ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Ноябрь—декабрь 2008. Том 15, № 6. С. 11–19

УДК 519.8

О ПОЛИНОМИАЛЬНОЙ РАЗРЕШИМОСТИ НЕКОТОРЫХ ЗАДАЧ ВЫБОРА ПОДМНОЖЕСТВА ВЕКТОРОВ В ЕВКЛИДОВОМ ПРОСТРАНСТВЕ ФИКСИРОВАННОЙ РАЗМЕРНОСТИ[*)]

Э. Х. Гимади, А. В. Пяткин, И. А. Рыков

Аннотация. Рассмотрены задачи, связанные с выбором из конечного семейства векторов в евклидовом пространстве $\mathbb{R}^k$ подмножества векторов. В качестве (максимизируемых) целевых функций задач выступают норма суммы и усреднённый квадрат нормы суммы. Для решения этих задач разработаны точные комбинаторные алгоритмы с временно́й сложностью $O(k^2 n^{2k})$. Тем самым доказана полиномиальная разрешимость данных задач при фиксированном k.

Ключевые слова: подмножество векторов, евклидово пространство, полиномиальная разрешимость.

Введение

В настоящей статье исследуются дискретные оптимизационные задачи «ПОДМНОЖЕСТВО ВЕКТОРОВ с максимальной нормой суммы» (далее m-ПВ и ПВ в зависимости от того, задано или нет число выбираемых векторов) и «ПОДМНОЖЕСТВО ВЕКТОРОВ с максимумом УСРЕДНЁННОГО квадратата нормы суммы» (далее ПВУ).

Под нормой в работе понимается евклидова норма в k-мерном пространстве $\mathbb{R}^k$, т. е. $\|x\| = \sqrt{x_1^2 + \ldots + x_k^2}$.

Задача ПВ рассматривалась в [1], задача m-ПВ — в [2–4, 6], а задача ПВУ — в [4].

Задача m-ПВ. *В евклидовом пространстве $\mathbb{R}^k$ заданы конечное семейство векторов $V = \{v_1, v_2, \ldots, v_n\}$ и число $m < n$. Требуется найти подсемейство векторов $U \subseteq V$ мощности m, обладающее максимальной нормой суммы $\|\sum\limits_{v\in U} v\|$.*

[*)]Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 08–01–00516 и 07–07–00222).

Задача ПВУ. *В евклидовом пространстве* $\mathbb{R}^k$ *задано конечное семейство векторов* $V = \{v_1, v_2, \ldots, v_n\}$. *Требуется найти непустое подсемейство векторов* $U \subseteq V$, *максимизирующее функцию*

$$F(U) = \Big\| \sum_{v \in U} v \Big\|^2 \Big/ |U|. \tag{1}$$

Заметим, что в целевой функции задачи ПВУ происходит в некотором роде усреднение квадрата нормы суммы векторов из выбраного подмножества.

Эти задачи возникают при решении прикладной задачи обнаружения квазипериодически повторяющегося вектор-фрагмента в зашумлённой числовой последовательности. Подобная ситуация типична для таких приложений, как электронная разведка, радиолокация, телекоммуникация, геофизика, обработка речевых сигналов, медицинская и техническая диагностика и др. [2, 4].

В общем случае задачи m-ПВ и ПВУ NP-трудны [4, 6]. Однако вопрос об их сложностном статусе при фиксированной размерности пространства оставался открытым.

В [6] для задачи m-ПВ разработан алгоритм, решающий задачу приближённо с произвольной наперёд заданной гарантированной относительной погрешностью $\varepsilon > 0$ за время $O(nk^{(k+3)/2}(1/\varepsilon)^{(k-1)/2})$. При фиксированной размерности k пространства $\mathbb{R}^k$ это позволило построить путём подходящего выбора ε вполне полиномиальную аппроксимационную схему (FPTAS) и получить условия асимптотической точности алгоритмов с полиномиальной временно́й сложностью.

В случае векторов с целочисленными координатами с использованием комбинаторной техники и техники динамического программирования построены точные псевдополиномиальные алгоритмы временно́й сложности $O(nk^{k+1}(mb)^{k-1})$ [6] и $O(nkm(2mb)^{k-1})$ [2] соответственно. Через b обозначена максимальная абсолютная величина координат векторов $\{v_1, v_2, \ldots, v_n\}$.

Для задачи ПВУ в [4] был предложен приближённый алгоритм, находящий решение задачи с погрешностью не более ε за время

$$O\left(nk(k + \log n)\left(\frac{k-1}{2\varepsilon} + 1\right)^{k-1}\right).$$

В [1] для задачи ПВ (без ограничения на мощность выбираемого подмножества векторов) был получен точный алгоритм временно́й слож-

ности $O(k^2n^k)$, полиномиальный в случае фиксированного k. На доказанной в этой работе лемме о числе областей принадлежности решения (ОПР) базируются предложенные в настоящей статье точные алгоритмы решения задач m-ПВ и ПВУ.

В разд. 1 приведены известные факты о сложности и алгоритмах решения рассматриваемых задач.

Разд. 2 посвящён построению точных алгоритмов для решения этих задач, полиномиальных при фиксированной размерности евклидова пространства $\mathbb{R}^k$. Алгоритмы для решения каждой из рассматриваемых задач — m-ПВ и ПВУ — имеют одинаковую временну́ю сложность $O(k^2n^{2k})$. Тем самым закрыт вопрос о полиномиальной разрешимости этих задач при фиксированном k.

1. Известные факты о задачах выбора подмножества векторов

Напомним некоторые результаты, полученные для рассматриваемых задач в [4, 6].

Теорема 1 [6]. *Задача* m-ПВ NP-*трудна.*

Теорема 1′ [4]. *Задача* ПВУ NP-*трудна.*

Заметим, что утверждения теорем 1 и 1′ верны при условии, что размерность пространства k является входным параметром. Как будет показано далее, при фиксированном k эти задачи полиномиально разрешимы.

Теорема 2 [6]. *Задача* m-ПВ *решается с гарантированной относительной погрешностью, не превышающей величины* $(k-1)/(8L^2)$, *за время* $O(nk^2(2L+1)^{k-1})$, *где* L *— параметр алгоритма.*

Теорема 2′ [4]. *Задача* ПВУ *решается с гарантированной относительной погрешностью, не превышающей* $(k-1)/(4L^2)$, *за время*

$$O(nk(k+\log n)(2L+1)^{k-1}).$$

Из теорем 2 и 2′ вытекает существование для задач m-ПВ и ПВУ вполне полиномиальной аппроксимационной схемы (FPTAS) для случая фиксированной размерности k пространства $\mathbb{R}^k$. Действительно, положим, например, для задачи m-ПВ относительную погрешность равной $\varepsilon = (k-1)/(8L^2)$. Тогда $L = \sqrt{(k-1)/(8\varepsilon)}$ и для времени решения задачи m-ПВ получим оценку $O(nk^2((k-1)/\varepsilon)^{(k-1)/2})$, полиномиальную относительно n и $1/\varepsilon$.

Из теоремы 2 следует

Теорема 3 [6]. *При фиксированной размерности k пространства $\mathbb{R}^k$ задача m-ПВ решается асимптотически точно при выборе параметра $L = L(n)$, где $L(n)$ — произвольная неограниченно растущая функция от n.*

Теорема 4 [2]. *Если размерность k пространства $\mathbb{R}^k$ фиксирована, то задачи m-ПВ с целочисленными координатами векторов разрешимы за псевдополиномиальное время $O(nm^k b^{k-1})$, где b — максимальная по абсолютной величине координата векторов из семейства V.*

2. Алгоритмы решения задач выбора подмножеств векторов, полиномиальные при фиксированном k

Для построения точных полиномиальных алгоритмов решения задач m-ПВ и ПВУ при фиксированной размерности k пространства $\mathbb{R}^k$ нам понадобятся сведения об областях принадлежности решений (ОПР), введённых в [1].

Ортогональной гиперплоскостью для ненулевого вектора $v \in \mathbb{R}^k$ называется гиперплоскость, заданная уравнением $(v, x) = 0$. Она является $(k-1)$-мерным подпространством, состоящим из векторов, ортогональных вектору v.

Семейством ОПР для данных ненулевых векторов $u_1, u_2, \ldots, u_t$ назовём семейство максимальных по включению связных подмножеств пространства $\mathbb{R}^k$, не пересекающихся с ортогональными гиперплоскостями для векторов $u_1, u_2, \ldots, u_t$. Заметим, что векторы v и w лежат в одной области принадлежности решения тогда и только тогда, когда $(v, u_i) \cdot (w, u_i) > 0$ для любого $i = 1, 2, \ldots, t$.

Набор векторов пространства $\mathbb{R}^k$, содержащий ровно по одному вектору из каждой области семейства ОПР, назовём *семейством представителей областей принадлежности решения* для векторов $u_1, u_2, \ldots, u_t$.

Характеризацию введённого семейства представителей ОПР представляет следующая

Лемма 1 [1]. *Семейство представителей* ОПР *для ненулевых векторов $u_1, u_2, \ldots, u_t$ в пространстве $\mathbb{R}^k$ имеет мощность $O(kt^{k-1})$ и может быть найдено за время $O(k^2 t^k)$.*

Далее рассмотрим множество векторов $u_{ij} = v_i - v_j$, где $1 \leqslant i < j \leqslant n$. Построим для них семейство ОПР и дополним некоторые из его областей точками гиперплоскостей, добавляя точки гиперплоскости $(u_{ij}, x) = 0$ к полупространству $(u_{ij}, x) < 0$ для всех $1 \leqslant i < j \leqslant n$. Полученное семейство областей назовём *семейством расширенных* ОПР для этих векто-

ров. Очевидно, что любой вектор пространства $\mathbb{R}^k$ принадлежит одной из областей этого семейства.

Пусть $v \in \mathbb{R}^k$. Введём на множестве V следующее отношение: будем считать, что $v_i \leqslant_v v_j$, если либо $(v_i - v_j, v) < 0$, либо $(v_i - v_j, v) = 0$ и $i \leqslant j$.

Лемма 2. *Для любого $v \in R^k$ отношение $\leqslant_v$ задаёт на множестве V линейный порядок.*

Доказательство. Рефлексивность отношения $\leqslant_v$ очевидна. Если $v_i \leqslant_v v_j$ и $v_j \leqslant_v v_i$, то $(v_i - v_j, v) = 0$, $i \leqslant j$ и $j \leqslant i$, т. е. $i = j$. Следовательно, отношение антисимметрично. Предположим, что $v_i \leqslant_v v_j$ и $v_j \leqslant_v v_k$. Тогда $(v_i - v_j, v) \leqslant 0$ и $(v_j - v_k, v) \leqslant 0$. Если хотя бы одно из этих неравенств строгое, то $(v_i - v_k, v) < 0$, т. е. $v_i \leqslant_v v_k$. В противном случае имеем $(v_i - v_j, v) = (v_j - v_k, v) = 0$ и $i \leqslant j \leqslant k$. Отсюда $(v_i - v_k, v) = 0$ и $i \leqslant k$, т. е. $v_i \leqslant_v v_k$. Таким образом, $\leqslant_v$ является частичным порядком. Поскольку любые два вектора из V, очевидно, сравнимы, то $\leqslant_v$ — линейный порядок. Лемма 2 доказана.

Формально векторы v и w лежат в одной расширенной области принадлежности решения тогда и только тогда, когда для всех $1 \leqslant i < j \leqslant n$ выполняются либо неравенства $(v, u_{ij}) > 0$ и $(w, u_{ij}) > 0$, либо неравенства $(v, u_{ij}) \leqslant 0$ и $(w, u_{ij}) \leqslant 0$. Отсюда, в частности, следует, что если векторы v и w лежат в одной расширенной ОПР, то линейные порядки $\leqslant_v$ и $\leqslant_w$ совпадают. Кроме того, объединение всех расширенных ОПР даёт $\mathbb{R}^k$. Поскольку множество расширенных ОПР находится во взаимно однозначном соответствии с множеством ОПР, то для расширенных ОПР выполняется лемма 1.

Следующая лемма служит основой для построения полиномиального алгоритма решения задачи m-ПВ.

Лемма 3. *Пусть U^* — оптимальное решение задачи m-ПВ и $v^* = \sum_{v \in U^*} v$. Тогда если $v_i \neq v_j$ и $v_i \leqslant_{v^*} v_j$, то либо $v_i \notin U^*$, либо $v_j \in U^*$.*

Доказательство. Предположим, что $v_i \in U^*$ и $v_j \notin U^*$. Положим $U = (U^* \setminus \{v_i\}) \cup \{v_j\}$. Ясно, что $|U| = |U^*| = m$. Из неравенства $(v_i - v_j, v^*) \leqslant 0$ получаем

$$F(U) = \Big\| \sum_{v \in U} v \Big\|^2 = \Big\| \sum_{v \in U^*} v - v_i + v_j \Big\|^2$$
$$= \|v^*\|^2 + \|v_i - v_j\|^2 - 2(v_i - v_j, v^*) > \|v^*\|^2 = F(U^*),$$

что противоречит оптимальности U^*. Лемма 3 доказана.

Из леммы 3 следует, что оптимальное решение задачи m-ПВ состоит из m максимальных относительно порядка $\leqslant_{v^*}$ векторов, который, по сказанному выше, совпадает с линейным порядком $\leqslant_w$ для представителя w той расширенной ОПР, которая содержит вектор v^*.

Опишем теперь алгоритм A решения задачи m-ПВ.

Алгоритм A

Шаг 1. Для множества векторов $u_{ij}, 1 \leqslant i < j \leqslant n$, строится семейство W представителей ОПР. Для каждого $w \in W$ выполняются шаги 2–3.

Шаг 2. Векторы $v_1, v_2, \ldots, v_n$ переупорядочиваются по убыванию в соответствии с линейным порядком $\leqslant_w$; обозначим их через $v'_1, v'_2, \ldots, v'_n$, т. е. $v'_1 \geqslant_w v'_2 \geqslant_w \ldots \geqslant_w v'_n$.

Шаг 3. Полагается $U^w = \{v'_1, v'_2, \ldots, v'_m\}$.

Шаг 4. Выбирается такое $U \in \{U^w \mid w \in W\}$, что

$$F(U) = \max\{F(U^w) \mid w \in W\}.$$

Множество U является результатом работы алгоритма.

Имеет место следующая

Теорема 5. *Алгоритм A находит оптимальное решение задачи m-ПВ за время $O(k^2n^{2k})$.*

Доказательство. Покажем, что найденное алгоритмом решение является оптимальным. Пусть U^* — оптимальное решение задачи m-ПВ. Положим $v^* = \sum\limits_{v \in U^*} v$. Пусть w является представителем той расширенной ОПР, которая содержит v^*. Тогда линейные порядки $\leqslant_{v^*}$ и $\leqslant_w$ совпадают. Обозначим через $v'_1 \geqslant_w v'_2 \geqslant_w \ldots \geqslant_w v'_n$ переупорядоченные по убыванию в соответствии с этим линейным порядком векторы из V. По лемме 3 оптимальное решение задачи m-ПВ имеет вид $U^* = \{v'_1, v'_2, \ldots, v'_m\}$. Но тогда $F(U^*) = F(U^w) \leqslant F(U)$, откуда следует оптимальность решения U.

Определим трудоёмкость алгоритма. На шаге 1 ищется семейство представителей ОПР для $(n^2 - n)/2$ векторов. По лемме 1 оно содержит $O(kn^{2k-2})$ векторов, а для его нахождения требуется $O(k^2n^{2k})$ операций. Шаги 2 и 3 выполняются $O(kn^{2k-2})$ раз. При этом шаг 2 требует $O(kn \log n)$ операций (поскольку при каждом сравнении производится вычитание и скалярное перемножение k-мерных векторов), а шаг 3 — $O(n)$ операций. Шаг 4 выполняется за время $O(k^2n^{2k-1})$ (для каждого из векторов $w \in W$ вычисляется $F(U^w)$ за время kn). Таким образом,

временна́я сложность всего алгоритма не превосходит $O(k^2n^{2k})$. Теорема 5 доказана.

Для задачи ПВУ, очевидно, имеет место следующая

Лемма 3$'$. *Пусть U^* — оптимальное решение задачи* ПВУ *и $v^* = \sum\limits_{v\in U^*} v$. Тогда если $v_i \neq v_j$ и $v_i \leqslant_{v^*} v_j$, то либо $v_i \notin U^*$, либо $v_j \in U^*$.*

Доказательство. Предположим, что $v_i \in U^*$ и $v_j \notin U^*$. Положим $U = (U^* \setminus \{v_i\}) \cup \{v_j\}$. Очевидно, $|U| = |U^*|$. Тогда из неравенства $(v_i - v_j, v^*) \leqslant 0$ получаем

$$F(U) = \frac{\big\|\sum\limits_{v\in U} v\big\|^2}{|U|} = \frac{\big\|\sum\limits_{v\in U^*} v - v_i + v_j\big\|^2}{|U^*|}$$
$$= \frac{\|v^*\|^2 + \|v_i - v_j\|^2 - 2(v_i - v_j, v^*)}{|U^*|} > \frac{\|v^*\|^2}{|U^*|} = F(U*),$$

что противоречит оптимальности U^*. Лемма 3 доказана.

В задаче ПВУ количество выбираемых векторов в оптимальном решении $|U^*|$ является заранее неизвестной (переменной) величиной. Понятно, что оптимум этой задачи может быть найден в результате решения задач m-ПВ с $|U| = m$, $m = \overline{1,n}$, и выбором среди них решения с максимальным значением функции (1). Отсюда следует, что задача ПВУ разрешима за время, превышающее временну́ю сложность $O(k^2n^{2k})$ задачи m-ПВ не более чем в n раз. Однако далее покажем, что обе задачи разрешимы с одинаковой временно́й сложностью $O(k^2n^{2k})$. Действительно, алгоритм A' для решения задачи ПВУ получается из алгоритма A заменой шага 3 на следующий

Шаг 3$'$. Для каждого $i = 1, 2, \ldots, n$ рассматривается множество $U_i^w = \{v'_1, v'_2, \ldots, v'_i\}$ и вычисляется $F(U_i^w)$. Через U^w обозначим то U_i^w, в котором достигается максимум $F(U_i^w)$.

Теорема 5$'$. *Алгоритм A' находит оптимальное решение задачи* ПВУ *за время $O(k^2n^{2k})$.*

Доказательство. Оптимальность решения следует из леммы 3$'$ так же, как в теореме 5.

Трудоёмкость шага 3$'$ равна $O(kn^2)$ (вычисление n значений целевой функции, каждое за $O(kn)$ операций, выбор максимального). Таким образом, изменение шага не увеличивает итоговую трудоёмкость алгоритма. Теорема 5$'$ доказана.

Следствие. *При фиксированной размерности пространства* $\mathbb{R}^k$ *имеет место полиномиальная разрешимость задач* m-ПВ *и* ПВУ.

Заключительные замечания. Отметим некоторые нерешённые вопросы, связанные с задачами выбора подмножества векторов в евклидовом пространстве $\mathbb{R}^k$.

Обозначим через m-ПВО задачу m-ПВ с дополнительными ограничениями на номера выбираемых векторов. *Заданы натуральные числа* m *и* l, *удовлетворяющие условию* $lm < n$, *и конечная последовательность векторов* $V = (v_1, v_2, \ldots, v_n)$ *в евклидовом пространстве* $\mathbb{R}^k$. *Требуется выделить подпоследовательность векторов* $U = (v_{a_1}, v_{a_2}, \ldots, v_{a_m})$, *обладающую максимальной нормой суммы, при соблюдении ограничений на номера соседних векторов выделенной подпоследовательности:*

$$a_{i+1} - a_i \geqslant l \text{ для } i = 1, 2, \ldots, m-1. \qquad (2)$$

Сформулируем также задачу ПВУО, получаемую из ПВУ так же, как m-ПВО из m-ПВ. *Заданы натуральное число* l *и конечная последовательность векторов* $V = (v_1, v_2, \ldots, v_n)$ *в евклидовом пространстве* $\mathbb{R}^k$. *Требуется выделить подпоследовательность векторов* $U = (v_{a_1}, v_{a_2}, \ldots, v_{a_m})$, *максимизирующую функцию* $F(U)$ *из* (1) *с соблюдением ограничений* (2).

Подобного рода задачи возникают при анализе зашумлённых числовых последовательностей.

В [6] показано, что задача m-ПВО NP-трудна. Было бы интересно установить сложностной статус и для задачи ПВУО.

Поскольку в настоящей статье вопрос о полиномиальной разрешимости задач m-ПВ и ПВУ при фиксированной размерности пространства $\mathbb{R}^k$ удалось решить положительно, то представляется интересным разрешение этого вопроса и для задач m-ПВО и ПВУО.

Отметим, что вопрос о сложностном статусе задачи ПВ, рассмотренной в [1], к настоящему времени также остаётся открытым.

ЛИТЕРАТУРА

1. **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.
2. **Гимади Э. Х., Глазков Ю. В., Рыков И. А.** Задача выбора подмножества векторов с целочисленными координатами в евклидовом пространстве с максимальной нормой суммы // Дискрет. анализ и исслед. операций. — 2008. — Т. 15, № 4. — С. 31–43.

3. **Гимади Э. Х., Кельманов А. В., Кельманова М. А., Хамидуллин С. А.** Апостериорное обнаружение в числовой последовательности квазипериодически повторяющегося фрагмента при заданном числе повторов // Сиб. журн. индустр. математики. — 2006. — Т. 9, № 1. — С. 55–74.
4. **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.
5. **Кельманов А. В., Хамидуллин С. А., Окольнишникова Л. В.** Апостериорное обнаружение одинаковых подпоследовательностей-фрагментов в квазипериодической последовательности // Сиб. журн. индустр. математики. — 2002. — Т. 5, № 2. — С. 94–108.
6. **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, N 1. — P. 32–38.

Гимади Эдуард Хайрутдинович,
e-mail: gimadi@math.nsc.ru

Пяткин Артём Валерьевич,
e-mail: artem@math.nsc.ru

Рыков Иван Александрович,
e-mail: rykov@ngs.ru

Статья поступила
18 июля 2008 г.
Переработанный вариант —
14 октября 2008 г.