МАТЕМАТИЧЕСКИЕ ОСНОВЫ ИНФОРМАТИКИ И ПРОГРАММИРОВАНИЯ

УДК 510.52

О ГЕНЕРИЧЕСКОЙ СЛОЖНОСТИ ПРОБЛЕМЫ ПРЕДСТАВИМОСТИ НАТУРАЛЬНЫХ ЧИСЕЛ СУММОЙ ДВУХ КВАДРАТОВ[1]

А. Н. Рыбалов

Институт математики им. С. Л. Соболева СО РАН, г. Омск, Россия

Генерический подход к алгоритмическим проблемам был предложен Мясниковым, Каповичем, Шуппом и Шпильрайном в 2003 г. В рамках этого подхода рассматривается поведение алгоритмов на множествах почти всех входов. В работе изучается генерическая сложность проблемы представимости натуральных чисел суммой двух квадратов. Данная проблема, восходящая ещё к Ферма и Эйлеру, тесно связана с проблемой факторизации целых чисел и проблемой распознавания квадратичности вычетов по составным модулям, для решения которых не известно эффективных алгоритмов. Доказывается, что, при условии трудноразрешимости этой проблемы в худшем случае и P = BPP, для её решения не существует полиномиального сильно генерического алгоритма. Сильно генерический алгоритм решает проблему не на всём множестве входов, а на подмножестве, последовательность относительных плотностей которого при увеличении размера экспоненциально быстро сходится к единице. Для доказательства используется метод генерической амплификации, который позволяет строить генерически трудные проблемы из проблем, трудных в худшем случае. Основным ингредиентом метода является объединение эквивалентных входов в достаточно большие множества. Эквивалентность входов означает, что рассматриваемая проблема на них решается одинаково.

Ключевые слова: *генерическая сложность, суммы квадратов, диофантовы уравнения.*

DOI 10.17223/20710410/48/8

ON GENERIC COMPLEXITY OF THE PROBLEM OF REPRESENTATION OF NATURAL NUMBERS BY SUM OF TWO SQUARES

A. N. Rybalov

Sobolev Institute of Mathematics, Omsk, Russia

E-mail: alexander.rybalov@gmail.com

[1]Исследование поддержано Программой фундаментальных научных исследований СО РАН I.1.1.4, проект № 0314-2019-0004.

Generic-case approach to algorithmic problems was suggested by Miasnikov, Kapovich, Schupp and Shpilrain in 2003. This approach studies behavior of an algorithm on typical (almost all) inputs and ignores the rest of inputs. In this paper, we study the generic complexity of the problem of representation of natural numbers by sum of two squares. This problem, going back to Fermat and Euler, is closely related to the problem of integer factorization and the quadratic residuosity problem modulo composite numbers, for which no efficient algorithms are known. We prove that under the condition of worst-case hardness and P = BPP, for the problem of representation of natural numbers by sum of two squares there is no polynomial strongly generic algorithm. A strongly generic algorithm solves a problem not on the whole set of inputs, but on a subset, the sequence of frequencies which with increasing size converges exponentially fast to 1. To prove this theorem we use the method of generic amplification, which allows to construct generically hard problems from the problems hard in the classical sense. The main ingredient of this method is a technique of cloning, which unites inputs of the problem together in the large enough sets of equivalent inputs. Equivalence is understood in the sense that the problem is solved similarly for them.

Keywords: *generic complexity, sums of squares, Diophantine equations.*

Введение

Проблема представимости натуральных чисел суммой двух квадратов состоит в том, чтобы по любому заданному натуральному числу N определить, разрешимо ли в натуральных числах диофантово уравнение $x^2 + y^2 = N$. Эта задача восходит ещё к Ферма, который в 1640 г. сформулировал (см. [1, 2]) следующее красивое утверждение: любое простое число вида $p = 4n + 1$ представимо в виде суммы квадратов двух натуральных чисел. Эта гипотеза впоследствии была доказана Эйлером и называется теперь теоремой Ферма — Эйлера (см. [1, 2]). В дальнейшем был получен критерий Ферма — Эйлера разрешимости диофантова уравнения $x^2 + y^2 = N$ для любого натурального N. Однако этот критерий сводит проблему к задаче факторизации (разложения на множители) целых чисел, которая на текущий момент считается трудноразрешимой [3]. Таким образом, критерий Ферма — Эйлера не может быть проверен эффективно (за полиномиальное от размера входа время). Кроме того, проблема представимости натуральных чисел суммой двух квадратов тесно связана с проблемой распознавания квадратичности вычетов по составным модулям, для которой тоже не известно эффективных алгоритмов [3].

Генерический подход к алгоритмическим проблемам предложен в [4]: алгоритмическая проблема рассматривается не на всём множестве входов, а на некотором подмножестве «почти всех» входов. Такие входы образуют генерическое множество. Понятие «почти все» формализуется введением естественной меры на множестве входных данных. Проблема может быть трудноразрешимой или вообще неразрешимой в классическом смысле, но легкоразрешимой на генерическом множестве.

В данной работе изучается генерическая сложность проблемы представимости натуральных чисел суммой двух квадратов. Доказывается, что, при условии трудноразрешимости этой проблемы в худшем случае и P = BPP, для неё не существует полиномиального сильно генерического алгоритма. Сильно генерический алгоритм решает проблему не на всём множестве входов, а на подмножестве, последовательность относительных плотностей которого при увеличении размера экспоненциально быстро сходится к единице. Класс BPP состоит из проблем, разрешимых за полиномиальное время на вероятностных машинах Тьюринга. Одной из важных гипотез в теории слож-

ности вычислений является гипотеза о совпадении классов P и BPP. Из неё следует, что любой полиномиальный вероятностный алгоритм $\mathcal{A}$ можно эффективно дерандомизировать, то есть построить полиномиальный алгоритм $\mathcal{B}$, не использующий генератор случайных чисел и решающий ту же проблему, что и алгоритм $\mathcal{A}$. В [5] доказано, что равенство P = BPP следует из весьма правдоподобных гипотез о вычислительной сложности некоторых трудных проблем.

1. Генерические алгоритмы

Пусть I — некоторое множество входов. Для подмножества $S \subseteq I$ определим *последовательность относительных плотностей*

$$\rho_n(S) = \frac{|S_n|}{|I_n|}, \ n = 1, 2, 3, \ldots,$$

где I_n — множество входов размера n; $S_n = S \cap I_n$. Заметим, что $\rho_n(S)$ — это вероятность попасть в S при случайной и равновероятной генерации входов из I_n. В данной работе множеством входов для алгоритмов является множество натуральных чисел, записанных в двоичной форме. Под размером натурального числа понимается длина его двоичной записи.

Асимптотической плотностью множества S назовём верхний предел

$$\rho(S) = \overline{\lim_{n\to\infty}} \rho_n(S).$$

Множество S называется *генерическим*, если $\rho(S) = 1$, и *пренебрежимым*, если $\rho(S) = 0$. Очевидно, что S генерическое тогда и только тогда, когда его дополнение $I \setminus S$ пренебрежимо.

Следуя [4], назовём множество S *сильно пренебрежимым*, если последовательность $\rho_n(S)$ экспоненциально быстро сходится к нулю, т. е. существуют константы σ, $0 < \sigma < 1$, и $C > 0$, такие, что для любого n

$$\rho_n(S) < C\sigma^n.$$

Теперь S называется *сильно генерическим*, если его дополнение $I \setminus S$ сильно пренебрежимо.

Алгоритм $\mathcal{A}$ с множеством входов I и множеством выходов $J \cup \{?\}$ $(? \notin J)$ называется (*сильно*) *генерическим*, если

1) $\mathcal{A}$ останавливается на всех входах из I;
2) множество $\{x \in I : \mathcal{A}(x) \neq ?\}$ является (сильно) генерическим.

Генерический алгоритм $\mathcal{A}$ вычисляет функцию $f : I \to J$, если $(\mathcal{A}(x) = y \in J) \Rightarrow (f(x) = y)$ для всех $x \in I$. Ситуация $\mathcal{A}(x) = ?$ означает, что $\mathcal{A}$ не может вычислить функцию f на аргументе x. Но условие 2 гарантирует, что $\mathcal{A}$ корректно вычисляет f на почти всех входах (входах из генерического множества). Множество $S \subseteq I$ называется (*сильно*) *генерически разрешимым за полиномиальное время*, если существует (сильно) генерический полиномиальный алгоритм, вычисляющий его характеристическую функцию.

2. Проблема представимости натуральных чисел суммой двух квадратов

Проблема представимости натуральных чисел суммой двух квадратов состоит в следующем. Дано натуральное число N, записанное в двоичной системе. Нужно

определить, разрешимо ли в натуральных числах диофантово уравнение $x^2 + y^2 = N$. Классический критерий Ферма — Эйлера (см. [1, 2]) связывает эту проблему с известной проблемой факторизации целых чисел.

Теорема 1 (Ферма, Эйлер). Пусть N — натуральное число. Диофантово уравнение $N = x^2 + y^2$ разрешимо в натуральных числах тогда и только тогда, когда каждый простой делитель N вида $4k + 3$ входит в разложение N в чётной степени.

Если бы проблема факторизации решалась эффективно, то этот критерий давал бы эффективный алгоритм для проблемы представимости натуральных чисел суммой двух квадратов. Однако до сих пор неизвестно эффективных алгоритмов факторизации [3]. Кроме того, проблема представимости натуральных чисел суммой двух квадратов тесно связана с проблемой распознавания квадратичности вычетов по составным модулям, которая тоже считается трудноразрешимой [3].

В дальнейшем нам потребуется простое следствие из теоремы Ферма — Эйлера.

Лемма 1. Пусть N и M — натуральные числа, такие, что диофантовы уравнения $M = x^2 + y^2$ и $NM = x^2 + y^2$ разрешимы в натуральных числах. Тогда диофантово уравнение $N = x^2 + y^2$ тоже разрешимо в натуральных числах.

Доказательство. По теореме Ферма — Эйлера и в разложение M, и в разложение NM любой простой делитель вида $4k + 3$ входит в чётной степени. Отсюда следует, что то же верно и для числа N. Поэтому по теореме Ферма — Эйлера уравнение $N = x^2 + y^2$ разрешимо в натуральных числах. ■

Потребуется также следующая классическая теорема [6].

Теорема 2 (Гаусс). Пусть $N(r)$ есть число решений неравенства $x^2 + y^2 \leqslant r$ в натуральных числах. Тогда имеет место оценка

$$|N(r) - \pi r/4| \leqslant \frac{\sqrt{2}\pi\sqrt{r}}{2}.$$

Непосредственно из теоремы Гаусса выводится следующая

Лемма 2. Пусть S_n есть число решений неравенства $2^n \leqslant x^2 + y^2 \leqslant 2^{n+1}$ в натуральных числах. Тогда при $n > 8$ имеет место

$$S_n > \frac{\pi 2^n}{8}.$$

Доказательство. Очевидно, что $S_n = N(2^{n+1}) - N(2^n)$. С помощью теоремы Гаусса дадим верхнюю и нижнюю оценки:

$$N(2^{n+1}) \geqslant \frac{\pi 2^{n+1}}{4} - \frac{\sqrt{2}\pi 2^{(n+1)/2}}{2}, \quad N(2^n) \leqslant \frac{\pi 2^n}{4} + \frac{\sqrt{2}\pi 2^{n/2}}{2}.$$

Поэтому $S_n \geqslant \frac{\pi 2^n}{4} - \pi 2^{(n+2)/2} = \pi(2^{n-2} - 2^{(n+2)/2}) > \pi 2^{n-3}$ при $n > 8$. ■

3. Основной результат

Теорема 3. Если существует сильно генерический полиномиальный алгоритм, решающий проблему представимости натуральных чисел суммой двух квадратов, то существует вероятностный полиномиальный алгоритм, разрешающий эту проблему на всем множестве входов.

Доказательство. Допустим, что существует сильно генерический полиномиальный алгоритм $\mathcal{A}$, решающий проблему представимости натуральных чисел суммой двух квадратов. Построим вероятностный полиномиальный алгоритм $\mathcal{B}$, решающий эту проблему на всём множестве входов. На натуральном числе N размера n ($2^n \leqslant N < 2^{n+1}$) алгоритм $\mathcal{B}$ будет работать следующим образом:

1) Генерирует случайно и равновероятно натуральное число M вида $x^2 + y^2$ размера n^2 за n раундов:
   а) случайно, равновероятно и независимо генерируются натуральные числа x и y из отрезка $[0, \ldots, [\sqrt{2^{n^2+1}}]]$;
   б) проверяется условие $2^{n^2} \leqslant x^2 + y^2 < 2^{n^2+1}$;
   в) если условие выполнено, процесс генерации заканчивается;
   г) если условие не выполнено, то алгоритм возвращается к п. а;
   д) если n раундов не дали результата, то алгоритм выдаёт 0.
2) Запускает алгоритм $\mathcal{A}$ на числе NM.
3) Если $\mathcal{A}(NM) \neq ?$, то, по лемме 1, ответ $\mathcal{A}(NM)$ является решением проблемы представимости натуральных чисел суммой двух квадратов и для числа N.
4) Если $\mathcal{A}(NM) = ?$, то выдаёт ответ 0.

Заметим, что полиномиальный вероятностный алгоритм $\mathcal{B}$ выдаёт правильный ответ на шаге 3, а на шагах 1 и 4 может выдать неправильный ответ. Нужно доказать, что вероятность того, что ответ выдаётся на шагах 1 или 4, меньше $1/2$.

Процесс генерации на шаге 1 можно мыслить как бросание случайной точки (x, y) с целыми координатами внутрь квадрата $[0, \ldots, [\sqrt{2^{n^2+1}}]] \times [0, \ldots, [\sqrt{2^{n^2+1}}]]$. При этом генерация успешна, если точка попала в область $2^{n^2} \leqslant x^2 + y^2 < 2^{n^2+1}$. По лемме 2, вероятность успеха больше

$$\frac{\pi 2^{n^2}}{8 \cdot 2^{n^2+1}} = \frac{\pi}{16}.$$

Поэтому вероятность того, что в течение всех n раундов мы не получим успешной генерации числа N, меньше

$$\left(1 - \frac{\pi}{16}\right)^n < \frac{1}{4}$$

при достаточно больших n.

Теперь оценим вероятность выдачи ответа на шаге 4. Число A имеет размер n^2, значит, размер числа NM равен $m = n^2 + n$. Вероятность того, что для NM имеет место $\mathcal{A}(NM) = ?$, не больше

$$\frac{|\{K \in \mathbb{N} : \mathcal{A}(K) \neq ?\}_m|}{|\{N(x^2+y^2) : 2^{n^2} \leqslant x^2 + y^2 \leqslant 2^{n^2+1}\}_m|} =$$
$$= \frac{|\{K \in \mathbb{N} : \mathcal{A}(K) \neq ?\}_m|}{|\mathbb{N}_m|} \frac{|\mathbb{N}_m|}{|\{N(x^2+y^2) : 2^{n^2} \leqslant x^2 + y^2 \leqslant 2^{n^2+1}\}_m|}.$$

Так как множество $\{K \in \mathbb{N} : \mathcal{A}(K) \neq ?\}$ сильно пренебрежимое, то существует константа $\alpha > 0$, такая, что

$$\frac{|\{K \in \mathbb{N} : \mathcal{A}(K) \neq ?\}_m|}{|\mathbb{N}_m|} < \frac{1}{2^{\alpha m}} = \frac{1}{2^{\alpha(n^2+n)}}$$

для любого n. С другой стороны, по лемме 2

$$|\{N(x^2+y^2) : 2^{n^2} \leqslant x^2 + y^2 \leqslant 2^{n^2+1}\}_m| > \frac{\pi 2^m}{8N} > \frac{\pi 2^m}{8 \cdot 2^{n+1}}.$$

Отсюда

$$\frac{|\mathbb{N}_m|}{|\{N(x^2+y^2) : 2^{n^2} \leqslant x^2+y^2 \leqslant 2^{n^2+1}\}_m|} < \frac{2^m \cdot 8 \cdot 2^{n+1}}{\pi 2^m} = \frac{2^{n+4}}{\pi}.$$

Искомая вероятность ответа на шаге 4 не больше $\dfrac{2^{n+4}}{\pi 2^{\alpha(n^2+n)}} < \dfrac{1}{4}$ при больших n. Итого, вероятность ответа на шагах 1 или 4 меньше $1/4 + 1/4 = 1/2$. ■

Напомним, что класс BPP состоит из проблем, разрешимых за полиномиальное время на вероятностных машинах Тьюринга. Одной из важных гипотез в теории сложности вычислений является гипотеза о совпадении классов P и BPP. Из неё следует, что любой полиномиальный вероятностный алгоритм $\mathcal{A}$ можно эффективно дерандомизировать, то есть построить полиномиальный алгоритм $\mathcal{B}$, не использующий генератор случайных чисел и решающий ту же проблему, что и алгоритм $\mathcal{A}$. В [5] доказано, что равенство P = BPP следует из весьма правдоподобных гипотез о вычислительной сложности некоторых трудных проблем.

Теорема 4. Если проблема представимости натуральных чисел суммой двух квадратов не лежит в классе P и P = BPP, то не существует сильно генерического полиномиального алгоритма для этой проблемы.

Доказательство. Пусть существует сильно генерический алгоритм, решающий проблему представимости натуральных чисел суммой двух квадратов. Тогда по теореме 3 существует вероятностный полиномиальный алгоритм, решающий её на всём множестве входов, т. е. эта проблема лежит в классе BPP. Так как P = BPP, то она лежит и в классе P, что противоречит посылке теоремы. ■

Автор выражает благодарность рецензенту за полезные замечания и предложения по улучшению текста статьи.

ЛИТЕРАТУРА

1. *Dickson L. E.* History of the Theory of Numbers. V. II. N.Y.: Dover Publications, 2005. 803 p.
2. *Сендеров В., Спивак А.* Суммы квадратов и целые гауссовы числа // Квант. 1999. № 3. С. 14–22.
3. *Adleman L. M. and McCurley K. S.* Open problems in number theoretic complexity, II // Proc. First Intern. Symp. Algorithmic Number Theory. N.Y., USA, May 6–9, 1994. P. 291–322.
4. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks // J. Algebra. 2003. V. 264. No. 2. P. 665–694.
5. *Impagliazzo R. and Wigderson A.* P = BPP unless E has subexponential circuits: Derandomizing the XOR Lemma. Proc. 29th STOC. El Paso: ACM, 1997. P. 220–229.
6. *Hardy G. H. and Ramanujan S.* Twelve Lectures on Subjects Suggested by His Life and Work. N.Y.: Chelsea, 1999. 67 p.

REFERENCES

1. *Dickson L. E.* History of the Theory of Numbers, vol. II. N.Y., Dover Publications, 2005. 803 p.
2. *Senderov V. and Spivak A.* Summy kvadratov i celye gaussovy chisla [Sums of squares and gaussian integers]. Quant, 1999, no. 3, pp. 14–22. (in Russian)
3. *Adleman L. M. and McCurley K. S.* Open problems in number theoretic complexity, II. Proc. First Intern. Symp. Algorithmic Number Theory, N.Y., USA, May 6–9, 1994, pp. 291–322.
4. *Kapovich I., Miasnikov A., Schupp P., and Shpilrain V.* Generic-case complexity, decision problems in group theory and random walks. J. Algebra, 2003, vol. 264, no. 2, pp. 665–694.

5. *Impagliazzo R. and Wigderson A.* P = BPP unless E has subexponential circuits: Derandomizing the XOR Lemma. Proc. 29th STOC, El Paso, ACM, 1997, pp. 220–229.
6. *Hardy G. H. and Ramanujan S.* Twelve Lectures on Subjects Suggested by His Life and Work. N.Y., Chelsea, 1999. 67 p.