ДИСКРЕТНЫЙ АНАЛИЗ И ИССЛЕДОВАНИЕ ОПЕРАЦИЙ
Июль—август 2014. Том 21, № 4. С. 3–11

УДК 519.2+621.391

СЛОЖНОСТЬ ЗАДАЧИ О РАЗРЕЗЕ МАКСИМАЛЬНОГО ВЕСА В ЕВКЛИДОВОМ ПРОСТРАНСТВЕ *)

А. А. Агеев, А. В. Кельманов, А. В. Пяткин

Аннотация. Рассматривается задача поиска разреза максимального веса в полном неориентированном графе, вершинами которого являются точки q-мерного пространства. Анализируются два случая, в которых длины рёбер равны (i) евклидовым расстояниям между точками пространства и (ii) квадратам этих расстояний. Доказано, что в обоих случаях задача NP-трудна в сильном смысле. Показано также, что для них в предположении P≠NP не существует полностью полиномиальной приближённой схемы (FPTAS).

Ключевые слова: граф, разрез, евклидово пространство, NP-трудная задача.

Введение

Настоящая работа ориентирована на исследование вычислительной сложности задачи о разрезе максимального веса (задача Max-Cut) в двух геометрических случаях. В обоих случаях вершины графа заданы точками евклидова пространства, а длины рёбер — либо расстояниями (задача Euclidean Max-Cut), либо квадратами расстояний между этими точками (задача Quadratic Euclidean Max-Cut). Предполагается, что размерность пространства является частью входа задачи.

В теоретическом плане геометрические случаи с евклидовой нормой естественны для задач на графах с заданными длинами рёбер и традиционно в первую очередь исследуются на предмет полиномиальной разрешимости. Анализируемые случаи актуальны, в частности, в компьютерной геометрии [5], анализе данных и распознавании образов [7, 15]. Среди других приложений можно отметить статистическую физику [13], астрофизику [14], биологию [12], проектирование микросхем [4] и телекоммуникационных сетей [8] и др.

*) Исследование выполнено при финансовой поддержке Российского фонда фундаментальных исследований (проекты 12–01–00090, 12–01–00093, 12–01–00184 и 13–07–00070).

При решении, например, проблем кластеризации в анализе данных и распознавании образов требуется разбить совокупность имеющихся данных на группы, состоящие из похожих элементов. При этом элементы в группах (кластерах) должны быть более похожи друг на друга, чем на элементы из других кластеров. Можно считать, что в этих проблемах вершинам графа соответствуют элементы совокупности, а вес ребра в графе определяет близость элементов.

Если данные числовые, то близость элементов может быть определена через евклидово расстояние. В этом случае каждый из двух кластеров, получаемых в результате решения задачи Euclidean Max-Cut, очевидно, будет содержать элементы такие, что сумма двух сумм всевозможных попарных внутрикластерных евклидовых расстояний минимальна. Тогда каждый из этих кластеров будет содержать элементы, наиболее отличающиеся от элементов другого кластера.

В случае, когда числовые данные содержат ошибку измерения или возмущены помехой с неизвестным или гауссовским вероятностным распределением, как правило, применяется критерий минимума суммы квадратов уклонений. В этом случае близость элементов определяется через квадрат невязки, т. е. через квадрат евклидова расстояния между точками (см., например, [1–3] и цитированные там работы). Задача Quadratic Euclidean Max-Cut соответствует этому случаю.

Напомним, что во взвешенной задаче Max-Cut дан полный граф с неотрицательными весами на рёбрах. Требуется разбить множество вершин на две части таким образом, чтобы суммарный вес рёбер, соединяющих вершины из разных частей, был максимален. В невзвешенном случае веса рёбер принимают значения из множества $\{0, 1\}$, т. е. можно считать, что граф произвольный неориентированный и требуется разбить множество его вершин на две части так, что число рёбер между ними максимально.

В верификационной версии взвешенной задачи дан произвольный граф и число K, требуется разбить множество вершин на две части таким образом, чтобы суммарный вес рёбер, соединяющих части, был не меньше K. Верификационная версия взвешенной задачи числится под номером 21 в классическом списке Карпа базовых NP-полных задач [11]. NP-полнота верификационной версии невзвешенной задачи установлена в [10].

Известно, что задача Max-Cut в случае произвольного метрического пространства (когда вершины графа — точки пространства, а длины рёбер — расстояния между этими точками) NP-трудна в сильном смыс-

ле [17]. В [5, с. 325] сложностной статус задачи Euclidean Max-Cut формулируется как открытый вопрос. Статус сложности задачи Quadratic Euclidean Max-Cut до настоящего времени также оставался неясным. Ниже показываем, что обе евклидовы задачи NP-трудны в сильном смысле.

1. Анализ вычислительной сложности

Рассматриваемые задачи имеют следующие формулировки.

Задача Euclidean Max-Cut. Дано: множество $\mathcal{X} = \{x_1, \ldots, x_N\}$ точек из $\mathbb{R}^q$.

Найти: разбиение множества $\mathcal{X}$ на два подмножества $\mathcal{Y}$ и $\mathcal{Z}$ таких, что
$$f(\mathcal{Y}, \mathcal{Z}) = \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} \|y - z\| \longrightarrow \max .$$

Задача Quadratic Euclidean Max-Cut. Дано: множество $\mathcal{X} = \{x_1, \ldots, x_N\}$ точек из $\mathbb{R}^q$.

Найти: разбиение множества $\mathcal{X}$ на два подмножества $\mathcal{Y}$ и $\mathcal{Z}$ таких, что
$$g(\mathcal{Y}, \mathcal{Z}) = \sum_{y\in\mathcal{Y}} \sum_{z\in\mathcal{Z}} \|y - z\|^2 \longrightarrow \max .$$

Для выяснения вычислительной сложности этих задач нам потребуется следующая NP-трудная [9]

Задача Minimum Bisection. Дано: неориентированный граф G.

Найти: разбиение множества вершин графа G на две части равного размера (бисекцию) такое, что число рёбер между этими частями минимально.

Известно [6], что задача Minimum Bisection NP-трудна в случае кубических графов. Заметим, что кубический граф имеет чётное число вершин.

Статус сложности первой из рассматриваемых задач устанавливает

Теорема 1. *Задача* Euclidean Max-Cut NP-*трудна в сильном смысле.*

Доказательство. Построим полиномиальное сведение задачи Minimum Bisection для кубических графов к задаче Euclidean Max-Cut.

По примеру задачи Minimum Bisection на кубическом графе с числом вершин $n = 2k$ и числом рёбер m построим следующий пример задачи Euclidean Max-Cut. Пусть A — матрица с числом строк n и числом столбцов $n + m$, состоящая из двух частей: слева матрица инциденций кубического графа G размера $n \times m$, справа — диагональная матрица

размера $n \times n$ с диагональными элементами, равными положительному числу γ (позже определим его значение). В каждой строке матрицы A ровно 3 единицы (так как граф G кубический), число γ (в диагональной части строки), а остальные элементы равны нулю.

Положим $q = n+m$. Между строками матрицы A и входными точками задачи Euclidean Max-Cut установим взаимно однозначное соответствие. Координату $i = 1, \dots, q$ входной точки в задаче Euclidean Max-Cut положим равной i-му элементу строки матрицы A. Построим полный граф H, вершины которого — эти точки, а каждое ребро имеет вес, равный расстоянию между точками (в евклидовой метрике). Заметим, что вес разреза в графе H совпадает с значением целевой функцией задачи Euclidean Max-Cut.

Из построения следует, что две вершины в G смежны тогда и только тогда, когда в графе H расстояние между соответствующими точками (вершинами) равно $\sqrt{4+2\gamma^2}$. Для любой пары несмежных вершин в графе G расстояние между соответствующими точками в графе H равно $\sqrt{6+2\gamma^2}$.

По построению множества вершин графов G и H находятся во взаимно однозначном соответствии. Покажем, что бисекция графа G минимальна тогда и только тогда, когда соответствующая бисекция графа H имеет максимальный вес.

Пусть бисекция графа G содержит t рёбер. Тогда этой бисекции соответствует разрез в графе H с числом рёбер k^2, в котором t рёбер имеют одинаковый вес $\sqrt{4+2\gamma^2}$, а каждое из оставшихся рёбер имеет вес $\sqrt{6+2\gamma^2}$. Поэтому вес этого разреза равен

$$t\sqrt{4+2\gamma^2} + (k^2-t)(\sqrt{6+2\gamma^2}) = k^2\sqrt{6+2\gamma^2} - t(\sqrt{6+2\gamma^2} - \sqrt{4+2\gamma^2}).$$

Поскольку первое слагаемое в правой части — константа, при любом фиксированном γ бисекция в графе G минимальна тогда и только тогда, когда соответствующая бисекция графа H имеет максимальный вес.

Покажем, что при подходящем выборе значения константы γ в графе H любой разрез с неравными долями будет иметь вес, не превосходящий веса любой бисекции. Тем самым будет доказано, что бисекция максимального веса в графе H, соответствующая, как показано выше, минимальной бисекции кубического графа G, имеет максимальный вес среди всех разрезов графа H.

Пусть (U, W) — разрез с неравными размерами долей r_1 и r_2, $r_1 + r_2 = 2k$. Верхняя оценка на вес этого разреза равна $r_1 r_2(\sqrt{6+2\gamma^2})$, нижняя

оценка на вес бисекции равна $k^2\sqrt{4+2\gamma^2}$. Оценим их разность. Имеем

$$\begin{aligned} k^2\sqrt{4+2\gamma^2} - r_1r_2\sqrt{6+2\gamma^2} &\geqslant k^2\sqrt{4+2\gamma^2} - (k^2-1)\sqrt{6+2\gamma^2} \\ &= \sqrt{6+2\gamma^2} - k^2(\sqrt{6+2\gamma^2} - \sqrt{4+2\gamma^2}), \end{aligned}$$

поскольку для произведения размеров долей справедлива оценка $r_1r_2 \leqslant k^2 - 1$.

Возьмём $\gamma = k^2$. Тогда для рассматриваемой разности верна оценка снизу:

$$\sqrt{6+2k^4} - k^2(\sqrt{6+2k^4} - \sqrt{4+2k^4}) > k^2(\sqrt{2} - (\sqrt{6+2k^4} - \sqrt{4+2k^4})).$$

Легко показать (взятием производной), что $\sqrt{6+2h} - \sqrt{4+2h}$ — монотонно убывающая функция от h; даже при $h = 1$ её значение строго меньше $\sqrt{2}$. Поэтому при $\gamma = k^2$ вес произвольного разреза с неравными долями строго меньше веса любой бисекции, стало быть, разрез максимального веса в графе H обязан быть бисекцией.

Таким образом, при определённом выше значении числа γ разрезу максимального веса в построенном примере задачи Euclidean Max-Cut соответствует бисекция минимального веса в исходном примере задачи Minimum Bisection. Теорема 1 доказана.

Замечание 1. Сильная NP-трудность следует из того, что NP-трудная задача с нечисловым входом сведена к задаче Euclidean Max-Cut с числовыми параметрами, ограниченными полиномом от размера входа задачи.

Сложностной статус второй задачи устанавливает

Теорема 2. *Задача* Quadratic Euclidean Max-Cut NP-*трудна в сильном смысле.*

Доказательство теоремы проводится аналогичным сведéнием, а именно, по примеру задачи Minimum Bisection строится пример задачи Quadratic Euclidean Max-Cut с идентичной структурой матрицы A. Отличие состоит в том, что в полном графе H весами теперь являются квадраты евклидовых расстояний между точками (вершинами). При этом бисекции графа G, содержащей t рёбер, будет соответствовать бисекция графа H, вес которой в задаче Quadratic Euclidean Max-Cut равен

$$t(4+2\gamma^2) + (k^2 - t)(6+2\gamma^2) = k^2(6+2\gamma^2) - 2t.$$

Отсюда (как и при доказательстве теоремы 1) следует, что при любом фиксированном γ бисекция в графе G минимальна тогда и только тогда, когда соответствующая бисекция графа H имеет максимальный вес.

Далее, для разности нижней оценки на вес бисекции и верхней оценки на вес разреза с неравными долями в графе H аналогичным образом получим

$$\begin{aligned}k^2(4+2\gamma^2) - r_1 r_2(6+2\gamma^2) &\geqslant k^2(4+2\gamma^2) - (k^2-1)(6+2\gamma^2)\\ &= (6+2\gamma^2) - k^2((6+2\gamma^2)-(4+2\gamma^2)) = 6+2\gamma^2-2k^2.\end{aligned}$$

Возьмём $\gamma = k$. Тогда в графе H произвольный разрез с неравными долями будет строго меньше веса любой бисекции. Следовательно, разрез максимального веса в графе H является бисекцией.

Таким образом, при определённом выше значении числа γ разрезу максимального веса в построенном примере задачи Quadratic Euclidean Max-Cut соответствует бисекция минимального веса в исходном примере задачи Minimum Bisection. Теорема 2 доказана.

Сильная NP-трудность устанавливается в соответствии с замечанием 1.

2. Несуществование FPTAS

Для задачи Quadratic Euclidean Max-Cut несуществование FPTAS (в предположении P≠NP) следует из целочисленности целевой функции в примере, возникающем при сведе́нии в доказательстве теоремы 2 [16, теорема 8.5].

Для задачи Euclidean Max-Cut в соответствующем примере (см. доказательство теоремы 1) целевая функция нецелочисленная, однако нетрудно доказать, что FPTAS также не существует.

Рассуждаем от противного. Предположим, что для задачи Euclidean Max-Cut существует алгоритм, который при любом значении относительной погрешности ε за время, ограниченное полиномом от длины входа задачи и $1/\varepsilon$, находит разрез, вес которого не менее чем в $(1-\varepsilon)$ раз отличается от веса оптимального. Применим этот алгоритм к примеру, возникающему в доказательстве теоремы 1 при сведе́нии задачи Minimum Bisection на кубическом графе. Легко видеть, что для этого примера разрезы могут отличаться по весу как минимум на величину $\sqrt{6+2k^4} - \sqrt{4+2k^4}$.

Мы утверждаем что при $\varepsilon = 1/8k^6$

$$(1-\varepsilon)\mathrm{OPT} > \mathrm{OPT} - (\sqrt{6+2k^4} - \sqrt{4+2k^4}),$$

где OPT — вес оптимального разреза. Это неравенство очевидно влечёт существование точного полиномиального алгоритма для Euclidean Max-Cut, что противоречит утверждению теоремы 1.

В силу того, что $\mathrm{OPT} \leqslant k^2(\sqrt{6+2k^4})$, достаточно показать, что

$$\varepsilon = \frac{1}{8k^6} < \frac{1}{k^2}\left(1 - \sqrt{\frac{4+2k^4}{6+2k^4}}\right).$$

Далее следуют элементарные преобразования, которые опускаем (перегруппировка членов и возведение обеих частей неравенства в квадрат с целью избавления от радикалов). В итоге после замены $x = k^4$ останется проверить положительность квадратного трёхчлена $48x^2 - 47x + 3$ при $x \geqslant 1$ (например, взятием производной в точке $x = 1$).

Заключение

В соответствии с [9] из приведённых результатов следует, что для рассмотренных задач не существует точных полиномиального и псевдополиномиального алгоритмов (в предположении, что гипотеза P≠NP верна). Нами также показано, что для этих задач не существует полностью полиномиальной приближённой схемы (FPTAS).

Таким образом, сложностной статус рассмотренных специальных геометрических случаев задачи Max-Cut оказывается ничуть не лучше, чем у общего случая этой задачи.

ЛИТЕРАТУРА

1. **Кельманов А. В., Пяткин А. В.** О сложности одного из вариантов задачи выбора подмножества «похожих» векторов // Докл. АН. — 2008. — Т. 421, № 5. — С. 590–592.
   Kel'manov A. V., Pyatkin A. V. On the complexity of a search for a subset of "similar" vectors // Dokl. Math. — 2008. — Vol. 78, N 1. — P. 574–575.
2. **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.
   Kel'manov A. V., Pyatkin A. V. NP-completeness of some problems of choosing a vector subset // J. Appl. Industr. Math. — 2011. — Vol. 5, N 3. — P. 352–357.
3. **Кельманов А. В., Пяткин А. В.** О сложности некоторых задач кластерного анализа векторных последовательностей // Дискрет. анализ и исслед. операций. — 2013. — Т. 20, № 2. — С. 47–57.
   Kel'manov A. V., Pyatkin A. V. On complexity of some problems of cluster analysis of vector sequences // J. Appl. Industr. Math. — 2013. — Vol. 7, N 3. — P. 363–369.
4. **Barahona F., Grötschel M., Jünger M., Reinelt G.** An application of combinatorial optimization to statistical physics and circuit layout design // Oper. Res. — 1988. — Vol. 36. — P. 493–513.

5. **Bern M., Eppstein D.** Approximation algorithms for geometric problems // Approximation algorithms for NP-hard problems. — Boston: PWS Publ., 1997. — P. 296–345.
6. **Bui T. N., Chaudhuri S., Leighton F. T., Sipser M.** Graph bisection algorithms with good average case behavior // Combinatorica. — 1987. — Vol. 7, N 2. — P. 171–191.
7. **Ding C. H. Q., He X., Zha H., Gu M., Simon H. D.** A min-max cut algorithm for graph partitioning and data clustering // Proc. IEEE Int. Conf. Data Mining (San Jose, Nov. 29–Dec. 2, 2001). — Los Alamitos; Washington; Brussels; Tokyo: IEEE Comput. Soc., 2001. — P. 107–114.
8. **Eisenblätter A.** The semidefinite relaxation of the k-partition polytope is strong // Proc. 9th Int. IPCO Conf. Integer Programming and Combinatorial Optimization (Cambridge, MA, May 27–29, 2002). Berlin; New York: Springer-Verl., 2002. — P. 273–290. (Lect. Notes Comput. Sci.; Vol. 2337).
9. **Garey M. R., Johnson D. S.** Computers and intractability: a guide to the theory of NP-completeness. — San Francisco: Freeman, 1979. — 314 p.
10. **Garey M. R., Johnson D. S., Stockmeyer L.** Some simplified NP-complete graph problems // Theoret. Comput. Sci. — 1976. — Vol. 3, N 1. — P. 237–267.
11. **Karp R. M.** Reducibility among combinatorial problems // Complexity of computer computations. — New York: Plenum Press, 1972. — P. 85–103.
12. **Karp R. M.** The genomics revolution and its challenges for algorithmic research // Current trends in theoretical computer science: entering the 21st century. — River Edge, NJ: World Sci. Publ. Co., Inc., 2001. — P. 631–642.
13. **Liers F., Jünger M., Reinelt G., Rinaldi G.** Computing exact ground states of hard ising spin glass problems by branch-and-cut // New Optimization Algorithms in Physics. — Weinheim: Wiley-VCH Verl., 2004. — P. 47–68.
14. **Lupton R., Maley M. P., Young N. E.** Data collection for the Sloan digital sky survey: a network-flow heuristic // J. Algorithms. — 1998. — Vol. 27, N 2. — P. 339–356.
15. **De Sousa S., Haxhimusa Y., Kropatsch W. G.** Estimation of distribution algorithm for the Max-Cut problem // Proc. 9th IAPR-TC-15 Int. Workshop Graph-Based Representations in Pattern Recognition (Vienna, Austria, May 15–17, 2013). — Heidelberg; Dordrecht; London; New York: Springer-Verl., 2013. — P. 244–253.
16. **Vazirani V. V.** Approximation algorithms. — Berlin; Heidelberg; New York: Springer-Verl., 2001. — 380 p.

17. **De la Vega F., Kenyon C.** A randomized approximation scheme for metric max-cut // J. Comput. Syst. Sci. — 2001. — Vol. 63. — P. 531–541.

Агеев Александр Александрович,
e-mail: ageev@math.nsc.ru
Кельманов Александр Васильевич,
e-mail: kelm@math.nsc.ru
Пяткин Артём Валерьевич,
e-mail: kelm@math.nsc.ru

Статья поступила
3 декабря 2013 г.

Переработанный вариант —
10 февраля 2014 г.