О представлении чисел в виде суммы двух квадратов

М. Н. Вялый

Какие целые числа являются суммами двух квадратов? Ответ на этот вопрос известен: число представляется в виде суммы двух квадратов тогда и только тогда, когда в его разложение на простые множители каждое простое число вида $4k+3$ входит в четной степени.

Как найти представление числа n в виде суммы двух квадратов? Конечно, можно перебрать все пары $0 \leqslant x, y \leqslant n$ и проверить для каждой пары равенство $x^2 + y^2 = n$. Нельзя ли придумать алгоритм, который решает эту задачу быстрее?

Эффективных алгоритмов для представления числа в виде суммы двух квадратов пока нет. Но задача значительно упрощается, если известен корень из -1 по модулю n. В этом случае можно обойтись без перебора.

Алгоритм Эрмита – Серре.

1. Вход: n, z, причем $z^2 \equiv -1 \pmod{n}$.
2. Полагаем $a_0 = n$, $a_1 = z$.
3. К паре чисел a_0, a_1 применяем алгоритм Евклида, т. е. вычисляем последовательность остатков от деления a_i на a_{i+1}:
$$a_i = q_i a_{i+1} + a_{i+2}\,.$$
4. Останавливаемся в тот момент, когда два последовательных остатка не превосходят $\sqrt{n}$: $a_t \leqslant \sqrt{n}$, $a_{t+1} \leqslant \sqrt{n}$. Это и есть нужная пара чисел:
$$a_t^2 + a_{t+1}^2 = n.$$

Числа в последовательности остатков, возникающей в алгоритме Евклида, быстро убывают. Поэтому вычисления по алгоритму Эрмита – Серре занимают гораздо меньше времени, чем перебор всех возможностей.

Откуда берется связь между алгоритмом Евклида и представлениями числа в виде суммы двух квадратов? Есть два, содержательно очень близких, способа обосновать алгоритм Эрмита – Серре. Первый основан на связи между алгоритмом Евклида и цепными дробями и использует

симметричность разложения в цепную дробь числа n/z [2]. Второй предложен А. ван дер Портеном [4] (см. также [3]) и использует симметричные разложения матриц размера 2×2. Ниже это доказательство излагается подробно.

Мы опишем связь между алгоритмом Евклида и матрицами размера 2×2 с помощью функции «скобки», которая определена на последовательностях натуральных (положительных целых) чисел следующими рекуррентными соотношениями:

$$[\,] = 1, \quad [q] = q, \quad [q_0, q_1, q_2, \ldots, q_n] = q_0[q_1, q_2, \ldots, q_n] + [q_2, \ldots, q_n]. \tag{1}$$

Например, $[q_0, q_1] = q_0[q_1] + [\,] = q_0 q_1 + 1$.

Непосредственно из определения видно, что эта функция положительна на любой последовательности $q_0, q_1, q_2, \ldots, q_n$ положительных целых чисел. Поэтому

$$[q_0, q_1, \ldots, q_n] > [q_1, \ldots, q_n]. \tag{2}$$

Таким образом, из (1) следует, что применение алгоритма Евклида к паре чисел $[q_0, q_1, q_2, \ldots, q_n]$, $[q_1, q_2, \ldots, q_n]$ дает последовательность остатков $a_i = [q_i, \ldots, q_n]$, при этом последовательностью частных будет $q_0, \ldots, q_n$. Отсюда вытекает следующая

Лемма 1. *Представление пары взаимно простых чисел a, b в виде*

$$a = [q_0, q_1, q_2, \ldots, q_n], \quad b = [q_1, q_2, \ldots, q_n] \tag{3}$$

однозначно определено.

Заметим также, что та же самая последовательность $q_0, q_1, q_2, \ldots, q_n$ возникает при разложении a/b в цепную дробь.

Шаг алгоритма Евклида можно представить в матричном виде

$$\begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & -q_{i-1} \end{pmatrix} \begin{pmatrix} a_{i-1} \\ a_i \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} a_{i-1} \\ a_i \end{pmatrix} = \begin{pmatrix} q_{i-1} & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_i \\ a_{i+1} \end{pmatrix}.$$

Мы будем использовать произведения матриц второго вида. Матричные элементы таких произведений легко выражаются через скобки $[\cdot]$ с помощью матричного равенства

$$\begin{pmatrix} [q_0, \ldots, q_n] & [q_0, \ldots, q_{n-1}] \\ [q_1, \ldots, q_n] & [q_1, \ldots, q_{n-1}] \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} [q_1, \ldots, q_n] & [q_1, \ldots, q_{n-1}] \\ [q_2, \ldots, q_n] & [q_2, \ldots, q_{n-1}] \end{pmatrix}, \tag{4}$$

которое проверяется прямым вычислением. Из (4) по индукции следует, что

$$\begin{pmatrix} [q_0, \ldots, q_n] & [q_0, \ldots, q_{n-1}] \\ [q_1, \ldots, q_n] & [q_1, \ldots, q_{n-1}] \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix}. \tag{5}$$

Из (5) и леммы 1 следует единственность представления матрицы 2×2 вида

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \tag{6}$$

(конечно, такое представление существует не всегда).

Нам потребуется следующая лемма.

Лемма 2. *Пусть* $ad = b^2 + 1$, $a > b \geqslant d > 0$. *Тогда существует единственное натуральное* q *такое, что*

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} = \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \tag{7}$$

и либо $a' > b' \geqslant d' > 0$, *либо* $b' = 0$, $a' = d' = 1$.

Доказательство. Легко проверить, что из (7) следуют равенства $a' = d$, $b' = b - qd$, $d' = (a - qb) - q(b - qd)$:

$$\begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} d & b - qd \\ b - qd & (a - qb) - q(b - qd) \end{pmatrix} \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} = \\ = \begin{pmatrix} b & a - qb \\ d & b - qd \end{pmatrix} \begin{pmatrix} q & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} a & b \\ b & d \end{pmatrix}. \tag{8}$$

Поэтому условия $a' > b' > 0$ означают $d > b - qd > 0$. Таким образом, q должно быть частным от целочисленного деления b на d.

Рассмотрим вначале случай, когда b не делится на d: $b = qd + r$, $0 < r < d$. Поскольку $ad - b^2 = 1$,

$$\det \begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} = 1 \tag{9}$$

(эта матрица получается умножением матрицы с определителем 1 на две матрицы с определителем -1). Поэтому из $a' = d > 0$, $b' = r > 0$ следует, что и $d' > 0$. Кроме того, $a' = d > r = b'$. Осталось проверить неравенство $r = b' \geqslant d'$. Если $d' > b'$, то $a'd' \geqslant (b' + 1)^2 > b'^2 + 1$, что противоречит (9).

Теперь рассмотрим случай, когда b делится на d. Тогда

$$\begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix} = \begin{pmatrix} a' & 0 \\ 0 & d' \end{pmatrix}.$$

Определитель этой (диагональной) матрицы равен 1, поэтому $a' = d' = 1$.

Заметим, что лемму 2 можно применять индуктивно: полученная матрица $\begin{pmatrix} a' & b' \\ b' & d' \end{pmatrix}$ либо единичная, либо тоже удовлетворяет условию леммы. При последовательном применении леммы 2 внедиагональные элементы убывают, как видно из (8). Поэтому рано или поздно матрица станет

единичной. В результате такого процесса получается *симметричное разложение* исходной матрицы, имеющее вид

$$\begin{pmatrix} a & b \\ b & d \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \tag{10}$$

Теперь вернемся к обоснованию алгоритма Эрмита – Серре. Без ограничения общности можно считать, что $z < n/2$. Тогда $k = (z^2+1)/n \leqslant z$. Поэтому из леммы 2 следует существование симметричного разложения для матрицы

$$\begin{pmatrix} n & z \\ z & k \end{pmatrix} = \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix}. \tag{11}$$

Обозначим

$$A = \begin{pmatrix} x & u \\ y & v \end{pmatrix} = \begin{pmatrix} q_n & 1 \\ 1 & 0 \end{pmatrix} \cdot \ldots \cdot \begin{pmatrix} q_1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} q_0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Тогда

$$\begin{pmatrix} n & z \\ z & k \end{pmatrix} = A^t A = \begin{pmatrix} x & y \\ u & v \end{pmatrix} \begin{pmatrix} x & u \\ y & v \end{pmatrix}.$$

Значит, $n = x^2 + y^2$. Так как разложение вида (6) единственно, то из (5) заключаем, что $x = a_{n+1}$, $y = a_{n+2}$, где a_i — остатки в алгоритме Евклида, примененном к паре $a_0 = n$, $a_1 = z$. Ясно, что $a_{n+k}^2 \leqslant n$ при $k \geqslant 1$. Для завершения обоснования алгоритма Эрмита – Серре осталось проверить, что $a_i^2 > n$ при $i \leqslant n$. Для этого достаточно проверить, что $a_n = q_n a_{n+1} + a_{n+2} > \sqrt{n}$. И действительно,

$$a_n^2 = (q_n x + y)^2 \geqslant (x+y)^2 = n + 2xy > n.$$

Замечания.

1. У цепных дробей есть естественная геометрическая интерпретация (см., например, [1].) Интересно было бы выразить явно геометрический смысл леммы 2.
2. Из (11) можно вывести упоминавшийся выше факт, что разложение n/z в цепную дробь симметрично.
3. В книге Дэвенпорта [2] приведено несколько способов разложения числа в сумму двух квадратов. Все они так или иначе используют разложения в цепные дроби и неэффективны в смысле теории вычислительной сложности: время работы соответствующих алгоритмов экспоненциально зависит от длины входа (которая примерно равна логарифму числа, для которого находится представление в виде суммы двух квадратов).

4. В случае простого n имеет место замечательная формула Гаусса, доказанная Коши и Якобшталем. Пусть $n = 4k + 1$ — простое,

$$x \equiv \frac{1}{2}\binom{2k}{k} \pmod{n}, \quad y \equiv (2k)!x \pmod{n}, \quad |x|, |y| < n/2.$$

Тогда $x^2 + y^2 = n$. Несмотря на «явный» вид формулы Гаусса, она также не приводит к эффективному алгоритму.

К сожалению, в [2] не приводится доказательства формулы Гаусса, а лишь сообщается, что известные доказательства не очень просты. Возможно, кто-нибудь из читателей «Математического просвещения» придумает простое доказательство этого факта?

Автор благодарен Ю. А. Флёрову, заинтриговавшего его вопросом об алгоритмах разложения числа в сумму квадратов, а также И. Богданову, В. Бугаенко, В. Прасолову за интерес к статье и ценные замечания.

Список литературы

[1] Арнольд В. И. *Цепные дроби*. Биб-ка Матем. просв., вып. 14. М.: МЦНМО, 2001.

[2] Дэвенпорт Г. *Высшая арифметика*. М.: Наука, 1965.

[3] Butcher J. C. *MATHEMATICAL MINIATURE 14: Sums of two squares revisited*.
www.math.auckland.ac.nz/˜ butcher/miniature/miniature14.pdf

[4] van der Poorten A. J. *The Hermite–Serret Algorithm and* $12^2 + 33^2$ // Cryptography and Computational Number Theory. Springer Verlag, 2001. P. 129–136.

Вялый М. Н., ВЦ РАН, ВКМ НМУ
e-mail: vyalyi@mccme.ru