает с y. Пусть h — бисектор точек x и y, h^+ — полупространство, состоящее из точек, более близких к x, чем к y, $\mathbf{0}$ — нулевой вектор и $(\mathbf{0}, v)$ — прямая, проходящая через $\mathbf{0}$ и v (рис. 1).

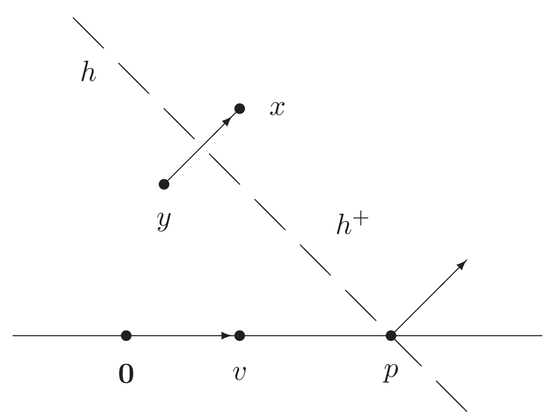

Рис. 1

Заметим, что вектор $x - y$ перпендикулярен гиперплоскости h, но не перпендикулярен прямой $(\mathbf{0}, v)$, поскольку $\langle x, v\rangle \neq \langle y, v\rangle$. Следовательно, прямая $(\mathbf{0}, v)$ пересекает гиперплоскость h. Обозначим точку их пересечения через p. Тогда полупространство h^+ может быть задано соотношением $h^+ = \{z \in \mathbb{R}^d \mid \langle z - p, x - y\rangle > 0\}$. С другой стороны, точка p лежит на прямой $(\mathbf{0}, v)$, следовательно, $p = tv$ для некоторого t. Положим $\alpha(x, y, v) = t$. Тогда если $\alpha > t$, то $\langle \alpha v - p, x - y\rangle = (\alpha - t)\langle v, x - y\rangle > 0$, поскольку $\langle x, v\rangle > \langle y, v\rangle$. Следовательно, точка αv лежит в полупространстве h^+. Лемма 1 доказана.

Лемма 2. *Пусть $v \in \mathbb{R}^d$ и X — конечное множество векторов из $\mathbb{R}^d$. Тогда существует вектор $u \in \mathbb{R}^d$ такой, что $\|x - u\| < \|y - u\|$ для всех пар $x, y \in X$ со свойством $\langle x, v\rangle > \langle y, v\rangle$.*

Доказательство. Положим

$$\alpha = 1 + \max\{\alpha(x, y, v) \mid x, y \in X,\ \langle x, v\rangle > \langle y, v\rangle\},$$

где числа $\alpha(x, y, v)$ определены в лемме 1. Тогда вектор $u = \alpha v$, очевидно, удовлетворяет утверждению леммы. Лемма 2 доказана.

Лемма 3. *Пусть $\mathcal{P}$ — произвольная задача Vector Subset, удовлетворяющая свойству* L2. *Тогда $\mathcal{P}$ удовлетворяет и свойству* L1.

Доказательство. Пусть X — множество, заданное на входе задачи $\mathcal{P}$. Согласно свойству L2 существуют точка $v \in \mathbb{R}^d$ и число $k \in [1, n]$ такие, что всякое k-элементное множество векторов из X, имеющих максимальные скалярные произведения с v, является оптимальным решением задачи $\mathcal{P}$. В силу леммы 2 существует вектор u такой, что $\|x - u\| < \|y - u\|$ для всех пар $x, y \in X$ со свойством $\langle x, v\rangle > \langle y, v\rangle$. Покажем, что вектор u — оптимальный центр задачи $\mathcal{P}$.

Пусть S — произвольное k-элементное множество векторов из X, ближайших к u. Если для некоторых векторов $x \in S$ и $y \in X \setminus S$ выполнено неравенство $\langle x, v\rangle < \langle y, v\rangle$, то согласно выбору вектора u имеем $\|x - u\| > \|y - u\|$, что противоречит выбору множества S. Следовательно, $\langle x, v\rangle \geqslant \langle y, v\rangle$ для любых $x \in S$ и $y \in X \setminus S$. Отсюда получаем, что множество S состоит из k векторов, имеющих максимальное скалярное произведение с v, а значит, образует оптимальное решение задачи $\mathcal{P}$ в силу выбора v и k. Таким образом, u и k — вектор и число, существование которых требуется в свойстве L1. Лемма 3 доказана.

Из леммы 3 и теоремы 1 вытекает

Теорема 4. *Если задача Vector Subset обладает свойством* L2, *то её оптимальное решение может быть найдено с помощью алгоритма* $\mathcal{A}$ *за время* $O(n^{d+1}\max\{d,T\})$.

Отметим, что в случае задач, обладающих свойством L2, алгоритм $\mathcal{A}$ на шаге 2 может рассматривать только неограниченные ячейки диаграмм Вороного, поскольку в качестве оптимального центра, заменяющего оптимальное направление (см. доказательство леммы 2), может быть выбрана сколь угодно далёкая точка.

Задачи 2–4. Примерами задач, обладающих свойством L2, являются задачи 2–4: оптимальным направлением в этих задачах является сумма векторов оптимального подмножества. Действительно, пусть $y^* = \sum\limits_{x\in S^*} x$, где S^* — оптимальное решение задачи 3, и $k = |S^*|$. Тогда

$$f_3(S) = \|\sum_{x\in S} x\| \geqslant \Big\langle \sum_{x\in S} x, y^*/\|y^*\|\Big\rangle,$$

поскольку $\|a\|\|b\| \geqslant \langle a,b\rangle$. Тем самым если множество S состоит из k векторов, имеющих максимальные скалярные произведения с y^*, то

$$f_3(S) \geqslant \Big\langle \sum_{x\in S^*} x, y^*/\|y^*\|\Big\rangle = f_3(S^*),$$

а значит, S — оптимальное решение. Таким образом, задача 3 обладает свойством L2. Аналогичное утверждение верно и для задач 2 и 4, поскольку

$$f_4(S) = f_3^2(S)/|S|, \quad f_2(S) = F - f_4(S), \tag{2}$$

где $F = \sum\limits_{x\in X} \|x\|^2$ [5].

Покажем, что вычисление целевых функций задач 2–4 на шаге 2 может быть выполнено за время $O(d)$. Пусть $S = S(C)$, $S' = S(C')$ и $s = s(C)$. Тогда $s \in S \setminus S'$ и $S = S' \cup \{s\}$, а значит, $\sum\limits_{x\in S} x = \sum\limits_{x\in S'} x + s$. Следовательно, если запоминать вектор $\sum\limits_{x\in S(C)} x$ для всех ячеек диаграммы $V_k(X)$, то вычисление такого же вектора для ячеек диаграммы $V_{k+1}(X)$ займёт время $O(d)$. Вычисление нормы вектора также выполнимо за время $O(d)$. Таким образом, вычисление целевой функции задачи 3 на шаге 2 может быть выполнено за время $O(d)$. Аналогичное утверждение верно и для задач 2 и 4 в силу равенств (2).

Проверка допустимости подмножества $S(C)$ на шаге 2 в случае задач 2 и 3 тривиальна, поскольку мощность $S(C)$ равна текущему значению индекса k, для которого выполняется данный шаг, а для задачи 4 любое рассматриваемое подмножество допустимо. Таким образом, с учётом замечания 1 получаем следующее утверждение.

Теорема 5. *Оптимальные решения задач* 2 *и* 3 *для всех* k *могут быть найдены за время* $O(dn^{d+1})$. *Такое же время требуется и для решения задачи* 4.

Заключение

В статье доказана возможность применения известной техники, связанной с диаграммами Вороного высших порядков, для решения задач 2–5. Предложенный метод существенно уменьшает время решения задач 2–4 по сравнению с известными алгоритмами, а также доказывает полиномиальную разрешимость задачи 5 в случае фиксированной размерности пространства.

ЛИТЕРАТУРА

1. **Бабурин А. Е., Гимади Э. Х., Глебов Н. И., Пяткин А. В.** Задача отыскания подмножества векторов с максимальным суммарным весом // Дискрет. анализ и исслед. операций. Сер. 2. 2007. Т. 14, № 1. С. 32–42.
2. **Бабурин А. Е., Пяткин А. В.** О полиномиальных алгоритмах решения одной задачи суммирования векторов // Дискрет. анализ и исслед. операций. Сер. 1. 2006. Т. 13, № 2. С. 3–10.
3. **Гимади Э. Х., Кельманов А. В., Кельманова М. А., Хамидуллин С. А.** Апостериорное обнаружение в числовой последовательности квазипериодического фрагмента при заданном числе повторов // Сиб. журн. индустр. математики. 2006. Т. 9, № 1. С. 55–74.
4. **Гимади Э. Х., Пяткин А. В., Рыков И. А.** О полиномиальной разрешимости некоторых задач выбора подмножества векторов в евклидовом пространстве фиксированной размерности // Дискрет. анализ и исслед. операций. 2008. Т. 15, № 6. С. 11–19.
5. **Долгушев А. В., Кельманов А. В.** Приближённый алгоритм решения одной задачи кластерного анализа // Дискрет. анализ и исслед. операций. 2011. Т. 18, № 2. С. 29–40.
6. **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.
7. **Кельманов А. В., Пяткин А. В.** Об одном варианте задачи выбора подмножества векторов // Дискрет. анализ и исслед. операций. 2008. Т. 15, № 5. С. 20–34.

8. **Кельманов А. В., Пяткин А. В.** NP-полнота некоторых задач выбора подмножества векторов // Дискрет. анализ и исслед. операций. 2010. Т. 17, № 5. С. 37–45.
9. **Кельманов А. В., Романченко С. М.** FPTAS для одной задачи поиска подмножества векторов // Дискрет. анализ и исслед. операций. 2014. Т. 21, № 3. С. 41–52.
10. **Шенмайер В. В.** Аппроксимационная схема для одной задачи поиска подмножества векторов // Дискрет. анализ и исслед. операций. 2012. Т. 19, № 2. С. 92–100.
11. **Aggarwal A., Imai H., Katoh N., Suri S.** Finding k points with minimum diameter and related problems // J. Algorithms. 1991. Vol. 12, No. 1. P. 38–56.
12. **Aronov B., Har-Peled S.** On approximating the depth and related problems // SIAM J. Computing. 2008. Vol. 38, No. 3. P. 899–921.
13. **Aurenhammer F., Klein R.** Voronoi diagrams // Handbook of computational geometry (J.-R. Sack, J. Urrutia, ed.). Amsterdam: Elsevier, 2000. P. 201–290.
14. **Edelsbrunner H., O'Rourke J., Seidel R.** Constructing arrangements of lines and hyperplanes with applications // SIAM J. Computing. 1986. Vol. 15, No. 2. P. 341–363.
15. **Johnson D. S., Preparata F. P.** The densest hemisphere problem // Theor. Comput. Sci. 1978. Vol. 6, No. 1. P. 93–107.
16. **Shamos M. I., Hoey D.** Closest-point problems // Proc. 16th IEEE Annu. Symp. Found. Comput. Sci. (Berkeley, Oct. 13–15, 1975). Piscaway: IEEE, 1975. P. 151–162.

Шенмайер Владимир Владимирович

Статья поступила
20 мая 2016 г.

Исправленный вариант —
15 июня 2016 г.

DISKRETNYI ANALIZ I ISSLEDOVANIE OPERATSII
October–December 2016. Volume 23, No. 4. P. 102–115

UDC 519.176 DOI: 10.17377/daio.2016.23.526

SOLVING SOME VECTOR SUBSET PROBLEMS BY VORONOI DIAGRAMS

Vladimir V. Shenmaier[1]
[1]Sobolev Institute of Mathematics,
4 Acad. Koptyug Ave., 630090 Novosibirsk, Russia
e-mail: shenmaier@mail.ru

Abstract. We propose a general approach to solving some vector subset problems in a Euclidean space that is based on higher-order Voronoi diagrams. In the case of a fixed space dimension, this approach allows us to find optimal solutions to these problems in polynomial time which is better than the runtime of available algorithms. Ill. 1, bibliogr. 16.

Keywords: computational geometry, vector subset problem, Euclidean space, Voronoi diagram, polynomial-time algorithm.

REFERENCES

1. **A. E. Baburin, E. Kh. Gimadi, N. I. Glebov,** and **A. V. Pyatkin,** The problem of finding a subset of vectors with the maximum total weight, *Diskretn. Anal. Issled. Oper., Ser. 2,* **14**, No. 1, 32–42, 2007. Translated in *J. Appl. Ind. Math.,* **2**, No. 1, 32–38, 2008.
2. **A. E. Baburin** and **A. V. Pyatkin**, Polynomial algorithms for solving the vector sum problem, *Diskretn. Anal. Issled. Oper., Ser. 1,* **13**, No. 2, 3–10, 2006. Translated in *J. Appl. Ind. Math.,* **1**, No. 3, 268–272, 2007.
3. **E. Kh. Gimadi, A. V. Kel'manov, M. A. Kel'manova,** and **S. A. Khamidullin,** A posteriori detection of a quasiperiodic fragment with a given number of repetitions in a numerical sequence, *Sib. Zh. Ind. Mat.,* **9**, No. 1, 55–74, 2006. Translated in *Pattern Recognit. Image Anal.,* **18**, No. 1, 30–42, 2008.
4. **E. Kh. Gimadi, A. V. Pyatkin,** and **I. A. Rykov,** On polynomial solvability of some problems of a vector subset choice in a Euclidean space of fixed dimension, *Diskretn. Anal. Issled. Oper.,* **15**, No. 6, 11–19, 2008. Translated in *J. Appl. Ind. Math.,* **4**, No. 1, 48–53, 2010.
5. **A. V. Dolgushev** and **A. V. Kel'manov,** An approximation algorithm for solving a problem of cluster analysis, *Diskretn. Anal. Issled. Oper.,* **18**, No. 2, 29–40, 2011. Translated in *J. Appl. Ind. Math.,* **5**, No. 4, 551–558, 2011.

6. **A. V. Dolgushev, A. V. Kel'manov,** and **V. V. Shenmaier,** Polynomial-time approximation scheme for a problem of partitioning a finite set into two clusters, *Tr. Inst. Mat. Mekh.,* **21**, No. 3, 100–109, 2015.
7. **A. V. Kel'manov** and **A. V. Pyatkin,** On a version of the problem of choosing a vector subset, *Diskretn. Anal. Issled. Oper.,* **15**, No. 5, 20–34, 2008. Translated in *J. Appl. Ind. Math.,* **3**, No. 4, 447–455, 2009.
8. **A. V. Kel'manov** and **A. V. Pyatkin,** NP-completeness of some problems of choosing a vector subset, *Diskretn. Anal. Issled. Oper.,* **17**, No. 5, 37–45, 2010. Translated in *J. Appl. Ind. Math.,* **5**, No. 3, 352–357, 2011.
9. **A. V. Kel'manov** and **S. M. Romanchenko,** An FPTAS for a vector subset search problem, *Diskretn. Anal. Issled. Oper.,* **21**, No. 3, 41–52, 2014. Translated in *J. Appl. Ind. Math.,* **8**, No. 3, 329–336, 2014.
10. **V. V. Shenmaier**, An approximation scheme for a problem of search for a vector subset, *Diskretn. Anal. Issled. Oper.,* **19**, No. 2, 92–100, 2012. Translated in *J. Appl. Ind. Math.,* **6**, No. 3, 381–386, 2012.
11. **A. Aggarwal, H. Imai, N. Katoh,** and **S. Suri**, Finding k points with minimum diameter and related problems, *J. Algorithms,* **12**, No. 1, 38–56, 1991.
12. **B. Aronov** and **S. Har-Peled**, On approximating the depth and related problems, *SIAM J. Comput.,* **38**, No. 3, 899–921, 2008.
13. **F. Aurenhammer** and **R. Klein**, Voronoi diagrams, in J.-R. Sack and J. Urrutia, eds., *Handbook of Computational Geometry,* pp. 201–290, Elsevier, Amsterdam, 2000.
14. **H. Edelsbrunner, J. O'Rourke,** and **R. Seidel**, Constructing arrangements of lines and hyperplanes with applications, *SIAM J. Comput.,* **15**, No. 2, 341–363, 1986.
15. **D. S. Johnson** and **F. P. Preparata**, The densest hemisphere problem, *Theor. Comput. Sci.,* **6**, No. 1, 93–107, 1978.
16. **M. I. Shamos** and **D. Hoey**, Closest-point problems, in *Proc. 16th IEEE Annu. Symp. Found. Comput. Sci., Berkeley, USA, Oct. 13–15, 1975,* pp. 151–162, IEEE, Piscataway, 1975.

Vladimir V. Shenmaier

Received
20 May 2016

Revised
15 June 2016