Г. ДЭВЕНПОРТ

ВЫСШАЯ АРИФМЕТИКА

ВВЕДЕНИЕ В ТЕОРИЮ ЧИСЕЛ

Перевод с английского
Б. З. МОРОЗА

под редакцией
Ю. В. ЛИННИКА

ИЗДАТЕЛЬСТВО «НАУКА»
ГЛАВНАЯ РЕДАКЦИЯ
ФИЗИКО-МАТЕМАТИЧЕСКОЙ ЛИТЕРАТУРЫ
МОСКВА 1965

THE HIGHER ARITHMETIC

An Introduction to the Theory of Numbers

H. DAVENPORT

ASTOR PROFESSOR OF MATHEMATICS
UNIVERSITY OF LONDON

HARPER TORCHBOOKS The Science Library

HARPER & BROTHERS, NEW YORK

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

Высшая арифметика, или теория чисел, изучает свойства натуральных чисел 1, 2, 3, ... Эти числа интересуют человека с давних времен. Античные летописи говорят о том, что уже тогда арифметику знали глубже и шире, чем это было необходимо для нужд повседневной жизни. Но систематической, самостоятельной наукой высшая арифметика становится лишь в новое время, начиная с открытий Ферма (Fermat, 1601—1665).

Многие простые и общие теоремы высшей арифметики естественно возникают из вычислений, однако при доказательстве этих теорем часто встречаются очень большие трудности. «Эта особенность, — по словам Гаусса, — вместе с неистощимым богатством высшей арифметики, которым она столь сильно превосходит другие области математики, придает высшей арифметике неотразимое очарование, сделавшее ее любимой наукой величайших математиков».

Теория чисел считается обычно «чистейшей» ветвью чистой математики. Она имеет очень немного прямых приложений к другим естественным наукам, но обладает одной общей с ними чертой: теория чисел развивается из *эксперимента*, роль которого играет проверка общих теорем на численных примерах. Такой эксперимент необходим в любой области математики, но в теории чисел он играет большую роль, чем где бы то ни было, ибо в других областях математики результаты, полученные таким способом, часто бывают неверными.

Автор этой книги хорошо понимает, что нематематик не сможет прочесть ее без труда. Трудность частично лежит в самом предмете. Этой трудности не избежать, пытаясь использовать несовершенные аналогии или проводя доказательства, выражающие основную мысль, но неточные в деталях. Такая попытка может лишь уменьшить интерес к этой наиболее точной из наук.

В этой книге теоремы и их доказательства часто иллюстрируются численными примерами. Примеры обычно очень просты и могут не удовлетворить читателя, который любит вычисления. Задача этих примеров — пояснить общую теорию. Вопрос о наиболее эффективном проведении арифметических вычислений выходит за рамки данной книги.

Автор признателен многим друзьям, особенно д-ру Эрдешу, проф. Морделлу и д-ру Роджерсу, за предложения и исправления. Он обязан также капитану Дрэму за разрешение включить описание его алгоритма.

ГЛАВА I

РАЗЛОЖЕНИЕ НА МНОЖИТЕЛИ И ПРОСТЫЕ ЧИСЛА

1. Законы арифметики. Высшая арифметика исследует общие предложения о натуральных числах 1, 2, 3, ... обычной арифметики. Примерами таких предложений могут служить фундаментальная теорема (I, 4)*) о том, что *каждое натуральное число разлагается на простые множители и это разложение единственно,* и теорема Лагранжа (V, 4) о том, что *любое натуральное число представимо в виде суммы не более четырех точных квадратов.* С арифметическими вычислениями мы встретимся только в иллюстративных примерах; мы не касаемся также числовых курьезов, не связанных с общей теорией.

В раннем детстве, играя в четки или в шарики, мы экспериментально изучаем арифметику. Сначала, объединяя два ряда предметов в один, мы учимся складывать; затем, многократно повторяя сложение, учимся умножать. Постепенно мы узнаем, как следует обращаться с числами, и хорошо знакомимся с законами арифметики — законами, которые, вероятно, наиболее близки нам из всех достояний человеческого знания.

Высшая арифметика — дедуктивная наука, основанная на законах арифметики. Мы все знаем эти законы, хотя, может быть, не видели их формулировки в общих терминах. Законы арифметики выражаются следующим образом.

Сложение. Любые два натуральных числа a и b имеют *сумму*, обозначаемую $a+b$, которая сама является натуральным числом. Операция сложения удовлетворяет двум законам:

$$a + b = b + a \qquad (\textit{коммутативный закон сложения}),$$
$$a + (b + c) = (a + b) + c \qquad (\textit{ассоциативный закон сложения}),$$

*) Ссылки такого рода относятся к главам и пунктам глав этой книги.

скобки в последней формуле указывают порядок выполнения операций.

Умножение. Любые два натуральных числа a и b имеют *произведение*, обозначаемое $a \cdot b$ или ab, которое само является натуральным числом. Операция умножения удовлетворяет двум законам:

$$ab = ba \qquad (\textit{коммутативный закон умножения}),$$
$$a(bc) = (ab)c \qquad (\textit{ассоциативный закон умножения}).$$

Имеется также закон, связывающий сложение и умножение:

$$a(b + c) = ab + ac \quad (\textit{дистрибутивный закон}).$$

Порядок. Если a и b — два натуральных числа, то или a равно b, или a *меньше* b, или b *меньше* a, и из этих трех возможностей осуществляется ровно одна. Утверждение «a меньше b» символически выражается в виде $a < b$, в этом случае мы говорим также, что b больше a, символически: $b > a$. Основной закон, управляющий этим отношением порядка, таков:

$$\textit{если} \quad a < b \quad \textit{и} \quad b < c, \quad \textit{то} \quad a < c.$$

Имеются также два закона, связывающих отношение порядка с операциями сложения и умножения:

$$\textit{если} \quad a < b, \quad \textit{то} \quad a + c < b + c \quad \textit{и} \quad ac < bc,$$

каково бы ни было натуральное число c.

Сокращение. Два закона сокращения логически вытекают из законов порядка; однако они достаточно важны, и мы их точно сформулируем. Первый закон гласит:

$$\textit{если} \quad a + x = a + y, \quad \textit{то} \quad x = y.$$

Это следует из того, что если $x < y$, то $a + x < a + y$, что противоречит предположению; невозможно также неравенство $y < x$; поэтому $x = y$. Тем же способом получаем и второй закон сокращения, утверждающий, что

$$\textit{если} \quad ax = ay, \quad \textit{то} \quad x = y.$$

Вычитание. Вычесть число b из числа a — значит найти, если это возможно, такое число x, что $b + x = a$. Возможность вычитания связана с отношением порядка следующим законом: *b можно вычесть из a тогда и только тогда, когда b меньше*

a. Из первого закона сокращения следует, что если вычитание возможно, то результат единственен; действительно, если $b + x = a$ и $b + y = a$, то $x = y$. Результат вычитания b из a обозначается $a - b$. Правила действий со знаком минус, например $a - (b - c) = a - b + c$, вытекают из определения вычитания и коммутативного и ассоциативного законов сложения.

Деление. Разделить число a на число b — значит найти, если это возможно, такое число x, что $bx = a$. Если такое число существует, то оно обозначается $\frac{a}{b}$ или a/b. Из второго закона сокращения следует, что если деление возможно, то результат единственен.

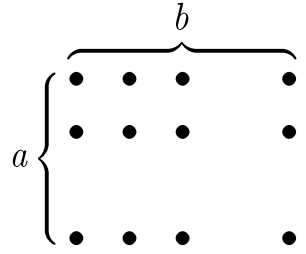

Рис. 1.

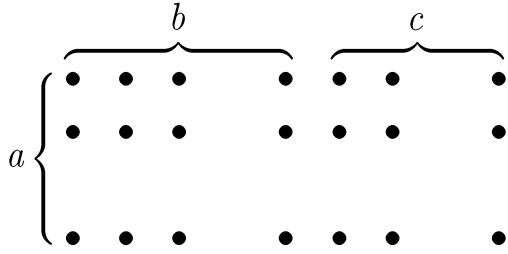

Рис. 2.

Все вышеупомянутые законы довольно очевидны, если сложение и умножение понимать как действия над совокупностями некоторых предметов. Например, коммутативный закон умножения становится очевидным, если рассмотреть прямоугольную таблицу (рис. 1), в которой предметы расположены в b столбцов и a строк; число предметов в ней равно ab или ba. Дистрибутивный закон очевиден, если рассматривать совокупность предметов на рис. 2; в этой совокупности имеется $a(b + c)$ предметов, их число складывается из ab и ac предметов. Несколько менее очевидным, возможно, является ассоциативный закон умножения, утверждающий, что $a(bc) = (ab)c$. Чтобы сделать ясным и этот закон, рассмотрим прямоугольник, изображенный на рис. 1, заменив в нем каждый предмет числом c. Тогда сумма всех чисел в каждой строке равна bc; так как имеется a строк, то полная сумма равна $a(bc)$. С другой стороны, имеется ab чисел, каждое из которых равно c, поэтому полная сумма есть $(ab)c$. Значит, $a(bc) = (ab)c$, что и требуется доказать.

Законы арифметики вместе с принципом индукции (который мы обсудим далее) образуют основу для логического развития теории чисел. Они дают возможность доказывать общие теоремы о натуральных числах, не возвращаясь к исходным значениям чисел и операций над ними. Правда, некоторые довольно глубокие результаты теории чисел проще всего получить, подсчитав определенное число предметов двумя различными способами, но таких результатов не очень много.

Хотя законы арифметики и образуют логическую основу теории чисел (на самом деле они являются основой и для большей части математики), было бы неудобным возвращаться к ним на каждом этапе доказательства, поэтому мы будем предполагать, что читатель уже обладает некоторыми знаниями в области элементарной математики. Мы детально обсудили эти законы, чтобы показать, где предмет действительно начинается.

Закончим этот пункт кратким рассмотрением соотношений между системой натуральных чисел и двумя другими числовыми системами, важными как в высшей арифметике, так и в математике вообще: *системой всех целых и системой всех рациональных чисел.*

В системе натуральных чисел всегда выполнимы действия сложения и умножения, но не всегда выполнимы вычитание и деление. Чтобы сделать вычитание всегда возможным, в математике вводятся число 0 и отрицательные целые -1, -2, ... Вместе с натуральными числами они образуют систему всех целых чисел: ..., -2, -1, 0, 1, 2, ..., в которой вычитание всегда выполнимо и результат вычитания определен единственным образом. Элементарная алгебра учит, как в этой расширенной числовой системе определить умножение («правило знаков»), чтобы законы арифметики, управляющие сложением и умножением натуральных чисел, оставались в силе. Отношение порядка распространяется на все целые числа так, что управляющие им в системе натуральных чисел законы сохраняются и здесь, кроме одного исключения: из $a < b$ вытекает $ac < bc$ только в случае, если c положительно. Расширение системы натуральных чисел приводит к исключению и во втором законе сокращения,

остающемся верным лишь при условии, что сокращаемый множитель отличен от нуля:

$$\textit{если} \quad ax = ay, \quad \textit{то} \quad x = y \quad \textit{при} \quad a \neq 0.$$

Таким образом, целые числа (положительные, отрицательные и нуль) удовлетворяют тем же арифметическим законам, что и натуральные числа, но теперь вычитание уже всегда выполнимо; кроме того, закон порядка и второй закон сокращения должны быть изменены только что указанным способом. Натуральные числа можно теперь описать как *целые положительные числа.*

Вернемся к натуральным числам. Как мы уже знаем, в системе натуральных чисел деление выполнимо не всегда. Если натуральное число b можно разделить на натуральное число a в системе натуральных чисел, то говорят, что a является *множителем* или *делителем* b или что b является *кратным* a. В качестве иллюстрации к этому определению отметим, что 1 является множителем каждого числа и что a является множителем a (отношение равно 1). В качестве другого примера заметим, что числа, делящиеся на 2, называются четными: 2, 4, 6, ..., а числа, не делящиеся на 2, — нечетными: 1, 3, 5, ...

Отношение делимости изучается в теории чисел и в некоторых других, близких к теории чисел областях математики. В этой главе будут рассмотрены некоторые непосредственные следствия определения делимости. Отметим, прежде всего, несколько очевидных фактов.

I. *Если a делит b, то $a \leqslant b$* (т. е. a или меньше, или равно b). Действительно, если $b = ax$, то $b - a = a(x - 1)$, где $x - 1$ либо равно 0, либо является натуральным числом.

II. *Если a делит b и b делит c, то a делит c.* В самом деле, при $b = ax$ и $c = by$ имеем $c = a(xy)$, где x и y — натуральные числа.

III. *Если каждое из чисел b и c делится на a, то $b + c$ и $b - c$ (если $c < b$) также делятся на a.* Действительно, при $b = ax$ и $c = ay$ будет $b + c = a(x + y)$ и $b - c = a(x - y)$.

Ограничение $b > c$ при рассмотрении $b - c$ в последнем утверждении становится излишним, если очевидным образом распространить отношение делимости на все целые числа: целое число

b делится на натуральное a, если отношение $\frac{b}{a}$ является целым числом. Таким образом, отрицательное число $-b$ делится на a тогда и только тогда, когда b делится на a. Заметим, что 0 делится на любое натуральное число (отношением служит целое число 0).

IV. *Если целые числа b и c делятся на натуральное число a, то и каждое число, представимое в виде $ub + vc$, где u и v — целые числа, делится на a.* При $b = ax$ и $c = ay$ получим $ub + vc = (ux + vy)a$. Это свойство включает результаты, сформулированные в III как частные случаи: полагая u и v равными 1, получаем $b + c = ub + vc$; полагая u равным 1, а v равным -1, получаем $ub + vc = b - c$.

Так же как с введением 0 и отрицательных чисел оказывается всегда возможным вычитание, расширение системы натуральных чисел путем введения положительных дробей, т. е. всевозможных дробей вида a/b, где a и b — натуральные числа, делает возможным деление на любое число. Комбинируя эти методы, получим *систему рациональных чисел*, которая состоит из всех целых чисел и всех дробей, как положительных, так и отрицательных. В этой числовой системе выполнимы все четыре арифметических действия — сложение, умножение, вычитание и деление, за исключением деления на нуль.

Основной предмет теории чисел — натуральные числа. Но часто бывает удобно работать в системе всех целых чисел или в системе всех рациональных чисел. Важно, конечно, чтобы читатель в каждом конкретном случае понимал, какие числа обозначаются теми или иными символами.

2. Доказательство по индукции. Большинство предложений теории чисел являются утверждениями о произвольном натуральном числе; например, теорема Лагранжа говорит о том, что каждое натуральное число есть сумма не более четырех квадратов. Как же доказать, что некоторое утверждение истинно для *любого натурального числа*? Конечно, некоторые утверждения непосредственно вытекают из законов арифметики; таковы, например, алгебраические тождества типа

$$(n + 1)^2 = n^2 + 2n + 1.$$

Но более интересные и более арифметические по своей природе утверждения не столь просты.

Ясно, что мы никогда не докажем общего предложения, последовательно убеждаясь в его истинности для 1, 2, 3 и так далее, ибо нельзя перебрать бесконечного числа возможностей. Установив, что утверждение верно для любого числа, меньшего миллиона или миллиона миллионов, мы не приблизимся к доказательству того, что оно верно всегда. (Иногда, правда, бывает, что некоторые предложения теории чисел, установленные путем вычислений для большого числа частных случаев, оказываются верными в довольно широкой области.)

Может быть, однако, мы найдем *общее доказательство* того, что *если* наше предложение верно для каждого из чисел $1, 2, \ldots, n-1$, *то* оно верно и для следующего натурального числа n.

Если это доказано, то из истинности нашего предложения для числа 1 следует, что оно верно для числа 2, из того, что оно верно для 1 и 2, вытекает, что оно верно для 3, и так далее до бесконечности. Это предложение будет поэтому верным для любого числа, если оно верно для числа 1.

В этом и состоит принцип доказательства по индукции, который относится к предложениям о том, что что-то верно для любого натурального числа. Чтобы применить этот принцип, необходимо доказать две вещи: во-первых, нужно доказать, что предложение верно для 1, а во-вторых, что *если* предложение верно для каждого из чисел $1, 2, \ldots, n-1$, меньших n, то оно верно и для числа n. Установив эти факты, мы заключаем, что доказываемое предложение верно для любого натурального числа.

Простой пример проиллюстрирует этот принцип. Будем изучать отрезки ряда $1 + 3 + 5 + \ldots$ последовательных нечетных чисел. Легко заметить, что

$$1 = 1^2, \quad 1 + 3 = 2^2, \quad 1 + 3 + 5 = 3^2, \quad 1 + 3 + 5 + 7 = 4^2$$

и т. д.

Этим подсказывается общее утверждение: *при любом натуральном n сумма первых n нечетных чисел равна n^2*. Докажем это общее предложение по индукции. Оно, конечно, верно, если

n равно 1. Мы должны доказать, что результат верен для любого натурального числа n; в силу принципа индукции можно предполагать, что предложение уже доказано для всех натуральных чисел, меньших n. Пусть, в частности, нам уже известно, что сумма первых $n-1$ нечетных чисел равна $(n-1)^2$. Сумма первых n нечетных чисел получится из нее добавлением n-го нечетного числа, равного $2n-1$. Таким образом, сумма первых n нечетных чисел есть

$$(n-1)^2+2n-1,$$

что равно n^2. Этим требуемое утверждение доказано.

Доказательства с помощью индукции иногда озадачивают неискушенных: «вы предсказываете, какое предложение следует доказывать». Предложения рассматриваемого типа состоят из бесконечного числа частных случаев, каждый из которых соответствует определенному натуральному числу 1, 2, 3, ...; принцип индукции лишь дает возможность предполагать при доказательстве одного из этих случаев, что предшествующие случаи уже рассмотрены.

Изложение доказательства по индукции в безупречной форме требует некоторого внимания. В приведенном примере доказывалось, что *сумма первых n нечетных чисел равна n^2*. Здесь n — любое натуральное число, и, конечно, утверждение не изменится, если всюду, где встречается n, употреблять какой-нибудь другой символ. Когда мы приступаем к доказательству, n есть вполне определенное число и имеется опасность употребить один и тот же символ в разных смыслах или даже высказать бессмыслицу типа «предложение верно при n, равном $n-1$». Чтобы избежать этого, нужно использовать в случае необходимости разные символы.

С общелогической точки зрения нет ничего более очевидного, чем законность доказательства по индукции. Тем не менее можно спорить, является ли этот принцип по своей природе *определением* или *аксиомой* или это *логический принцип*. Но, так или иначе, ясно, что принцип индукции служит для того, чтобы расположить натуральные числа в определенном порядке: установив, что вначале идут числа 1, 2, ..., $n-1$, мы объявляем следующим числом число n. Таким образом, этот

принцип объясняет, что означают на самом деле слова «и так далее», встречающиеся при попытке перечислить все натуральные числа.

3. Простые числа. Очевидно, что каждое натуральное число a делится на 1 (отношение равно a) и на a (отношение равно 1). Множитель a, отличный от 1 или a, называется *собственным* множителем. Известно, что существуют числа, не имеющие собственных множителей; они называются простыми числами или *простыми.* Первые несколько простых таковы:

$$2,\ 3,\ 5,\ 7,\ 11,\ 13,\ 17,\ 19,\ 23,\ 29,\ 31,\ \ldots$$

Считать ли простым числом 1 — вопрос соглашения, но удобнее (как мы увидим позднее) 1 не считать простым.

Число, не являющееся простым и не равное 1, называется *составным*; такое число можно представить в виде произведения двух чисел, каждое из которых больше 1. Хорошо известно, что любое составное число можно представить в виде произведения простых; при этом, конечно, некоторые простые могут встретиться по нескольку раз. Возьмем, например, число 666; ясно, что оно делится на 2, и мы получаем $666 = 2 \cdot 333$. Далее, 333 имеет очевидный множитель 3, откуда $333 = 3 \cdot 111$. Множитель 111 снова делится на 3, так что $111 = 3 \cdot 37$. Следовательно,

$$666 = 2 \cdot 3 \cdot 3 \cdot 37,$$

и мы получили представление составного числа 666 в виде произведения простых. Имеется общая теорема о том, что каждое число представимо в виде произведения простых, или, что то же самое, *любое число, большее* 1, *является простым либо разлагается в произведение простых.*

Для доказательства этого общего утверждения воспользуемся методом индукции. Докажем утверждение для числа n, считая, что оно уже доказано для любого числа, меньшего n. Если n простое, доказывать нечего. Если n составное, то его можно представить в виде произведения ab, где a и b больше 1 и меньше n. Мы уже знаем, что числа a и b или являются простыми, или разлагаются в произведение простых; подставив их разложения в равенство $n = ab$, получаем разложение n на простые множители. Это доказательство столь просто, что может

показаться читателю совершенно излишним. Другая общая теорема о разложении на простые будет доказываться уже не так просто.

Ряд простых 2, 3, 5, 7, ... издавна интересовал людей. В дальнейшем мы сформулируем некоторые из полученных в этом направлении результатов. В настоящий момент мы ограничимся доказательством того, что *ряд простых бесконечен.* Это доказательство Евклида (книга IX, предложение 20) может служить образцом изящества и простоты. Пусть 2, 3, 5, ..., P — ряд простых до некоторого простого P. Рассмотрим произведение всех этих простых, добавим к нему 1 и положим

$$N = 2 \cdot 3 \cdot 5 \ldots P + 1.$$

Это число не может делиться на 2, так как если бы оно делилось на 2, то и $2 \cdot 3 \cdot 5 \ldots P$, а потому и разность $N - 2 \cdot 3 \cdot 5 \ldots P$ делились бы на 2. Но разность этих чисел равна 1 и на 2 не делится. Аналогично убеждаемся в том, что N не может делиться на 3, на 5 и вообще ни на какое другое простое вплоть до P. С другой стороны, N должно делиться на *какое-нибудь* простое (на само себя, если N простое, или на любой простой делитель N, если N составное). Следовательно, существует простое число, отличное от любого из простых 2, 3, 5, ..., P и потому большее P. Таким образом, ряд простых оборваться не может.

4. Основная теорема арифметики. В предыдущем пункте было доказано, что любое составное число представимо в виде произведения простых. В качестве примера мы разложили на множители число 666: $666 = 2 \cdot 3 \cdot 3 \cdot 37$. Обратимся теперь к другому вопросу, также имеющему первостепенное значение. Можно ли осуществить разложение на простые более чем одним способом? (Понятно, конечно, что два представления, отличающиеся только порядком множителей, должны рассматриваться как одно, например разложение $3 \cdot 2 \cdot 37 \cdot 3$ отождествляется с разложением $2 \cdot 3 \cdot 3 \cdot 37$.) Может ли, например, число 666 иметь какое-нибудь другое представление в виде произведения простых? Читатель, не знающий теории чисел, вероятно, будет все же уверен, что другого такого представления нет, однако найти слишком простое общее доказательство ему не удастся.

Было бы удобно высказать это предложение в форме, применимой ко всем натуральным, а не только к составным числам. Будем рассматривать простые числа как «произведения» простых, состоящие только из одного сомножителя. Можно сделать еще один шаг и рассматривать число 1 как пустое произведение простых, считая по определению, что величина пустого произведения равна 1. Это соглашение полезно не только здесь, но и в других частях математики, так как оно дает возможность включать в общие теоремы частные случаи, которые иначе пришлось бы исключить или рассматривать особо.

При этих соглашениях общее предложение таково: *каждое натуральное число может быть представлено в виде произведения простых одним и только одним способом.* Эту теорему называют *основной теоремой арифметики*; она имеет довольно странную историю. В «*Элементах*» Евклида она еще не встречается, но некоторые арифметические предложения в книге VII уже эквивалентны ей. Точно она не формулируется даже во «*Введении в теорию чисел*», написанном в 1798 году Лежандром. Первую точную формулировку теоремы и ее доказательство дал Гаусс в 1801 году в своих знаменитых «*Арифметических исследованиях*». Вероятно, из-за отсутствия этой теоремы у Евклида она принимается без доказательства во многих школьных учебниках. В одном из учебников ее даже объявляют «законом мышления», каковым она, конечно, не является. Приведем здесь прямое доказательство единственности разложения на множители. Позднее (в п. 7) будет дано другое доказательство, полностью независимое от рассматриваемого в этом пункте.

Заметим прежде всего, что если разложение числа m на простые множители единственно, то каждый простой множитель m должен входить в это разложение. Действительно, пусть p — какое-нибудь простое, делящее m, тогда $m = pm'$, где m' — некоторое целое число; разложение m можно получить из разложения m', добавив простой множитель p. Так как по предположению имеется только одно разложение числа m на простые, то p должно встретиться в нем.

Будем доказывать единственность разложения по индукции, именно: докажем единственность разложения числа n в предпо-

ложении, что единственность разложения для всех чисел, меньших n, уже установлена. Если n простое, то доказывать нечего. Предположим, что n составное и имеются два различных представления n в виде произведения простых, скажем,

$$n = pqr \cdots = p'q'r' \ldots,$$

где p, q, r, ... и p', q', r', ... — простые. Одно и то же простое не может встретиться в двух разложениях, так как в этом случае мы сократили бы на это простое и получили бы два различных разложения меньшего числа, а это противоречит индуктивному предположению.

Не нарушая общности, можно предполагать, что p — наименьшее из простых, встречающихся в первом разложении. Так как n составное, имеется по меньшей мере один множитель в разложении, помимо p; поэтому $n \geqslant p^2$. Аналогично $n \geqslant p'^2$. Так как p и p' не одинаковы, то по крайней мере одно из этих неравенств строгое и, следовательно, $pp' < n$. Рассмотрим теперь число $n - pp'$. Это натуральное число меньше n, следовательно, оно может быть представлено как произведение простых одним и только одним способом. Так как p делит n, оно делит также $n - pp'$, поэтому, согласно сделанному ранее замечанию, p должно входить в разложение $n - pp'$. Аналогично убеждаемся, что в это разложение должно входить и p'. Следовательно, разложение $n - pp'$ на простые имеет вид

$$n - pp' = pp'QR \ldots,$$

где Q, R, ... — простые. Отсюда следует, что число pp' делит n. Но $n = pqr \ldots$, поэтому (после сокращения на p) получается, что p' делит $qr \ldots$. Ввиду предварительного замечания это невозможно, ибо $qr \ldots$ — число, меньшее n, и p' не является одним из простых q, r, ..., входящих в его разложение. Это противоречие доказывает, что n обладает только одним разложением на простые множители.

Читатель согласится, что это доказательство, не будучи очень длинным или трудным, является все же довольно тонким. То же верно и для других прямых доказательств единственности разложения, основанных обычно примерно на тех же идеях, что и приведенное доказательство. Важно заметить, что, в то время как *возможность* разложения на простые непосредственно

следует из определения простого, доказательство *единственности* разложения получается не сразу. Следующий пример, указанный Гильбертом, объясняет, почему эти два предложения так отличаются друг от друга. Определения множителей и простых используют лишь операцию умножения и не зависят от операции сложения. Посмотрим, что произойдет, если применить эти определения к системе чисел, замкнутой относительно умножения, но не замкнутой относительно сложения и вычитания. Рассмотрим систему всевозможных чисел вида $4x + 1$:

$$1, 5, 9, 13, 17, 21, 25, 29, \ldots$$

Произведение двух таких чисел снова является числом того же вида. Определим «псевдопростое» как число этой системы (отличное от 1), которое не разлагается собственно *в этой системе*. Числа 5, 9, 13, 17, 21 являются псевдопростыми; а 25 есть первое не псевдопростое число. Каждое число этой системы либо само псевдопростое, либо может быть разложено на псевдопростые множители, — это доказывается так же, как и раньше. Но в этой системе разложение числа на псевдопростые не является однозначным; например, число 693 имеет два разложения:

$$693 = 9 \cdot 77 = 21 \cdot 33,$$

где все числа 9, 77, 21, 33 — псевдопростые. Конечно, ясно, что эти числа можно разложить на множители вне рассматриваемой системы; точка зрения примера проливает свет на логическую структуру любого доказательства единственности разложения на множители. Такое доказательство не может опираться лишь на определение простого и свойства мультипликативных операций. Оно должно где-то использовать сложение или вычитание, ибо иначе его можно было бы применить к только что рассмотренной системе чисел. В приведенном доказательстве фундаментальной теоремы вычитание использовалось при образовании числа $n - pp'$.

Основная теорема арифметики вскрывает структуру натуральных чисел по отношению к операции умножения. Эта теорема показывает, что все натуральные числа получаются из простых с помощью всевозможных умножений, причем в результате различных умножений получаются различные числа. Теперь понятно, что число 1 неудобно считать простым, ибо это

нарушило бы единственность разложения на простые множители: к любому произведению можно присоединить множителем 1, не изменив значения этого произведения.

5. Следствия из основной теоремы. Основная теорема арифметики, доказанная в предыдущем пункте, устанавливает, что любое натуральное число может быть представлено как произведение простых одним и только одним способом, если в качестве разложения на простые простого числа брать само это число и считать пустое произведение равным 1.

Если известно разложение числа на простые, то сразу же можно ответить на различные вопросы, связанные с этим числом. Прежде всего можно указать все делители этого числа. Посмотрим сначала, как это сделать в одном частном случае. Возьмем тот же численный пример, что и раньше:

$$666 = 2 \cdot 3 \cdot 3 \cdot 37.$$

Делителем числа 666 является такое число d, для которого $666 = dd'$, где d' — другое натуральное число. По основной теореме арифметики разложения d и d' на множители должны давать в произведении

$$2 \cdot 3 \cdot 3 \cdot 37.$$

Поэтому d должно быть произведением некоторых из простых 2, 3, 3, 37, а d' — произведением остальных. (Сделанное ранее соглашение о том, что пустое произведение равно 1, полезно и здесь, ибо дает возможность включить в формулировку крайние случаи: d равно 1 или d' равно 1.) Выбирая простые всеми возможными способами, получаем все делители 666, именно:

$$1,\ 2,\ 3,\ 37, 2 \cdot 3,\ 2 \cdot 37,\ 3 \cdot 3,\ 3 \cdot 37,\ 2 \cdot 3 \cdot 3,$$
$$2 \cdot 3 \cdot 37,\ 3 \cdot 3 \cdot 37,\ 2 \cdot 3 \cdot 3 \cdot 37.$$

Эта ситуация является совершенно общей, и остается лишь выбрать обозначения, в которых ее проще всего описать. Пусть n — произвольное натуральное число, большее 1, и пусть p, q, r, ... — различные простые в его разложении. Предположим, что p встречается a раз в разложении n, q встречается b раз и так далее. Тогда

$$n = p^a q^b \ldots \qquad (1)$$

В этом случае всевозможные делители n суть произведения вида $p^\alpha q^\beta \ldots$, в которых показатель a принимает значения 0, 1, ..., a, показатель β — значения 0, 1, ..., b, и так далее *). Это доказывается так же, как и в предыдущем примере; доказательство основано на фундаментальной теореме арифметики. В примере с $n = 666$ имеются три различных множителя 2, 3, 37; их показатели равны соответственно 1, 2, 1. Поэтому все делители числа 666 задаются формулой

$$2^\alpha 3^\beta 37^\gamma,$$

в которой α равно 0 или 1, β равно 0, 1 или 2, γ равно 0 или 1. Выписав их, получим указанные ранее делители 666.

Чтобы найти количество всех делителей числа n, достаточно подсчитать число всевозможных значений показателей α, β, γ, ... В общем случае, когда n представлено в виде (1), показатель α равен одному из чисел 0, 1, ..., a и, значит, для него имеется $a + 1$ различных возможностей. Аналогично для β имеется $b + 1$ различных возможностей, и так далее. Выборы различных показателей α, β, ... не зависят один от другого, и разным наборам значений для α, β, ... соответствуют различные делители n в силу единственности разложения на простые. Следовательно, полное число делителей n равно

$$(a + 1)(b + 1) \ldots$$

Обычно полное число делителей числа n (включая 1 и n, как мы это и делали) обозначают через $d(n)$ **). Используя это обозначение, можно сказать, что если $n = p^a q^b \ldots$, где p, q, ... — различные простые, то

$$d(n) = (a + 1)(b + 1) \ldots$$

В рассмотренном примере показателями служат числа 1, 2, 1 и число делителей равно

$$2 \cdot 3 \cdot 2 = 12.$$

Можно рассмотреть также сумму делителей n, включая 1 и n. Эта сумма обозначается обычно через $\sigma(n)$. Если разложение n на простые написано в виде (1), то $\sigma(n)$ получается по фор-

*) Мы считаем, как обычно, что нулевая степень числа равна 1.

**) Принято также обозначение $\tau(n)$. (*Прим. перев.*)

муле

$$\sigma(n) = \{1 + p + \cdots + p^a\}\{1 + q + \cdots + q^b\} \ldots$$

Действительно, после раскрытия скобок это выражение представляется суммой всевозможных произведений вида $p^\alpha q^\beta \ldots$, где α принимает каждое из значений $0, 1, \ldots, a$, β — значения $0, 1, \ldots, b$ и так далее. Эти произведения образуют все делители n. В прежнем численном примере будет

$$\sigma(666) = (1 + 2)(1 + 3 + 3^2)(1 + 37) = 3 \cdot 13 \cdot 38 = 1482;$$

это также можно подсчитать, просто выписав все делители и сложив их. Арифметические функции $d(n)$ и $\sigma(n)$, а также функция $\varphi(n)$, с которой мы встретимся позднее, табулированы вплоть до $n = 10\,000$ (см. Number-divisor Tables, vol. VIII, British Assoc. Math. Tables, Cambridge, 1940).

Древние греки уделяли большое внимание *совершенным* числам, которые они определяли как числа n, обладающие тем свойством, что сумма делителей n, включая 1, но исключая n, равна самому n. Простейший пример — число 6:

$$1 + 2 + 3 = 6.$$

Другой способ определения — потребовать, чтобы выполнялось равенство

$$\sigma(n) = 2n,$$

ибо $\sigma(n)$ есть сумма всех делителей, включая само n. Евклидом (книга IX, предложение 36) было доказано, что если p — простое число, а число $p+1$ является степенью 2, скажем, $p+1 = 2^k$, то число $2^{k-1}p$ — совершенное. Действительно, из ранее выведенной формулы для $\sigma(n)$ находим

$$\sigma(2^{k-1}p) = \{1 + 2 + \cdots + 2^{k-1}\}\{1 + p\}.$$

Но

$$1 + 2 + \cdots + 2^{k-1} = 2^k - 1 = p, \quad 1 + p = 2^k,$$

откуда $\sigma(n) = 2n$ при $n = 2^{k-1}p$. Эйлер в работе, опубликованной после его смерти, дополнил результат Евклида, доказав, что *каждое* четное совершенное число имеет форму Евклида. Неизвестно, существуют ли какие-нибудь нечетные совершенные числа; неизвестно также, бесконечно ли количество четных

совершенных чисел. Первые пять четных совершенных чисел таковы:

$$6,\ 28,\ 496,\ 8128,\ 33\,550\,336.$$

До сих пор мы рассматривали делители одного числа. Но можно также исследовать общие делители двух или более чисел. Любой общий делитель двух чисел m и n должен состоять в точности из тех простых, которые входят и в m, и в n. Если таких простых нет, то m и n не имеют общего делителя, отличного от 1, и называются *взаимно простыми*. Например, числа

$$2829 = 3 \cdot 23 \cdot 41 \quad \text{и} \quad 6850 = 2 \cdot 5^2 \cdot 137$$

взаимно просты.

Если m и n имеют общие множители, то их *наибольший общий делитель* (Н. О. Д.) получится, если мы перемножим различные общие простые множители m и n, взяв каждый из них в наибольшей степени, делящей и m, и n.

Например, Н. О. Д. чисел

$$3132 = 2^2 \cdot 3^3 \cdot 29 \quad \text{и} \quad 7200 = 2^5 \cdot 3^2 \cdot 5^2$$

равен $2^2 \cdot 3^2$ или 36. Ясно, что показатель каждого простого в Н. О. Д. равен меньшему из показателей, с которыми это простое входит в числа m и n.

Очевидно также, что общие делители m иn — это в точности все делители их Н. О. Д. Формулируя все эти утверждения, мы, конечно, пользуемся основной теоремой арифметики.

Аналогичная ситуация возникает и при рассмотрении *общих кратных* данных двух чисел. Среди всевозможных общих кратных имеется *наименьшее общее кратное* (Н. О. К.); оно равно произведению простых, входящих хотя бы в одно из чисел m и n и взятых с наибольшим из двух показателей, с которыми они входят в эти числа. Например, для двух записанных выше чисел (3132 и 7200) Н. О. К. равно

$$2^5 \cdot 3^3 \cdot 5^2 \cdot 29.$$

Общие кратные двух данных чисел суть всевозможные кратные их Н. О. К.

Эти рассмотрения легко переносятся и на случай более чем двух чисел. Важно отметить, что здесь имеются два вида возможной взаимной простоты. Говорят, что некоторые числа яв-

ляются *взаимно простыми*, если не существует числа, большего 1, делящего каждое из них; эти числа называются *попарно взаимно простыми*, если никакие два из них не имеют общего множителя, большего 1. Чтобы имел место первый случай, достаточно, чтобы не было простого, делящего все числа; а во втором случае нужно, чтобы ни одно простое не делило сразу два из этих чисел.

Имеется ряд простых теорем, которые кажутся очевидными, но в действительности очевидны лишь благодаря единственности разложения на простые. Например, *если число делит произведение двух чисел и взаимно просто с одним из них, то оно должно делить другое.* Действительно, если a делит bc и взаимно просто с b, то разложение на простые a входит в разложение на простые bc, но не имеет общих множителей с b и поэтому содержится в разложении c.

6. Алгоритм Евклида. В предложении 2 книги VII Евклид дал систематический способ, или *алгоритм*, для нахождения наибольшего общего делителя двух данных чисел. Этот алгоритм приводит к новому — по сравнению с изложенным в двух предыдущих пунктах — подходу к вопросам делимости. Поэтому мы начинаем изложение заново, не предполагая известным ничего, кроме определения отношения делимости.

Пусть a и b — два данных натуральных числа; предположим, что $a > b$. Мы попытаемся исследовать общие делители a и b. Если a делится на b, то общие делители a и b просто совпадают с делителями b и этим полностью характеризуются. Если a не делится на b, то a можно представить в виде суммы некоторого кратного b и остатка, меньшего чем b:

$$a = qb + c, \quad \text{где} \quad c < b. \tag{2}$$

Процесс «деления с остатком» выражает тот факт, что всякое a, не кратное b, должно встретиться где-то между двумя последовательными кратными b. Если a лежит между qb и $(q+1)b$, то

$$a = qb + c, \quad \text{где} \quad 0 < c < b.$$

Из равенства (2) следует, что любой общий делитель b и c является также делителем a. Кроме того, любой общий делитель a и b делит c, так как $c = a - qb$.

Следовательно, общие делители a и b, если такие найдутся, совпадают с общими делителями b и c. Задача нахождения общих делителей a и b сводится поэтому к такой же задаче для чисел b и c, соответственно меньших, чем a и b.

Суть алгоритма состоит в повторении этого рассуждения. Если b делится на c, то общие делители b и c состоят просто из всех делителей c. Если нет, то представим b в виде

$$b = rc + d, \quad \text{где} \quad d < c. \tag{3}$$

Общие делители b и c совпадают с общими делителями c и d.

Этот процесс может закончиться, только когда осуществится точная делимость, т. е. когда мы дойдем в последовательности a, b, c, ... до некоторого числа, являющегося делителем предыдущего. С другой стороны, ясно, что процесс должен окончиться, так как убывающая последовательность a, b, c, ... натуральных чисел не может быть бесконечной.

Предположим для определенности, что процесс закончится, когда мы дойдем до числа h, являющегося делителем предыдущего числа g. Тогда последние два уравнения в ряду (2), (3), ... таковы:

$$f = vg + h, \tag{4}$$

$$g = wh. \tag{5}$$

Общие делители a и b являются также общими делителями b и c, c и d и так далее вплоть до g и h. Так как h делит g, общие делители g и h состоят просто из всех делителей h. Число h можно рассматривать как последний остаток в алгоритме Евклида перед осуществлением точной делимости, т. е. как последний ненулевой остаток.

Таким образом, мы доказали, что *общие делители двух данных натуральных чисел a и b состоят из всех делителей некоторого натурального числа h (Н. О. Д. чисел a и b); это число является последним ненулевым остатком при применении алгоритма Евклида к числам a и b.*

В качестве численного примера возьмем числа 3132 и 7200, которые уже были использованы в п. 5. Последовательные шаги алгоритма в этом случае таковы:

$$7200 = 2 \cdot 3132 + 936,$$
$$3132 = 3 \cdot \ 936 + 324,$$

$$936 = 2 \cdot 324 + 288,$$
$$288 = 8 \cdot 36;$$

Н. О. Д. равен последнему остатку 36. Часто можно немного сократить работу, используя отрицательные остатки, если они численно меньше соответствующих положительных остатков. В вышеупомянутом примере последние три шага можно заменить двумя:

$$936 = 3 \cdot 324 - 36,$$
$$324 = 9 \cdot 36.$$

Возможность использования отрицательных остатков основана на том, что рассуждение, примененное к уравнению (2), остается в силе, если это уравнение заменить на уравнение

$$a = qb - c.$$

Говорят, что два числа являются взаимно простыми *), если они не имеют общего делителя, отличного от 1, или, другими словами, если их Н. О. Д. равен 1. Это возможно тогда и только тогда, когда последний остаток алгоритма Евклида, примененного к рассматриваемым числам, равен 1.

7. Другое доказательство основной теоремы. Применим теперь алгоритм Евклида для того чтобы дать новое доказательство основной теоремы, которое не зависит от приведенного в п. 4.

Начнем с одного простого замечания; ввиду своей очевидности оно может даже показаться излишним. Пусть a, b, n — произвольные натуральные числа. *Наибольший общий делитель* na *и* nb *есть* n *раз взятый наибольший общий делитель* a *и* b. Хотя это и кажется очевидным, читатель может убедиться, что дать прямое доказательство этого факта, не используя ни алгоритма Евклида, ни основной теоремы арифметики, нелегко.

Однако из алгоритма Евклида результат следует немедленно. Можно предполагать, что $a > b$. Если разделить na на nb, то отношение будет тем же, что и раньше (т. е. q, а остаток заменится на nc вместо c. Уравнение (2) заменится на равенство

$$na = q \cdot nb + nc.$$

*) Это, конечно, то же определение, что и в п. 5; оно здесь повторено, потому что настоящее рассмотрение не зависит от предыдущего.

То же самое применимо и к дальнейшим уравнениям: все они просто умножатся на n. Наконец, последний остаток, дающий Н. О. Д. чисел na и nb, равен nh, где h — Н. О. Д. a и b.

Мы используем этот простой результат для доказательства следующей теоремы, которую часто называют теоремой Евклида (она встречается в качестве предложения 30 книги VII). *Если простое делит произведение двух чисел, то оно должно делить одно из этих чисел* (или, возможно, каждое из них). Предположим, что простое p делит произведение na, но не делит a. Единственными делителями p являются 1 и p, поэтому единственный общий множитель p и а равен 1. Следовательно, по только что доказанной теореме Н. О. Д. np и na равен n. Далее ясно, что p делит np, кроме того, p делит na по предположению. Следовательно, p есть общий делитель np и na и потому делит n, ибо каждый общий делитель двух чисел является также делителем их Н. О. Д. Таким образом, мы доказали, что если p делит na и не делит a, то оно должно делить n, а это и есть теорема Евклида.

Теперь уже можно доказать единственность разложения на простые. Предположим, что число n имеег два разложения, скажем,

$$n = pqr \ldots = p'q'r' \ldots,$$

где все числа p, q, r, $\ldots$, p', q', r', $\ldots$ простые. Так как p делит произведение $p'(q'r' \ldots)$, то оно должно делить p' или $q'r' \ldots$. Если p делит p', то $p = p'$, так как эти числа простые. Если p делит $q'r' \ldots$, повторим предыдущее рассуждение; в конце концов, мы придем к заключению, что p равно одному из простых p', q', r', $\ldots$. Мы сможем сократить на общее простое p оба представления и вернуться к какому-нибудь другому числу, стоящему в левой части равенства, скажем, к q. Таким образом, можно установить, что слева и справа одинаковые простые, и оба представления совпадают.

Этим получено доказательство единственности разложения на множители, упоминаемое в п. 4. Его достоинство в том, что оно получается из общей теории (теории алгоритма Евклида), а не искусственным приемом, использованным в п. 4. С другой стороны, оно длиннее и менее непосредственно.

8. Одно свойство Н. О. Д. Из алгоритма Евклида можно вывести одно замечательное свойство Н. О. Д., вовсе не очевидное из его построения с помощью разложения на простые. Это свойство таково: *наибольший общий делитель h двух натуральных чисел a и b представляется как разность между некоторым кратным a и некоторым кратным b, т. е.*

$$h = ax - by,$$

где x и y — натуральные числа.

Так как a и b кратны h, любое число вида $ax - by$ также кратно h; мы утверждаем, что найдутся такие значения x и y, для которых $ax - by$ в точности равно h.

Прежде чем давать доказательство, отметим некоторые свойства чисел, представимых в виде $ax - by$. Во-первых, число, представимое таким образом, может быть также представлено как $by' - ax'$, где x' и y' — натуральные числа. Действительно, эти два выражения равны, если

$$a(x + x') = b(y + y'),$$

а этого можно достичь, взяв произвольное число m и определив x' и y' с помощью равенств

$$x + x' = mb, \quad y + y' = ma.$$

Числа x' и y' будут натуральными числами, если m взять настолько большим, чтобы было $mb > x$ и $ma > y$. Если x' и y' определены таким образом, то

$$ax - by = by' - ax'.$$

Будем говорить, что *число линейно зависит от a и b*, если оно представляется в виде $ax - by$. Только что установленный результат показывает, что линейная зависимость от a и b не нарушается, если a и b поменять местами.

Имеют место следующие два простых факта о линейной зависимости. Если какое-нибудь число линейно зависит от a и b, то тем же свойством обладает и любое кратное этого числа; действительно,

$$k(ax - by) = a \cdot kx - b \cdot ky.$$

Сумма двух чисел, линейно зависящих от a и b, также линейно зависит от a и b, ибо

$$(ax_1 - by_1) + (ax_2 - by_2) = a(x_1 + x_2) - b(y_1 + y_2).$$

То же применимо и к разности двух чисел; чтобы доказать это, запишем второе число в виде $by_2' - ax_2'$ (это возможно благодаря сделанному ранее замечанию) и вычтем его из первого. Тогда

$$(ax_1 - by_1) - (by_2' - ax_2') = a(x_1 + x_2') - b(y_1 + y_2').$$

Таким образом, свойство линейной зависимости от a и b сохраняется при сложении, вычитании и умножении на любое число.

Исследуем теперь в свете этого понятия алгоритм Евклида. Сами числа a и b, конечно, линейно зависят от a и b, так как

$$a = a(b+1) - b(a), \quad b = a(b) - b(a-1).$$

Первым уравнением алгоритма было уравнение

$$a = qb + c.$$

Так как b линейно зависит от a и b, то и qb линейно зависит от a и b, а так как a также линейно зависит от a и b, то этим свойством обладает и $a - qb$, т. е. c. Из следующего уравнения тем же способом можно вывести, что d линейно зависит от a и b и так далее до тех пор, пока мы не дойдем до последнего остатка, равного h. Этим доказано, что h линейно зависит от a и b, а это и утверждалось.

В качестве иллюстрации рассмотрим пример, уже приводившийся в п. 6: положим $a = 7200$ и $b = 3132$. Используем уравнения алгоритма Евклида для того, чтобы выразить каждый из остатков через a и b. Первое уравнение

$$7200 = 2 \cdot 3132 + 936$$

показывает, что

$$936 = a - 2b.$$

Второе уравнение

$$3132 = 3 \cdot 936 + 324$$

дает

$$324 = b - 3(a - 2b) = 7b - 3a.$$

Из третьего уравнения

$$936 = 2 \cdot 324 + 288$$

получаем

$$288 = (a - 2b) - 2(7b - 3a) = 7a - 16b.$$

Четвертое уравнение

$$324 = 1 \cdot 288 + 36$$

дает

$$36 = (7b - 3a) - (7a - 16b) = 23b - 10a.$$

Это и есть искомое выражение наибольшего общего делителя 36 в виде разности двух кратных чисел a и b. Если мы хотим получить выражение, в котором a является первым слагаемым, достаточно положить

$$23b - 10a = (M - 10)a - (N - 23)b, \quad \text{где} \quad Ma = Nb.$$

Так как общий делитель a и b равен 36, то (после сокращения на него) условия, связывающие M и N, принимают вид

$$200M = 87N.$$

Простейший выбор M и N ($M = 87$, $N = 200$) после подстановки дает

$$36 = 77a - 177b.$$

Возвращаясь к общей теории, выразим полученный результат в другой форме. Пусть a, b, n — данные натуральные числа; требуется найти натуральные числа x и y так, чтобы выполнялось равенство

$$ax - by = n. \tag{6}$$

Такое уравнение называется *неопределенным* (ибо оно однозначно не определяет x и y), или *диофантовым,* уравнением в честь Диофанта из Александрии (третий век нашей эры), написавшего знаменитый трактат по арифметике. Уравнение (6) не имеет решений, если n не кратно h — наибольшему общему делителю a и b, так как этот наибольший общий делитель делит $ax - by$ при любых x и y. Предположим теперь, что n кратно h, скажем, $n = mh$. Тогда мы можем решить это уравнение; прежде всего решим уравнение

$$ax_1 - by_1 = h.$$

(мы уже видели, как это сделать), после чего, умножив x_1 и y_1 на m, получим решение $x = mx_1$, $y = my_1$ уравнения (6). Таким образом, *линейное неопределенное уравнение* (6) *разрешимо в натуральных числах x и y тогда и только тогда, когда n кратно h.* В частности, если a и b взаимно просты, так что $h = 1$, то уравнение разрешимо для любого n.

Мы нашли условие разрешимости линейного неопределенного уравнения

$$ax + by = n$$

в целых числах x и y противоположных знаков: одно из них положительно, другое отрицательно. Вопрос о том, когда это

уравнение разрешимо в натуральных числах, труднее, и на него нельзя полностью ответить столь простым способом. Конечно, n должно быть кратно h, но n должно быть также не слишком малым по отношению к a и b. Довольно легко доказать, что уравнение разрешимо в натуральных числах, если n кратно h и $n > 2ab$.

9. Разложение чисел на множители. Простейший способ разложить число на множители — это испытать, делится ли данное число на 2, на 3, на 5 и так далее, используя ряд простых. Если число N не делится ни на какое простое вплоть до $\sqrt{N}$, то оно само является простым, так как составное число имеет по крайней мере два простых множителя и они не могут быть оба больше $\sqrt{N}$.

Это очень трудоемкий процесс, если раскладываемое число велико, поэтому составлены таблицы разложений на множители. Самой большой из доступных таблиц является таблица Лемера (Carnegie Institute, Washington Pub. No. 105, 1909), в которой приводится наименьший простой множитель любого числа вплоть до 10 000 000. Если наименьший простой множитель некоторого числа известен, то это число можно разделить на него; повторение этого процесса дает полное разложение числа на простые множители.

Многими математиками, в том числе Ферма и Гауссом, были открыты способы, сокращающие число шагов при разложении на множители большого числа. Для понимания большинства из этих способов нужно знать теорию чисел лучше, чем это здесь предполагается. Имеется, однако, один довольно простой метод Ферма, который описывается в нескольких словах.

Пусть N — данное число и m — наименьшее целое, для которого $m^2 > N$. Образуем числа

$$m^2 - N, \quad (m+1)^2 - N, \quad (m+2)^2 - N, \ \ldots \tag{7}$$

Если какое-нибудь из этих чисел будет равно точному квадрату, мы получим уравнение $x^2 - N = y^2$, из которого следует, что $N = x^2 - y^2 = (x-y)(x+y)$. Вычисление чисел (7) облегчается тем, что их последовательные разности возрастают с постоянной скоростью. С помощью таблицы квадратов Барлоу (Barlow) можно легко определить, какие из этих чисел являются точными квадратами. Этот метод удобен, если число N разлагается

на два множителя почти одинаковой величины, так как тогда y мало. Если N простое, процесс будет продолжаться до тех пор, пока мы не дойдем до решения

$$x + y = N, \quad x - y = 1.$$

В качестве примера возьмем $N = 9271$. Оно лежит между 96^2 и 97^2, так что $m = 97$. Первым числом в ряду (7) будет число

$$97^2 - 9271 = 138.$$

Следующие получаются добавлением последовательно $2m + 1$, $2m + 3$ и так далее, т. е. 195, 197 и так далее. Это дает ряд чисел

$$138, 333, 530, 729, 930, \ldots$$

Четвертое из них — точный квадрат (оно равно 27^2), и мы получаем

$$9271 = 100^2 - 27^2 = 127 \cdot 73.$$

Интересный алгоритм разложения был открыт капитаном военно-морских сил США Дрэмом (N. A. Draim). В этом алгоритме результат каждого шага деления используется для того, чтобы подготовить число к следующему делению. Имеются различные формы этого алгоритма, простейшей, вероятно, является форма, в которой последовательными делителями служат нечетные числа 3, 5, 7, 9, ..., простые или составные. Поясним действие этого алгоритма на численном примере; возьмем, скажем, $N = 4511$. Первый шаг — деление на 3, частное 1503, остаток 2:

$$4511 = 3 \cdot 1503 + 2.$$

Следующий шаг — вычитание удвоенного частного из данного числа и добавление остатка:

$$4511 - 2 \cdot 1503 = 1505; \quad 1505 + 2 = 1507.$$

Последнее число надо разделить на следующее нечетное число 5:

$$1507 = 5 \cdot 301 + 2.$$

Следующий шаг — вычитание удвоенного частного из первого, получившегося на предыдущем шагу числа (из 1505 в данном случае) и добавление к нему последнего остатка:

$$1505 - 2 \cdot 301 = 903; \quad 903 + 2 = 905.$$

Это число надо разделить на следующее нечетное число 7. Да-

лее мы действуем таким же образом:

$$905 = 7 \cdot 129 + 2;$$
$$903 - 2 \cdot 129 = 645; \quad 645 + 2 = 647;$$
$$647 = 9 \cdot 71 + 8;$$
$$645 - 2 \cdot 71 = 503; \quad 503 + 8 = 511;$$
$$511 = 11 \cdot 46 + 5;$$
$$503 - 2 \cdot 46 = 411; \quad 411 + 5 = 416;$$
$$416 = 13 \cdot 32 + 0.$$

Мы получили нулевой остаток, и алгоритм говорит нам, что 13 является делителем данного числа 4511. Второй сомножитель находится в результате первой половины следующего шага:

$$411 - 2 \cdot 32 = 347.$$

Действительно,

$$4511 = 13 \cdot 347,$$

так как 347 — простое число, разложение закончено.

Обосновать действие алгоритма в общем случае — задача элементарной алгебры. Пусть N_1 — данное число; первый шаг состоял в том, чтобы представить N_1 в виде

$$N_1 = 3q_1 + r_1.$$

На следующем шагу мы строим числа

$$M_2 = N_1 - 2q_1 \quad \text{и} \quad N_2 = M_2 + r_1.$$

Затем делим число N_2 на 5:

$$N_2 = 5q_2 + r_2.$$

на третьем шагу мы строим числа

$$M_3 = M_2 - 2q_2 \quad \text{и} \quad N_3 = M_3 + r_2.$$

и так далее. Из написанных уравнений следует, что

$$N_2 = 2N_1 - 5q_1,$$
$$N_3 = 3N_1 - 7q_1 - 7q_2,$$
$$N_4 = 4N_1 - 9q_1 - 9q_2 - 9q_3$$

и так далее. Значит, N_2 делится на 5 тогда и только тогда, когда $2N_1$ делится на 5, или N_1 делится на 5. N_3 делится на 7 тогда и только тогда, когда $3N_1$ делится на 7, или N_1 делится на 7, и так далее. Когда мы дойдем до наименьшего простого делителя N_1 будет иметь место точная делимость и нулевой остаток.

Общее уравнение, аналогичное приведенным выше, имеет вид

$$N_n = nN_1 - (2n+1)(q_1 + q_2 + \cdots + q_{n-1}); \tag{8}$$

общее уравнение для M_n имеет вид

$$M_n = N_1 - 2(q_1 + q_2 + \cdots + q_{n-1}). \tag{9}$$

Если $2n+1$ — делитель данного числа N_1, то N_n делится на $2n+1$, так что

$$N_n = (2n+1)q_n,$$

откуда $nN_1 = (2n+1)(q_1 + q_2 + \cdots + q_n)$ в силу (8). Далее, в силу (9) мы имеем

$$M_{n+1} = N_1 - 2(q_1 + \cdots + q_n) = N_1 - 2\left(\frac{n}{2n+1}\right)N_1 = \frac{N_1}{2n+1}.$$

Таким образом, множителем, дополнительным к $2n+1$, является M_{n+1}, что и получилось в численном примере.

В рассмотренном численном примере N_1, N_2, ... убывают монотонно. В общем случае это свойство имеет место лишь в начале алгоритма и может не выполняться в дальнейшем. Тем не менее оказывается, что последующие числа всегда много меньше, чем исходное число.

10. Простые числа. Понятие простого числа очень естественно, однако вопросы, связанные с простыми, часто весьма трудны, и на многие из этих вопросов современная математика не в состоянии дать удовлетворительный ответ. Мы заканчиваем эту главу кратким перечислением некоторых результатов и гипотез, связанных с простыми числами.

В п. 3 мы привели евклидово доказательство бесконечности числа простых. То же рассуждение применимо для доказательства бесконечности числа простых некоторого специального вида. Так как каждое простое число, большее 2, нечетно, то любое простое принадлежит одной из двух арифметических прогрессий

(a) 1, 5, 9, 13, 17, 21, 25, ...,
(b) 3, 7, 11, 15, 19, 23, 27, ...;

прогрессия (a) состоит из всех чисел вида $4x+1$, а прогрессия (b) — из всех чисел вида $4x-1$ (или $4x+3$, что то же самое). Докажем сначала, что *в прогрессии* (b) *имеется бесконечно много*

простых. Пусть простые в (b) перенумерованы: q_1, q_2, ..., начиная с $q_1 = 3$. Рассмотрим число N, определенное равенством

$$N = 4(q_1 q_2 \dots q_n) - 1.$$

Это число также имеет вид $4x-1$. Каждый простой множитель N не может иметь вид $4x+1$, потому что произведение чисел вида $4x+1$ само является числом такого же вида:

$$(4x+1)(4y+1) = 4(4xy + x + y) + 1.$$

Значит, число N имеет простой множитель вида $4x-1$. Этот множитель не может равняться ни одному из чисел q_1, q_2, ..., q_n, ибо при делении на каждое из них число N дает остаток -1. Значит в ряду (b) найдется простое число, отличное от q_1, q_2, ..., q_n. Этим высказанное утверждение доказано.

Проведенное рассуждение неприменимо для доказательства бесконечности простых в ряду (a), ибо число вида $4x+1$ может не иметь ни одного простого множителя такого вида. Однако для доказательства этого можно воспользоваться другим методом. Обозначим простые ряда (a) через r_1, r_2, ... и рассмотрим число

$$M = (r_1 r_2 \dots r_n)^2 + 1.$$

Позднее (III, 3) мы увидим, что любое число вида a^2+1 имеет простой множитель вида $4x+1$ (число a^2+1 полностью состоит из множителей вида $4x+1$ и, быть может, множителя 2). Число M, очевидно, не делится ни на одно из простых r_1, r_2, ..., r_n; отсюда, как и раньше, следует, что *прогрессия* (a) *содержит бесконечно много простых чисел.*

Аналогичная ситуация возникает при рассмотрении прогрессий вида $6x+1$ и $6x-1$. Эти прогрессии включают все числа, не делящиеся на 2 или на 3, и поэтому каждое простое, большее 3, входит в одну из этих прогрессий. Методами, аналогичными только что использованным приемам, можно доказать бесконечность простых в каждой из этих прогрессий. Но такими методами нельзя решить вопрос об общей арифметической прогрессии. Эта прогрессия состоит из всех чисел $ax+b$, где a и b фиксированы, а $x = 0, 1, 2, \dots$, т. е. из чисел вида

$$b,\ b+a,\ b+2a,\ \dots$$

Если a и b имеют общий множитель, то каждый член прогрес-

сии имеет такой же множитель и, значит, не является простым (за исключением, быть может, первого числа b). Поэтому нужно предположить, что a и b взаимно простые. Кажется правдоподобным, что такая прогрессия будет содержать бесконечно много простых, т. е. что *если a и b взаимно просты, то существует бесконечно много простых вида $ax + b$.*

Вероятно, Лежандр был первым, кто понял важность этого предложения. Одно время он предполагал, что обладает доказательством, но ошибся. Первое доказательство было получено Дирихле и опубликовано в мемуарах, появившихся в 1836 году. Это доказательство использовало аналитические методы (функции комплексной переменной, пределы и бесконечные ряды) и явилось первым действительно ценным приложением таких методов к теории чисел. Оно открыло совершенно новые пути развития теории чисел; идеи, лежащие в основе методов Дирихле, носили очень общий характер и явились основой для большой последующей работы по приложению аналитических методов к теории чисел.

О представимости простых другими выражениями известно мало. Предполагается, например, что имеется бесконечно много простых вида $x^2 + 1$; вот несколько первых из них:

$$2,\ 5,\ 17,\ 37,\ 101,\ 197,\ 257,\ \ldots$$

Но в доказательстве этого утверждения не было достигнуто ни малейшего успеха, и вопрос этот до сих пор остается безнадежно трудным. Дирихле доказал тем не менее, что любая квадратичная форма от двух переменных, т. е. любая форма вида $ax^2 + bxy + cy^2$, в которой коэффициенты a, b, c взаимно просты, представляет бесконечно много простых.

В более позднее время был глубоко исследован вопрос о том, как часто встречаются простые, т. е. вопрос о том, сколько простых имеется в ряду целых чисел $1, 2, \ldots, X$ при большом X. Это число, зависящее, конечно, от X, обозначается обычно через $\pi(X)$. Первая гипотеза о величине $\pi(X)$ как функции от X, по-видимому, была сделана независимо Лежандром и Гауссом около 1800 года. Она состояла в том, что $\pi(X)$ примерно равно $X/\ln X$. Здесь $\ln X$ обозначает натуральный (так называемый неперов) логарифм X, т. е. логарифм X по основанию e. Это

предположение было, вероятно, основано на вычислениях. Например, при $X = 1\,000\,000$ найдено, что $\pi(1\,000\,000) = 78\,498$, в то время как величина $X/\ln X$ (округленная до ближайшего целого) равна 72 382, их отношение равно $1,084\ldots$. Вычисления такого рода могут, конечно, ввести в заблуждение. Но здесь предполагаемый результат оказался верен в том смысле, что отношение $\pi(X)$ и $X/\ln X$ стремится к 1 при X, стремящемся к бесконечности. Это составляет содержание известного асимптотического закона распределения простых чисел, впервые независимо доказанного Адамаром и Валле-Пуссеном в 1896 году с помощью новых и сильных аналитических методов.

Здесь невозможно перечислить многие другие теоремы, связанные с распределением простых. Теоремы, доказанные в девятнадцатом столетии, являются лишь несовершенными приближениями к асимптотическому закону; результаты, полученные в двадцатом веке, включают различные усовершенствования этой теоремы. Имеется, однако, один недавний результат, о котором следует упомянуть. Мы уже говорили, что доказательство теоремы Дирихле о простых в арифметических прогрессиях и доказательство асимптотического закона были аналитическими и использовали внешние по отношению к теории чисел методы. Сами же утверждения относятся непосредственно к натуральным числам, и казалось разумным искать доказательства, не использующие чуждых теории чисел идей. Поиски элементарных доказательств этих двух теорем были безуспешными до самого последнего времени. В 1948 году А. Сельберг нашел первое элементарное доказательство теоремы Дирихле, а затем с помощью П. Эрдеша он нашел и первое элементарное доказательство асимптотического закона распределения простых чисел. Под «элементарным» доказательством здесь подразумевается доказательство, которое оперирует только с натуральными числами. Такое доказательство не обязательно просто; оба этих доказательства весьма трудны.

Упомянем еще об одной известной проблеме, связанной с простыми, которая была предложена Гольдбахом в письме к Эйлеру в 1742 году. Гольдбах предположил (в несколько других терминах), что всякое четное число, большее 6, представляется как сумма двух простых, отличных от 2, например

$$6 = 3 + 3,\ \ 8 = 3 + 5,\ \ 10 = 3 + 7,\ \ 12 = 5 + 7,\ \ \ldots$$

Всякая задача, которая подобно этой имеет дело с *аддитивными* свойствами простых, по необходимости трудна, так как определение простого и естественные свойства простых выражаются в терминах умножения. Важный вклад в этот предмет сделали Харди и Литлвуд в 1923 году, но до 1930 года не было ни одного строго доказанного утверждения, которое можно было бы рассматривать хотя бы как далекое приближение к решению проблемы Гольдбаха. В 1930 году русский математик Шнирельман доказал существование такого числа N, что *любое число, начиная с некоторого места, представимо в виде суммы не более чем N простых.* Много ближе подошел к решению этого вопроса в 1937 году Виноградов. Он доказал с помощью очень тонких аналитических методов, что *каждое достаточно большое нечетное число представимо в виде суммы трех простых.* Его доказательство явилось исходным пунктом большой новой работы в аддитивной теории простых, по ходу которой были решены многие задачи, недоступные никаким методам, предшествующим методу Виноградова. Новейший результат Реньи, связанный с проблемой Гольдбаха, состоит в том, что всякое достаточно большое четное число представило в виде суммы двух чисел, одно из которых является простым, а другое имеет ограниченное число простых множителей.

Замечания к главе I. п. 1. Главная трудность при перечислении законов арифметики, подобном приведенному здесь, состоит в решении того, какие из понятий являются первичными. Различные подходы к этому вопросу зависят от вкусов авторов.

Нашей целью не является дальнейший анализ понятий и законов арифметики. Мы придерживаемся точки зрения здравого смысла (или наивной точки зрения), что все «знают» натуральные числа и удовлетворены истинностью законов арифметики и принципа индукции. Читатель, интересующийся основаниями математики, может обратиться к книгам: B. Russell, Introduction to Mathematical Philosophy (Allen and Unwin, London); M. Black, The Nature of Mathematics (Harcourt and Brace, New York).

Рассел определяет натуральные числа, выделяя их из чисел более общего рода. Эти более общие числа — (конечные или бесконечные) кардинальные числа, определяемые с помощью общих понятий

«класса» и «взаимно однозначного отображения». Выделение производится определением натуральных чисел как чисел, обладающих всеми индуктивными свойствами (Russell, p. 27). Но разумно ли базировать теорию натуральных чисел на таком неясном и неудовлетворительном понятии, как класс, — вопрос спорный. Dolus latet in universalibus *).

п. 2. Возражением против использования принципа индукции для определения натуральных чисел является необходимость ссылки на «любое предложение о натуральном числе n». Кажется ясным, что «предложения», упоминаемые здесь, должны быть утверждениями, имеющими смысл, когда в них говорится о натуральных числах. Неясно, однако, как эта осмысленность может быть испытана или оценена, если неизвестно, что такое натуральное число.

п. 4. Я не встречал этого доказательства единственности разложения, но маловероятно, чтобы оно было новым. О других прямых доказательствах см. ([11], стр. 2) и ([9], стр. 21).

п. 6. Критически настроенный читатель мог бы заметить, что в двух местах этого пункта я использовал принципы, которые не были точно сформулированы в пп. 1 и 2. В обоих случаях можно было дать доказательство по индукции, но, сделав так, я отвлек бы внимание читателя от главных вопросов. Вопрос о длине алгоритма Евклида обсуждается в ([14], глава 3) и в ([16], том I, гл. 4).

п. 9. О методах разложения на множители см. ([7], том I, гл. 14); об алгоритме Дрэма см. Mathematics Magazine **25** (1952), 191—194.

п. 10. Превосходное изложение вопроса о распределении простых дано в книжке Ингама «Распределение простых чисел» (М.—Л., 1936) **). Доказательство Дирихле его теоремы (с усовершенствованием, принадлежащим Мертенсу) дано в качестве приложения к ([8]). Для ссылок на элементарные доказательства теоремы Дирихле и асимптотического закона распределения простых см. Math. Reviews, **10** (1949), 595—596. Элементарное доказательство асимптотического закона дано в главе 22 ([9]) и в главе 3 ([5]). Для обозрения работ, связанных с проблемой Гольдбаха, см. James, Bull. Amer. Math. Soc. **55** (1949). Доказательство результата Виноградова можно найти у Эстермана (Estermann, Introduction to modern prime number theory, Cambridge, 1952). О результате Реньи см. Math. Reviews **9** (1948), 413.

*) Общие понятия полны коварства. (*Прим. перев.*)

**) Автор ссылается на английское издание этой книги. The Distribution of Prime Numbers (Cambridge, 1932). (*Прим. перев.*)

ГЛАВА II

СРАВНЕНИЯ

1. Понятие сравнения. В некоторых случаях два числа, отличающиеся на кратное какого-нибудь фиксированного числа, эквивалентны, так что вычисления с этими числами приводят к одному и тому же результату. Например, значение $(-1)^n$ зависит лишь от четности или нечетности n, так что два значения n, отличающиеся на кратное 2, приводят к одинаковому результату. Если мы интересуемся только последней цифрой числа, то числа, отличающиеся на кратное 10, для нас, по существу, совпадают.

Понятие сравнения, введенное Гауссом, выражает в удобной форме то, что два целых числа a и b отличаются на кратное фиксированного натурального числа m. Мы говорим в этом случае, что a *сравнимо с* b *по модулю* m или, символически,

$$a \equiv b \pmod{m},$$

Это выражение означает просто, что $a - b$ делится на m. Такое обозначение благодаря аналогии между сравнениями и равенствами упрощает вычисления, в которых числа, отличающиеся на кратное m, фактически не различаются. Сравнение есть «равенство с точностью до некоторого кратного m». Несколько примеров истинных сравнений:

$$63 \equiv 0 \pmod{3}, \quad 7 \equiv -1 \pmod{8}, \quad 5^2 \equiv -1 \pmod{13}.$$

Сравнение по модулю 1 выполняется для любых двух чисел, так как каждое число кратно 1. Два числа сравнимы по модулю 2, если они имеют одинаковую четность, т. е. оба четны или оба нечетны.

Два сравнения по одинаковому модулю можно складывать, вычитать или перемножать так же, как и равенства. Если

$$a \equiv \alpha \pmod{m} \quad \text{и} \quad b \equiv \beta \pmod{m},$$

то

$$a+b \equiv \alpha+\beta \pmod m,$$
$$a-b \equiv \alpha-\beta \pmod m,$$
$$ab \equiv \alpha\beta \pmod m.$$

Первые два из этих утверждений получаются немедленно; например, $(a+b)-(\alpha+\beta)$ кратно m, так как $a-\alpha$ и $b-\beta$ кратны m. Третье сравнение лучше всего доказать в два шага. Во-первых, $ab \equiv \alpha b$, потому что $ab-\alpha b=(a-\alpha)b$, и $a-\alpha$ кратно m. Далее, по тем же соображениям $ab \equiv \alpha\beta$. Значит, $ab \equiv \alpha\beta \pmod m$.

Сравнение всегда можно умножить на целое число: если $a \equiv \alpha \pmod m$, то $ka \equiv k\alpha \pmod m$. Это — частный случай третьего из только что установленных результатов (если b и β оба равны k). Однако сокращение сравнения на какой-либо множитель не всегда возможно. Например, $42 \equiv 12 \pmod{10}$, но нельзя сократить числа 42 и 12 на множитель 6; такое сокращение приводит к неверному сравнению $7 \equiv 2 \pmod{10}$. Причина очевидна: первое сравнение утверждает, что $42-12$ кратно 10, но отсюда не вытекает, что $\frac{1}{6}(42-12)$ кратно 10. Сокращение на множитель допустимо, если этот множитель *взаимно прост с модулем*. В самом деле, пусть имеется сравнение $ax \equiv ay \pmod m$ и пусть a — множитель, на который его нужно сократить; предположим, что a взаимно просто с m. Сравнение утверждает, что $a(x-y)$ делится на m, а из последнего предложения в (I, 5) тогда следует, что $x-y$ делится на m.

Пример использования сравнений доставляется хорошо известными признаками делимости на 3, на 9 и на 11. Обычное представление числа n цифрами в десятичной системе счисления есть просто представление n в виде

$$n = a+10b+100c+\ldots,$$

где a, b, c, $\ldots$ — цифры числа, прочитанные справа налево, так что a — число единиц, b — число десятков и так далее. Так как $10 \equiv 1 \pmod 9$, то $10^2 \equiv 1 \pmod 9$, $10^3 \equiv 1 \pmod 9$ и так далее. Поэтому из только что написанного представления следует, что

$$n \equiv a+b+c+\ldots \pmod 9.$$

Другими словами, любое число отличается от суммы своих цифр на кратное 9; в частности, n делится на 9 тогда и только тогда, когда сумма его цифр делится на 9. Это рассуждение проходит

также и при замене 9 на 3.

Признак делимости на 11 основан на том, что $10 \equiv -1 \pmod{11}$, так что $10^2 \equiv +1 \pmod{11}$, $10^3 \equiv -1 \pmod{11}$ и так далее. Поэтому

$$n \equiv a - b + c - \ldots \pmod{11}.$$

Следовательно, n делится на 11 тогда и только тогда, когда $a - b + c - \ldots$ делится на 11. Например, чтобы проверить, делится ли число 9581 на 11, образуем сумму $1 - 8 + 5 - 9$, равную -11. Так как эта сумма делится на 11, то и 9581 делится на 11.

2. Линейные сравнения. Ясно, что каждое целое число сравнимо по mod m точно с одним из следующих чисел:

$$0, 1, 2, \ldots, m - 1, \tag{1}$$

так как всякое целое число представимо в виде $qm + r$, где $0 \leqslant r < m$, и, значит, сравнимо с r по mod m. Очевидно, что, помимо ряда (1), имеются и другие ряды чисел, обладающие тем же свойством; так, например, любое целое число сравнимо по mod 5 точно с одним из чисел 0, 1, -1, 2, -2. Любой такой ряд чисел называется *полной системой вычетов* по модулю m. Таким образом, полной системой вычетов по mod m называется любой ряд из m чисел, никакие два из которых не сравнимы друг с другом.

Под *линейным* сравнением, по аналогии с линейными уравнениями в элементарной алгебре, понимается сравнение вида

$$ax \equiv b \pmod{m}. \tag{2}$$

Важно отметить, что любое такое сравнение разрешимо относительно x при условии, что a взаимно просто с m. Простейший способ доказательства этого утверждения состоит в следующем: если x пробегает полную систему вычетов, то соответствующие значения ax также образуют полную систему вычетов. Действительно, таких чисел ровно m и никакие два из них не сравнимы между собой, так как из сравнения $ax_1 \equiv ax_2 \pmod{m}$ следует сравнение $x_1 \equiv x_2 \pmod{m}$. (Здесь на множитель a можно сократить, так как a и m взаимно просты.) Так как числа ax образуют полную систему вычетов, то среди них найдется (и притом в точности одно) число, сравнимое с

данным числом b.

Рассмотрим, например, сравнение

$$3x \equiv 5 \pmod{11}.$$

Если придавать x значения 0, 1, 2, ..., 10 (полная система вычетов по модулю 11), то $3x$ будет принимать значения 0, 3, 6, ..., 30. Они образуют другую полную систему вычетов по mod 11; эти числа сравнимы соответственно с 0, 3, 6, 9, 1, 4, 7, 10, 2, 5, 8. Число 5 встречается при $x = 9$, значит, $x = 9$ есть решение сравнения. Каждое число, сравнимое с 9 по модулю 11, также, конечно, будет удовлетворять сравнению, но тем не менее мы говорим, что у этого сравнения есть только *одно* решение, имея в виду, что существует лишь одно решение в каждой полной системе вычетов, т. е. все решения сравнимы между собой. То же самое применимо и к общему сравнению (2); такое сравнение (если a взаимно просто с m) в точности эквивалентно сравнению $x \equiv x_0 \pmod{m}$, где x_0 какое-нибудь частное решение сравнения.

Можно смотреть на линейное сравнение (2) по-другому. Это сравнение эквивалентно *уравнению* $ax = my + b$ *или* $ax - my = b$. Мы доказали в (I, 8), что такое линейное диофантово уравнение разрешимо относительно x и y, если a и m взаимно просты; отсюда вытекает другое доказательство разрешимости линейного сравнения. Но приведенное выше доказательство проще и показывает преимущества, которые мы получаем, используя понятие сравнения.

Тот факт, что сравнение (2) имеет единственное решение (в объясненном ранее смысле), дает возможность использовать это решение для интерпретации дроби b/a по модулю m. Сделав это, мы получим арифметику по mod m, в которой всегда выполнимы сложение, вычитание и умножение, а также возможно деление, если делитель взаимно прост с m. В этой арифметике имеется только конечное число различных чисел (их ровно m), ибо два числа, сравнимых друг с другом по mod m, отождествляются. Возьмем модуль m равным 11; примерами «арифметики по модулю 11» могут служить сравнения

$$5 + 7 \equiv 1, \quad 5 \cdot 6 \equiv 8, \quad \frac{5}{3} \equiv 9 \equiv -2.$$

Любое соотношение, имеющееся между целыми числами или дробями обычной арифметики, остается верным и при интерпретации в новой арифметике. Например, соотношение

$$\frac{1}{2}+\frac{2}{3}=\frac{7}{6}$$

по mod 11 переходит в

$$6+8=3,$$

потому что решением сравнения $2x \equiv 1$ является $x \equiv 6$, решением сравнения $6x \equiv 7$ является $x \equiv 3$, а решением сравнения $3x \equiv 2$ является $x \equiv 8$. Естественно, что интерпретация дробей зависит от модуля, например

$$\frac{2}{3} \equiv 8 \pmod{11}, \quad \text{но} \quad \frac{2}{3} \equiv 3 \pmod{7}.$$

Уже отмечалось, что на эти вычисления накладывается единственное ограничение: знаменатель каждой дроби должен быть взаимно прост с модулем. Если модуль простой (как это было в примерах с модулем 11), то указанное ограничение принимает совсем простой вид: знаменатель не должен быть сравним с 0 (mod m), что в точности аналогично требованию обычной арифметики, в которой знаменатель всегда отличен от 0. Мы еще вернемся к этому вопросу (п. 7).

3. Теорема Ферма. Тот факт, что в арифметике по модулю m имеется только конечное число существенно различных чисел, означает, что имеются алгебраические соотношения, справедливые для *каждого* числа в этой арифметике. Ничего аналогичного этим соотношениям в обычной арифметике нет.

Возьмем число x и рассмотрим его степени x, x^2, x^3, ... Так как для них имеется лишь конечное число возможностей по модулю m, то на некотором месте должна стоять степень, уже встречавшаяся ранее; пусть, скажем,

$$x^h \equiv x^k \pmod{m},$$

где $k < h$. Если x взаимно просто с m, на множитель x^k можно сократить, значит, в этом случае $x^l \equiv 1 \pmod{m}$, где $l = h - k$. Отсюда следует, что каждое число x, взаимно простое с m, удовлетворяет некоторому сравнению такого вида. *Наи-*

меньший показатель l, для которого $x^l \equiv 1 \pmod m$, называется *порядком x по модулю m*. Если x равен 1, его порядок, очевидно, также равен 1. Для иллюстрации этого определения вычислим порядки нескольких элементов по модулю 11. Степени 2, взятые по модулю 11, таковы:

$$2,\ 4,\ 8,\ 5,\ 10,\ 9,\ 7,\ 3,\ 6,\ 1,\ 2,\ 4,\ \ldots$$

Каждая из них есть удвоенная предыдущая, из которой вычитается 11 или кратное 11, если после умножения на два получается число, большее 11. Первая степень 2, сравнимая с 1, равна 2^{10}, значит, порядок 2 (mod 11) равен 10. В качестве другого примера рассмотрим степени 3:

$$3,\ 9,\ 5,\ 4,\ 1,\ 3,\ 9,\ \ldots$$

Первая степень 3, сравнимая с 1, есть 3^5, так что порядок 3 (mod 11) равен 5. Можно найти, что порядок 4 равен 5, таков же и порядок 5.

Легко заметить, что последовательные степени x периодичны; если мы достигли первого числа l, для которого $x^l \equiv 1$, то $x^{l+1} \equiv x$, и предыдущий цикл повторяется. Ясно, что $x^n \equiv \equiv 1 (\mathrm{mod}\ m)$ *тогда и только тогда, когда n кратно порядку x.* В последнем примере $3^n \equiv 1 \pmod{11}$ тогда и только тогда, когда n кратно 5. Так как $3^0 = 1$, то это верно и для $n = 0$; это верно и для отрицательных показателей, если 3^{-n} или $\frac{1}{3^n}$ интерпретируется описанным в п. 2 способом как дробь по mod 11. Отрицательные степени 3 (mod 11) получаются чтением ряда положительных степеней в обратном порядке, так что таблица степеней 3 по модулю 11 имеет вид

$$\begin{array}{rccccccccccc} n = \ldots & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 & 6 \ldots \\ 3^n = \ldots & 9 & 5 & 4 & 1 & 3 & 9 & 5 & 4 & 1 & 3 \ldots \end{array}$$

Ферма заметил, что если модуль простой, скажем p, то каждое целое x, не сравнимое с 0, удовлетворяет сравнению

$$x^{p-1} \equiv 1 \pmod p. \tag{3}$$

Ввиду сказанного ранее этот факт эквивалентен утверждению: порядок любого числа является делителем $p - 1$. Результат (3) был сформулирован Ферма в письме к Френиклю де Бесси (Frenicle de Bessy) 18 октября 1640 года; в этом письме Фер-

ма утверждал также, что обладает доказательством указанного факта. Но, как и в большинстве открытий Ферма, доказательство не было опубликовано и не сохранилось. Первое известное доказательство, по-видимому, принадлежит Лейбницу (1646—1716). Лейбниц доказал, что имеет место сравнение

$$x^p \equiv x \pmod{p}.$$

эквивалентное (3), представив x в виде суммы $1+1+\ldots+1$ из x единиц (x предполагается положительным) и затем раскрыв $(1+1+\ldots+1)^p$ по полиномиальной теореме. Члены $1^p+1^p+\ldots+1^p$ дают p; кроме того, легко доказать, что биномиальные коэффициенты при остальных членах делятся на p.

Совершенно другое доказательство было дано в 1806 году Ивори (Ivory). Если

$$x \not\equiv 0 \pmod{p},$$

то целые числа

$$x,\ 2x,\ 3x,\ \ldots,\ (p-1)x$$

сравнимы (в некотором порядке) с числами 1, 2, 3, ..., $p-1$, так как каждое из этих множеств представляет собой полную систему вычетов, за исключением 0. Так как все элементы этих множеств, взятые в некотором порядке, сравнимы друг с другом, то сравнимы и их произведения, поэтому

$$x \cdot 2x \cdot 3x \ldots (p-1)x \equiv 1 \cdot 2 \cdot 3 \ldots (p-1) \pmod{p}.$$

Сокращая на множители 2, 3, ..., $p-1$ (что допустимо), мы получаем (3).

Одно из достоинств этого доказательства состоит в том, что оно может быть перенесено на более общий случай непростого модуля. Обобщение результата (3) на любой модуль впервые дал Эйлер в 1760 году. Прежде чем формулировать полученный им результат, рассмотрим вопрос о том, сколько чисел в ряду 0, 1, 2, ..., $m-1$ взаимно просто с m. Обозначим количество этих чисел через $\varphi(m)$. Если m простое, то все числа в этом ряду, за исключением 0, взаимно просты с m, так что $\varphi(p) = p-1$ для любого простого p. Эйлеровское обобщение теоремы Ферма для любого модуля от таково:

$$x^{\varphi(m)} \equiv 1 \pmod{m}, \tag{4}$$

если x взаимно просто с m.

Для доказательства достаточно слегка видоизменить метод Ивори, опустив среди чисел $0, 1, \ldots, m-1$ не только число 0, но и все числа, имеющие общие множители с m. Останется $\varphi(m)$ чисел, обозначим их соответственно через $a_1, a_2, \ldots, a_\mu$, здесь $\mu = \varphi(m)$. Тогда числа

$$a_1 x,\ a_2 x,\ \ldots,\ a_\mu x$$

сравнимы (в некотором порядке) с предыдущими числами; перемножив их и затем сократив на $a_1 a_2 \ldots a_\mu$ (что допустимо), получим $x^\mu \equiv 1 \pmod m$; это дает (4).

Для иллюстрации доказательства положим $m = 20$. Числа, меньшие 20 и взаимно простые с 20, таковы:

$$1,\ 3,\ 7,\ 9,\ 11,\ 13,\ 17,\ 19,$$

поэтому $\varphi(20) = 8$. Если мы умножим эти числа на любое число, взаимно простое с 20, то новые числа будут сравнимы с исходными числами, взятыми в некотором другом порядке. Например, если x равно 3, новые числа сравнимы соответственно с

$$3,\ 9,\ 1,\ 7,\ 13,\ 19,\ 11,\ 17 \pmod{20};$$

наше рассуждение доказывает, что $3^8 \equiv 1 \pmod{20}$. Фактически $3^8 = 6561$.

4. Функция Эйлера $\varphi(m)$. Как только что было сказано, $\varphi(m)$ — это количество чисел, меньших m и взаимно простых с m. Естественно спросить, какие соотношения связывают $\varphi(m)$ и m. Мы видели, что $\varphi(p) = p - 1$ для любого простого p. Легко вычислить также $\varphi(p^\alpha)$ для степени простого числа p^α. В ряду $0, 1, \ldots, p^\alpha - 1$ не взаимно просты с p только числа, делящиеся на p. Это — числа вида p^t, где $t = 0, 1, 2, \ldots, p^{\alpha-1}$. Их число равно $p^{\alpha-1}$, а когда мы вычтем это из полного количества p^α всех чисел, меньших p^α, мы получим

$$\varphi(p^\alpha) = p^\alpha - p^{\alpha-1} = p^{\alpha-1}(p-1). \tag{5}$$

Чтобы найти значение $\varphi(m)$ для любого m, докажем, что эта функция *мультипликативна*. Это означает, что для любых *взаимно простых чисел* a и b выполняется равенство

$$\varphi(ab) = \varphi(a)\varphi(b). \tag{6}$$

Начнем с одного общего замечания: *если a и b взаимно просты, то система двух сравнений вида*

$$x \equiv \alpha \pmod{a}, \quad x \equiv \beta \pmod{b} \tag{7}$$

в точности эквивалентна одному сравнению по модулю ab. Действительно, первое сравнение означает, что $x = \alpha + at$, где t — какое-нибудь целое. Число x удовлетворяет второму сравнению тогда и только тогда, когда

$$\alpha + at \equiv \beta \pmod{b} \quad \text{или} \quad at \equiv \beta - \alpha \pmod{b}.$$

Последнее сравнение, будучи линейным, разрешимо относительно t. Значит, система сравнений (7) также разрешима. Если x и x' — два решения этой системы, то $x \equiv x' \pmod{a}$ и $x \equiv x' \pmod{b}$, и поэтому $x \equiv x' \pmod{ab}$. Таким образом, имеется ровно одно решение системы по модулю ab. Этот принцип, обобщенный на несколько сравнений с попарно взаимно простыми модулями, иногда называют «китайской теоремой об остатках». Эта теорема показывает, что имеются числа с любыми предписанными остатками от деления на заданные попарно взаимно простые модули.

Обозначим решение системы сравнений (7) через $x \equiv [\alpha, \beta] \pmod{a}$, так что $[\alpha, \beta]$ — это некоторое число, зависящее от α и β (а также, конечно, от a и b) и однозначно определенное по модулю ab. Различные пары значений α и β порождают разные значения для $[\alpha, \beta]$. Если придавать α значения $0, 1, \ldots, a-1$ (образующие полную систему вычетов по модулю a), а β придавать значения $0, 1, \ldots, b-1$, то получающиеся значения $[\alpha, \beta]$ образуют полную систему вычетов по модулю ab.

Ясно, что если a и α имеют общий множитель, то x из (7) делится на этот множитель; иными словами, $[\alpha, \beta]$ делится на общий множитель α и a. Таким образом, $[\alpha, \beta]$ взаимно просто с ab, только если α взаимно просто с a и β взаимно просто с b, и, наоборот, эти условия гарантируют, что $[\alpha, \beta]$ взаимно просто с ab. Следовательно, если α принимает $\varphi(a)$ значений, меньших a и взаимно простых с a, а β принимает $\varphi(b)$ меньших и взаимно простых с b значений, то для $[\alpha, \beta]$ получается $\varphi(a)\varphi(b)$ значений и среди значений $[\alpha, \beta]$ содержатся все меньшие ab и взаимно простые с ab числа. Значит, $\varphi(ab) = \varphi(a)\varphi(b)$, и равенство (6) доказано.

Для иллюстрации возникающей в только что приведенном доказательстве ситуации составим таблицу величин $[\alpha, \beta]$ при $a = 5$ и $b = 8$. Возможные значения для α — числа 0, 1, 2, 3, 4; возможные значения для β — числа 0, 1, 2, 3, 4, 5, 6, 7. Из них для α имеется четыре значения, взаимно простых с a (в согласии с тем, что $\varphi(5) = 4$); для β также имеется 4 значения,

α \ β	0	1	2	3	4	5	6	7
0	0	25	10	35	20	5	30	15
1	16	*1*	26	*11*	36	*21*	6	*31*
2	32	*17*	2	*27*	12	*37*	22	*7*
3	8	*33*	18	3	28	*13*	38	*23*
4	24	*9*	34	*19*	4	*29*	14	*39*

взаимно простых с b (это согласуется с формулой (5), по которой $\varphi(8) = 4$). Эти значения выделены в таблице курсивом, как и соответствующие им значения $[\alpha, \beta]$. Выделенные курсивом значения $[\alpha, \beta]$ дают 16 чисел, меньших числа 40 и взаимно простых с ним; таким образом,

$$\varphi(40) = \varphi(5)\varphi(8) = 4 \cdot 4 = 16.$$

Вернемся теперь к вопросу о вычислении $\varphi(m)$ для произвольного числа m. Пусть разложение m в произведение степеней простых имеет вид $m = p^\alpha q^\beta \dots$. Тогда из (5) и (6) следует, что

$$\varphi(m) = (p^\alpha - p^{\alpha-1})(q^\beta - q^{\beta-1}) \dots$$

или в более изящной форме

$$\varphi(m) = m\left(1 - \frac{1}{p}\right)\left(1 - \frac{1}{q}\right)\dots \tag{8}$$

Например,

$$\varphi(40) = 40\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{5}\right) = 16;$$

$$\varphi(60) = 60\left(1 - \frac{1}{2}\right)\left(1 - \frac{1}{3}\right)\left(1 - \frac{1}{5}\right) = 16.$$

В своих «*Арифметических исследованиях*» Гаусс впервые приводит одно замечательное свойство функции $\varphi(m)$: сумма чисел $\varphi(\alpha)$, где α пробегает всевозможные делители числа m, равна самому m. Например, при $m = 12$ делителями m будут

числа 1, 2, 3, 4, 6, 12 и

$$\varphi(1)+\varphi(2)+\varphi(3)+\varphi(4)+\varphi(6)+\varphi(12)= \\ =1+1+2+2+2+4=12.$$

Общее доказательство можно дать, используя (8) или исходя непосредственно из определения функции $\varphi(m)$.

В (I, 5) мы уже ссылались на таблицу значений $\varphi(m)$ для всех $m \leqslant 10\,000$. В том же томе имеется таблица, в которой приводятся все значения m с данной величиной $\varphi(m)$ для всех $\varphi(m) \leqslant 2500$. Из этой таблицы видно, что при $\varphi(m) \leqslant 2500$ каждое значение функции $\varphi(m)$ принимается этой функцией не менее, чем дважды. Естественно предположить, что это верно и в общем случае, т. е. что *для любого натурального m найдется отличное от m число m', для которого $\varphi(m')=\varphi(m)$*. Но это утверждение не доказано, и всякая попытка дать общее доказательство встречает непреодолимые трудности. Для чисел некоторых частных видов результат прост; например, если m нечетно, то $\varphi(m)=\varphi(2m)$, если m не делится на 2 и на 3, то $\varphi(3m)=\varphi(4m)=\varphi(6m)$.

5. Теорема Вильсона. Эта теорема впервые была опубликована Варингом в его «*Алгебраических размышлениях*» (Meditationes Algebraicae) в 1770 году; он приписал ее сэру Джону Вильсону (1741—1793), юристу, изучавшему математику в Кэмбридже. Теорема утверждает, что для любого простого p имеет место сравнение

$$(p-1)! \equiv -1 \pmod{p}. \tag{9}$$

Следующее простое доказательство принадлежит Гауссу. Это доказательство основано на сопоставлении каждого из чисел $1, 2, \ldots, p-1$ с числом, обратным ему по mod p в смысле п. 2. Обратное к a — это такое число a', для которого $aa' \equiv 1 \pmod{p}$. Каждое число из ряда $1, 2, \ldots, p-1$ имеет в этом ряду точно одно обратное. Число, обратное к a, может быть равно a; тогда $a^2 \equiv 1 \pmod{p}$, т. е. $a \equiv \pm 1 \pmod{p}$, откуда следует, что $a = 1$ или $p-1$. Все остальные числа $2, 3, \ldots, p-2$ (кроме 1 и $p-1$) могут быть разбиты на пары, так что произведение чисел в каждой паре сравнимо с 1 по mod p. Следовательно,

$$2 \cdot 3 \cdot 4 \ldots (p-2) \equiv 1 \pmod{p}.$$

Умножив это сравнение на $p-1$, получим (так как $p-1 \equiv \equiv -1 \pmod p$) сравнение (9). Если p равно 2 или 3, то приведенное доказательство не проходит; в этих случаях, однако, результат устанавливается непосредственно.

Теорема Вильсона — одна из теорем, относящихся к симметрическим функциям чисел 1, 2, ..., $p-1$. Она устанавливает, что произведение этих чисел сравнимо с $-1 \pmod p$. Известны также некоторые результаты, касающиеся других симметрических функций. В качестве иллюстрации рассмотрим сумму k-х степеней

$$S_k = 1^k + 2^k + \ldots + (p-1)^k,$$

где p — простое, большее 2. Можно доказать, что если k не кратно $p-1$, то $S_k \equiv 0 \pmod p$. Если k делится на $p-1$, то каждое слагаемое суммы S_k по теореме Ферма сравнимо с 1; а так как в этой сумме $p-1$ слагаемых, то она сравнима с $p-1 \equiv -1 \pmod p$.

6. Алгебраические сравнения. По аналогии с теорией уравнений напрашивается рассмотрение алгебраических сравнений, т. е. сравнений вида

$$a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0 \equiv 0 \pmod m, \qquad (10)$$

где a_n, a_{n-1}, ..., a_0 — данные целые числа, а x — неизвестное. Прежде всего интересно, в какой мере теорию алгебраических уравнений можно распространить на алгебраические сравнения, ибо изучение алгебраических сравнений (в той или иной форме) является важной частью теории чисел.

Если степень сравнения n равна 1, то (10) принимает вид

$$a_1 x + a_0 \equiv 0 \pmod m;$$

это — линейное сравнение рассмотренного в п. 2 вида.

Если число x_0 удовлетворяет какому-нибудь алгебраическому сравнению по модулю m, то любое число x, сравнимое с x_0 по $\bmod m$, также удовлетворяет этому сравнению. Поэтому два сравнимых между собой решения можно рассматривать как одинаковые и *числом* решений сравнения по $\bmod m$ называть число решений этого сравнения в какой-нибудь полной системе вычетов по $\bmod m$, скажем в ряду 0, 1, ..., $m-1$. Например, сравнение $x^3 \equiv 8 \pmod{13}$ удовлетворяется при $x \equiv 2$, 5,

$6 \pmod{13}$ и только в этих случаях, поэтому оно имеет три решения.

Начнем с установления одного важного принципа, который позволяет свести определение числа решений алгебраического сравнения по произвольному модулю к подсчету числа решений в случае, когда модуль равен степени простого числа.

Чтобы убедиться в этом, допустим, что модуль m раскладывается в произведение $m_1 m_2$, в котором m_1 и m_2 взаимно просты. Алгебраическое сравнение

$$f(x) \equiv 0 \pmod{m} \tag{11}$$

удовлетворяется тогда и только тогда, когда удовлетворяется каждое из сравнений

$$f(x) \equiv 0 \pmod{m_1} \quad \text{и} \quad f(x) \equiv 0 \pmod{m_2}. \tag{12}$$

Если хотя бы одно из них неразрешимо, то неразрешимо и данное сравнение. Если они оба разрешимы, обозначим решения первого через

$$x \equiv \xi_1,\ x \equiv \xi_2,\ \ldots \pmod{m_1},$$

а решения второго через

$$x \equiv \eta_1,\ x \equiv \eta_2,\ \ldots \pmod{m_2}.$$

Каждое решение (11) порождает какое-нибудь ξ и какое-нибудь η. Обратно, если выбрать одно из ξ, скажем ξ_i, и одно из η, скажем η_j, то, как мы видели в предыдущем пункте, пара сравнений $x \equiv \xi_i \pmod{m_1}$ и $x \equiv \eta_j \pmod{m_2}$ эквивалентна ровно одному сравнению по модулю m. Следовательно, если $N(m)$ — число решений сравнения (11), а $N(m_1)$ и $N(m_2)$ обозначают количества решений двух сравнений (12), то

$$N(m) = N(m_1)N(m_2).$$

Другими словами, $N(m)$ является мультипликативной функцией от m. Если m разложено, как обычно, в произведение степеней простых, то

$$N(m) = N(p^a)N(q^b)\ldots \tag{13}$$

Таким образом, если известно число решений алгебраического сравнения по модулям, равным степеням простых, то благодаря мультипликативности легко найти и число решений этого сравнения по произвольному модулю. В частности, если хотя бы одно из чисел $N(p^a)$ равно нулю и p^a — входящая в m степень

простого, то сравнение неразрешимо (это, конечно, очевидно и само по себе).

Аналогичный результат имеет место для алгебраических сравнений от большего числа неизвестных. Число решений сравнения

$$f(x, y) \equiv 0 \pmod{m}$$

от двух неизвестных (и аналогично от любого числа неизвестных) — мультипликативная функция модуля.

7. Сравнения по простому модулю. Имеются два обстоятельства, в силу которых теория сравнений наибольшее внимание уделяет сравнениям по простому модулю.

Как мы только что видели, для определения числа решений сравнения достаточно рассматривать случай, когда модулем служит степень простого числа. Оказывается, что поведение сравнения по модулю p^a, являющемуся степенью простого, выводится обычно из его поведения в случае, когда модуль просто равен p. Такова первая причина, из-за которой теория сравнений по простому модулю имеет первостепенное значение.

Вторая причина в том, что арифметика по простому модулю, как это уже отмечалось в п. 2, особенно проста. В этой арифметике имеется p элементов, представляемых числами 0, 1, 2, ..., $p - 1$; здесь определены все четыре арифметических действия — сложение, умножение, вычитание и деление, кроме деления на 0. Первые три действия производятся как обычно, причем получающееся в результате число переводится в исходную совокупность прибавлением к нему или вычитанием из него соответствующего кратного p; последняя операция (деление) выполняется путем решения линейного сравнения.

Множество элементов произвольной природы, в котором определены операции, аналогичные четырем арифметическим действиям, удовлетворяющие тем же законам и выполнимые (кроме операции деления на нулевой элемент) внутри множества, называется *полем*. Самый известный пример поля — система рациональных чисел. Числа 0, 1, ..., $p - 1$, комбинируемые, как объяснено выше, также образуют поле; этот пример менее известен, но получающееся поле проще, ибо оно содержит конечное число элементов. Простейшим является случай $p = 2$.

В этом случае получается арифметика с двумя элементами. Если мы назовем их O и I (что соответствует 0 и 1), то получим такие правила действий:

$$O+O=O; \quad O+I=I; \quad I+O=I; \quad I+I=I;$$
$$O\cdot O=O; \quad O\cdot I=O; \quad I\cdot O=O; \quad I\cdot I=I.$$

Можно сказать, что эта арифметика представляет собой вырожденную форму обычной арифметики, в которой каждое четное число заменено на O, а каждое нечетное число — на I.

Часть теорем элементарной алгебры сохраняет силу для элементов любого поля. Одной из таких теорем является теорема о том, что алгебраическое уравнение степени n имеет не более n решений. В частности, эта теорема имеет место и в поле вычетов по mod p, где она принимает следующую форму: *сравнение степени n, скажем*

$$a_nx^n+a_{n-1}x^{n-1}+\ldots+a_1x+a_0\equiv 0 \pmod p, \qquad (14)$$

имеет не более n решений. Мы считаем, что старший коэффициент не сравним с $0 \pmod p$, ибо в противном случае соответствующее слагаемое можно было бы опустить.

Эту теорему впервые сформулировал и доказал Лагранж в 1768 году. Доказательство такое же, как и доказательство аналогичного факта для уравнений. Оно основано на том, что если x_1 является решением сравнения, то полином в левой части сравнения разлагается на множители и одним из них служит линейный полином $x-x_1$. Действительно, если x_1 удовлетворяет сравнению, то

$$a_nx_1^n+a_{n-1}x_1^{n-1}+\ldots+a_1x_1+a_0\equiv 0 \pmod p.$$

Если вычесть это сравнение из (14), то разность членов степени k будет иметь вид $a_k(x^k-x_1^k)$ при любом k из ряда 0, 1, …, $n-1$. Каждая такая разность содержит линейный множитель $x-x_1$. Таким образом, сравнение (14) может быть записано в виде

$$(x-x_1)(b_{n-1}x^{n-1}+b_{n-2}x^{n-2}+\ldots+b_0)\equiv 0 \pmod p,$$

где b_{n-1}, …, b_0 — некоторые целые числа, зависящие от a_n, …, a_0 и от x_1. Любое другое решение, скажем x_2, сравнения (14) должно (так как p простое) удовлетворять сравнению

$$b_{n-1}x^{n-1}+b_{n-2}x^{n-2}+b_0\equiv 0 \pmod p$$

и, значит, порождает множитель $x-x_2$ в стоящем здесь полино-

ме; так что в этом случае мы выделяем два линейных множителя исходного полинома. Это продолжается до тех пор, пока левая часть (14) полностью не разложится на множители или пока мы не придем к неразрешимому сравнению. В первом случае сравнение (14) имеет точно n решений *), во втором случае — меньше, чем n решений.

Простота модуля в теореме Лагранжа существенна. Например, сравнение $x^2 - 1 \equiv 0 \pmod 8$ степени 2 имеет 4 решения $x \equiv 1, 3, 5, 7 \pmod 8$. (Это сравнение выполняется для любого нечетного числа.)

Мы видели, что каждое решение алгебраического сравнения отвечает некоторому линейному множителю соответствующего полинома. Можно рассматривать и более общий вопрос о разложении полинома с приведенными по модулю p коэффициентами на другие полиномы. Легко видеть, что любой полином $f(x)$ можно разложить на *неприводимые* полиномы, т. е. на полиномы, которые далее не разлагаются. Другими словами, для всякого полинома $f(x)$ существуют неприводимые полиномы $f_1(x)$, $f_2(x)$, ..., $f_r(x)$ такие, что

$$f(x) \equiv f_1(x) f_2(x) \dots f_r(x) \pmod p$$

тождественно по x. Здесь, конечно, идет речь о неприводимости *по отношению* к данному простому p. Все линейные полиномы, встречающиеся в разложении, отвечают решениям сравнения $f(x) \equiv 0 \pmod p$; если линейных множителей нет, то сравнение неразрешимо. Вот два примера разложения на неприводимые полиномы:

$$x^4 + 3x^2 + 3 \equiv (x-1)(x+1)(x^2-3) \pmod 7,$$
$$x^4 + 2x^3 - x^2 + 2 \equiv (x^2+x+1)(x^2+x+2) \pmod 5,$$

Возникает вопрос, единственно ли это разложение. Очевидно, что полиномы $f_1(x)$, ..., $f_r(x)$ можно умножить на числа, произведение которых сравнимо с 1 по mod p, не изменив значения произведения $f_1(x) \cdot \ldots \cdot f_r(x)$. С другой стороны, можно доказать, что с точностью до этой возможности *разложение единственно*. Рассматриваемая теория очень похожа на теорию разложения натуральных чисел на простые. Здесь снова важную роль играет алгоритм Евклида, основанный на процессе

*) Здесь каждое решение x_1 засчитывается m раз, если множитель $x - x_1$ входит в разложение левой части (14) m раз. (*Прим. перев.*)

деления одного полинома на другой с остатком (степень остатка должна быть меньше степени делителя). Недостаток места препятствует более подробному изложению этой теории.

8. Сравнения от нескольких переменных. Очень простая и общая теорема, принадлежащая Шевалле, устанавливает разрешимость широкого класса сравнений от нескольких переменных. Пусть $f(x_1, \ldots, x_n)$ — произвольный полином от n переменных, не обязательно однородный, степень которого меньше n, а свободный член равен 0. Под *степенью* многочлена понимается наивысшая из степеней отдельных членов, а степень одночлена типа $x_1 x_2^3 x_3^4$ считается равной $1 + 3 + 4 = 8$. Теорема Шевалле утверждает, что существует такое решение сравнения

$$f(x_1, \ldots, x_n) \equiv 0 \pmod{p}, \tag{15}$$

в котором не все неизвестные сравнимы с нулем.

Прежде чем приводить доказательство, сделаем одно предварительное замечание. При каких условиях сравнение, скажем

$$\varphi(x_1, x_2, \ldots, x_n) \equiv 0 \pmod{p},$$

выполняется при *всех* целых $x_1, \ldots, x_n$? По теореме Ферма (п. 3) $x^p \equiv x \pmod{p}$ для всех x. Поэтому, не изменяя значения произвольного сравнения, можно в каждом его члене заменить все показатели на числа ряда $1, \ldots, p-1$ (вычитая из этих показателей соответствующее кратное $p-1$). Если сделать это, то получившееся сравнение будет выполняться при всех $x_1, x_2, \ldots, x_n$, только если оно сводится к тождеству, т. е. если все его коэффициенты будут сравнимы с нулем. Действительно, в силу теоремы Лагранжа такое сравнение, имея по x_1 степень не выше $p-1$, может иметь не более $p-1$ решений для x_1 если только не все его коэффициенты (когда оно рассматривается как полином от x_1) сравнимы с нулем. Эти коэффициенты являются полиномами от $x_2, \ldots, x_n$ не выше $(p-1)$-й степени по каждой из неизвестных, и мы можем применить к этим полиномам то же рассуждение. Общее утверждение получается повторением приведенного аргумента.

Теорема Шевалле доказывается сведением сравнения (15), которое предполагается неразрешимым при отличных от нуля значениях неизвестных, к другому сравнению, разрешимому для всех значений неизвестных. Рассмотрим сравнение

$$1-[f(x_1,\ldots,x_n)]^{p-1}\equiv(1-x_1^{p-1})\ldots(1-x_n^{p-1})\pmod p. \quad (16)$$

Если $x_1, \ldots, x_n$ все сравнимы с нулем, то обе части этого сравнения сравнимы с 1. Если какое-нибудь из $x_1, \ldots, x_n$ не сравнимо с нулем, то и левая, и правая части сравнимы с нулем по теореме Ферма. Значит, по предположению, которое нужно опровергнуть, (16) имеет место для всех целых $x_1, \ldots, x_n$. Как мы видели, это соотношение должно свестись к тождеству, если, выписав все его члены, каждый показатель каждой из переменных заменить на одно из значений $1, 2, \ldots, p-1$, вычитая подходящее кратное $p-1$. Справа такая редукция невозможна, ибо показатель каждой из переменных в любом члене не превосходит $p-1$. Слева, быть может, редукция и возможна. Но полная степень каждого члена в левой части по предположению меньше, чем $(p-1)n$; редукция показателей может ее лишь уменьшить. Отсюда ясно, что это соотношение нельзя свести к тождеству; действительно, ни одно слагаемое в левой части сравнения не может иметь такую высокую степень, как слагаемое $x_1^{p-1}x_2^{p-1}\ldots x_n^{p-1}$ в правой части. Это и доказывает теорему.

В качестве простого примера рассмотрим сравнение

$$x^2+y^2+z^2\equiv 0\pmod p.$$

Полином в левой части этого сравнения зависит от трех переменных x, y, z имеет степень 2 и не имеет свободного члена, так что выполняются все условия теоремы Шевалле. Следовательно, это сравнение разрешимо с не сравнимыми с нулем x, y, z. Этот частный результат полезен в связи с представлением числа суммой четырех квадратов (V, 4); но его можно легко доказать и без применения теоремы Шевалле.

9. Сравнения, покрывающие все числа. Любопытна задача построения такого ряда сравнений по различным модулям, что каждое число удовлетворяет по крайней мере одному из этих сравнений. Такую совокупность сравнений можно назвать покрывающим множеством. Модуль 1 должен быть, конечно, исключен. Сравнения

$$x\equiv 0\pmod 2,\ 0\pmod 3,\ 1\pmod 4,\ 1\pmod 6,\ 11\pmod{12}$$

образуют покрывающее множество. В самом деле, первые два из них покрывают все числа, за исключением сравнимых с 1,

5, 7, 11 (mod 12). Из этих чисел 1 и 5 покрываются сравнением $x \equiv 1 \pmod 4$, 7 покрываются сравнением $x \equiv 1 \pmod 6$, а 11 покрываются последним сравнением.

П. Эрдёш (P. Erdös) поставил следующую задачу: дано произвольное число N; существует ли множество покрывающих сравнений, использующих лишь модули, бóльшие N? Возможно, это верно, но доказательство найти нелегко. Сам Эрдёш построил множество, не использующее модуля 2; модулями у него являются различные множители 120. Д. Свифт (D. Swift) построил множество, для которого наименьшим модулем является 4; в качестве модулей здесь фигурируют различные сомножители 2880.

Замечания к главе II. п. 3. Обычная фраза — «x принадлежит показателю l по модулю m» — в этом контексте излишне громоздка.

п. 4. Число $[\alpha, \beta]$, введенное для представления совместного решения сравнений (7), выражается по следующей формуле. Определим a' и b' так, что $aa' \equiv 1 \pmod b$ и $bb' \equiv 1 \pmod a$, тогда $[\alpha, \beta] \equiv aa'\beta + bb'\alpha \pmod{ab}$.

п. 5. Теорему Вильсона можно обобщить на случай составного модуля; см. ([9], § 8.8) или ([13], p. 266).

Обычное доказательство того, что $S_k \equiv 0 \pmod p$, использует примитивные корни — см., например, ([9], § 7.10), но можно также дать и более непосредственные доказательства этого утверждения. Обширную литературу о симметрических функциях от чисел 1, 2, ..., $p-1$ можно найти в книге Диксона ([7], том I, гл. 3).

п. 7. Полное определение всех типов полей, состоящих из конечного числа элементов, было дано американским математиком E. X. Moope (E. H. Moore) в 1893 году. Число элементов конечного поля должно быть непременно степенью простого p^n; само же конечное поле является полем вычетов по mod p (при $n = 1$) или его алгебраическим расширением. Об этой теории см. Dickson, Linear Groups, ch. 1, или MacDuffee, Introduction to Abstract Algebra, pp. 174—180, или Birkhoff and MacLane, Survey of Modern Algebra, pp. 428—431, или Б. Л. Ван-дер-Варден, Современная алгебра *), М.—Л., 1947, стр. 153—157. Некоторые таблицы неприводимых полиномов для первых четырех простых модулей можно найти в статье R. Church, Annals of Math., **36** (1935), 198—209.

п. 8. О теореме Шевалле см. работу в журнале Abhandlungen Math. Seminar, Hamburg **11** (1936), 73—75.

*) Эта книга добавлена мною. (*Прим. перев.*)

ГЛАВА III

КВАДРАТИЧНЫЕ ВЫЧЕТЫ

1. Первообразные корни. В этой главе мы будем исследовать алгебраические сравнения по простому модулю, содержащие только два члена, т. е. кроме константы один член. Такое *двучленное* сравнение можно записать в виде

$$ax^k \equiv b \pmod{p},$$

где степень сравнения k положительна. Если a' обозначает число, обратное к a по модулю p, так что $aa' \equiv 1 \pmod{p}$, то, умножая обе части записанного выше сравнения на a', получим

$$x^k \equiv a'b \pmod{p}.$$

Мы можем поэтому привести любое двучленное сравнение к простейшему виду:

$$x^k \equiv c \pmod{p}. \tag{1}$$

Число, для которого сравнение (1) разрешимо, называется *k-степенным вычетом по модулю p*; аналогично, если это сравнение неразрешимо, говорят, что c есть *k-степенной невычет*. (Удобно, однако, числа c, сравнимые с $0 \pmod{p}$, не относить к k-степенным вычетам, хотя в этом случае сравнение и разрешимо.) Если k равно 2, мы получаем квадратичные вычеты и невычеты; так как в этом случае теория может быть развита дальше, чем в общем случае, то в настоящей главе будет рассматриваться именно эта возможность.

Чтобы проиллюстрировать это определение, положим p равным 13, а k равным 2 или 3. Значения x^2 и x^3 по модулю 13 таковы:

x:	1	2	3	4	5	6	7	8	9	10	11	12
x^2:	1	4	9	3	12	10	10	12	3	9	4	1
x^3:	1	8	1	12	8	8	5	5	1	12	5	12

Таким образом, по модулю 13 числа 1, 3, 4, 9, 10, 12 — квадратичные вычеты, а оставшиеся числа 2, 5, 6, 7, 8, 11 — квадратичные невычеты. Числа 1, 5, 8, 12 — кубические вычеты, а оставшиеся числа 2, 3, 4, 6, 7, 9, 10, 11 — кубические невычеты.

Теория k-степенных вычетов и невычетов связана с понятием *порядка* числа по модулю p, который был определен в (II, 3). Порядок любого числа a (предполагается, что a не сравнимо с 0) — это наименьшее натуральное число, для которого $a^l \equiv 1 \pmod{p}$. Мы доказали, что l является делителем $p-1$, а в примере с $p = 11$ нашли, что порядок 2 точно равен $p-1$. Эйлер первым сформулировал теорему о том, что *для любого простого числа p найдется число, порядок которого равен $p-1$*, и назвал такое число *первообразным корнем* для простого p. Но его доказательство существования первообразного корня было неудовлетворительным; первое удовлетворительное доказательство было дано Лежандром. К изложению этого доказательства мы теперь и приступаем.

Первый шаг доказательства состоит в установлении одного общего принципа, касающегося порядка произведения двух чисел. Если число a имеет порядок l, а число b — порядок k, *причем* l и k взаимно просты, то произведение ab имеет порядок lk. Конечно, число ab, возведенное в степень lk, дает $1 \pmod{p}$, потому что

$$(ab)^{lk} \equiv (a^l)^k (b^k)^l \equiv 1 \pmod{p}$$

(так как $a^l \equiv 1 \pmod{p}$ и $b^k \equiv 1 \pmod{p}$). Этот факт не зависит от взаимной простоты a и b, но он показывает лишь, что порядок ab является делителем lk. Может быть, этот порядок — собственный делитель lk; мы должны доказать, что на самом деле это не так. Предположим, что порядок ab равен $l_1 k_1$, где l_1 — делитель l, а k_1 — делитель k. Тогда

$$a^{l_1 k_1} b^{l_1 k_1} \equiv 1 \pmod{p}.$$

Возведем обе части этого сравнения в степень l_2, где $l_1 l_2 = l$. Так как $a^l \equiv 1$, то мы получаем, что $b^{lk_1} \equiv 1$. Из этого следует, что lk_1 кратно порядку b, который равен k. Но l взаимно просто с k, значит k_1 кратно k; будучи также делителем k, k_1 должно быть равно k. Аналогично $l_1 = l$ и, таким образом, порядок ab в точности равен lk.

Этот принцип дает возможность строить первообразный корень шаг за шагом. Пусть $p-1$ разложено в произведение степеней простых, скажем

$$p-1=q_1^{a_1}q_2^{a_2}\ldots \tag{2}$$

Допустим, что мы нашли число x_1 порядок которого равен $q_1^{a_1}$, число x_2, порядок которого равен $q_2^{a_2}$, и так далее. Тогда, применив несколько раз только что доказанный принцип, мы убедимся, что произведение найденных чисел имеет порядок $p-1$ и, значит, является первообразным корнем. Поэтому остается лишь доказать, что *если q^a — одна из степеней простых, составляющих $p-1$, то существует число, порядок которого по* $\bmod p$ *в точности равен q^a.*

Число, порядок которого равен q^a, должно удовлетворять сравнению

$$x^{q^a}\equiv 1 \pmod p. \tag{3}$$

Но число, удовлетворяющее этому сравнению, не обязано иметь порядок q^a; его порядок может быть любым делителем q^a, т. е. он может быть равен 1 или q, или q^2 и так далее до q^{a-1}. Но если порядок не равен q^a, то он является делителем q^{a-1}, и число x должно удовлетворять сравнению

$$x^{q^{a-1}}\equiv 1 \pmod p. \tag{4}$$

Следовательно, нам нужно найти число, удовлетворяющее сравнению (3), но не удовлетворяющее сравнению (4).

Мы можем доказать существование такого числа, подсчитав, сколько решений имеют эти сравнения. Конечно, по теореме Лагранжа сравнение (3) имеет не более q^a решений, а сравнение (4) имеет не более q^{a-1} решений. Само по себе это не помогло бы нам, но, к счастью, можно доказать, что эти сравнения имеют *в точности* q^a и q^{a-1} решений. Отсюда уже следует, что имеется q^a-q^{a-1} чисел, удовлетворяющих (3) и не удовлетворяющих (4), а так как $q^a>q^{a-1}$, то это дает нужный результат и заканчивает доказательство.

Мы рассмотрим более общий случай, именно, рассмотрим сравнение

$$x^d-1\equiv 0 \pmod p,$$

где d — любой делитель $p-1$. По теореме Лагранжа это срав-

нение имеет не более d решений; мы должны доказать, что оно имеет точно d решений. Доказательство опирается на то, что полином $x^d - 1$ делит полином $x^{p-1} - 1$. Если мы напишем y вместо x^d и положим $p - 1 = dl$, то получим

$$x^{p-1} - 1 = y^l - 1 = (y - 1)(y^{l-1} + y^{l-2} + \ldots + y + 1).$$

Так как $y - 1 = x^d - 1$, это дает тождество вида

$$x^{p-1} - 1 = (x^d - 1)f(x),$$

где $f(x)$ — некоторый полином от x степени $p - 1 - d$. Но сравнение $x^{p-1} - 1 \equiv 0 \pmod{p}$ имеет $p - 1$ решений: оно удовлетворяется для всех x, не сравнимых с 0 (II, 3). Все $p - 1$ решений должны удовлетворять одному из сравнений

$$x^d - 1 \equiv 0 \pmod{p} \quad \textit{или} \quad f(x) \equiv 0 \pmod{p}.$$

Последнее из них по теореме Лагранжа имеет не более $p - 1 - d$ решений, поэтому первое должно иметь *не менее* d решений, а значит, оно имеет *ровно* d решений. Взяв d равным q^a или q^{a-1}, получаем то, что нам осталось доказать.

Мы проиллюстрируем доказательство, взяв $p = 19$. Имеем $p - 1 = 2 \cdot 3^2$. Найдем прежде всего число x_1 порядка 2, т. е. число x, для которого $x^2 \equiv 1$ и $x \not\equiv 1$. Очевидно, что x_1 должно быть равно -1 или (что то же самое) 18. Найдем, далее, число x_2 порядка 9, т. е. число, удовлетворяющее сравнениям $x^9 \equiv 1$ и $x^3 \not\equiv 1$. Можно установить, что решениями $x^9 \equiv 1 \pmod{19}$ являются числа 1, 4, 5, 6, 7, 9, 11, 16, 17. Из них числа 1, 7, 11 должны быть выкинуты, ибо для них будет выполняться сравнение $x^3 \equiv 1$. Остается шесть возможностей для x_2 это соответствует тому, что в общем случае имеется $q^a - q^{a-1}$ выборов. Умножая эти оставшиеся числа на x_1, получаем первообразные корни -4, -5, -6, -9, -16, -17 или 2, 3, 10, 13, 14, 15. Чтобы установить, что 2 есть первообразный корень, найдем последовательные степени 2 по модулю 19, ими являются числа: 2, 4, 8, 16, 13, 7, 14, 9, 18, 17, 15, 11, 3, 6, 12, 5, 10, 1; впервые число 1 встречается на восемнадцатом месте. Вышеуказанный метод практически не очень удобен для нахождения первообразного корня: много проще испытывать числа 2, 3, ... подряд. Но это, конечно, не приводит к общему доказательству существования первообразного корня.

Перемножив соответствующие количества для x_1, x_2, ...,

легко заметить, что получается не более

$$(q_1^{a_1}-1)(q_2^{a_2}-1)\dots$$

возможных значений для x_1, x_2, ... Найденные таким способом первообразные корни действительно различны, и так получаются все первообразные корни, но мы не будем останавливаться на доказательстве этих фактов. Число первообразных корней совпадает с указанным выше произведением, которое равно $\varphi(p-1)$ (в силу равенства (8) главы II). Например, когда $p = 19$, имеется $\varphi(18) = 6$ первообразных корней.

2. Индексы. Существование первообразного корня имеет не только теоретический интерес, но и дает новое средство для вычислений по простому модулю p. Эти вычисления аналогичны вычислениям с помощью логарифмов в обычной арифметике.

Пусть g — первообразный корень по mod p. Тогда числа

$$g,\ g^2,\ \dots,\ g^{p-1}\ (\equiv 1) \tag{5}$$

не сравнимы между собой, так как g^{p-1} — *первая* степень g, сравнимая с 1. Ни одно из этих чисел не сравнимо также с 0. Поэтому они должны быть сравнимы с числами 1, 2, ..., $p-1$ в некотором порядке. Пример в предыдущем пункте иллюстрирует этот факт: степени 2 от 2^1 до 2^{18} ($\equiv 1$) сравнимы с 1, 2, ..., 18 по модулю 19 (в другом порядке).

Любое число, не сравнимое с 0 по mod p, сравнимо поэтому с одним из чисел ряда (5). Если $a \equiv g^\alpha \pmod p$, мы говорим, что α есть *индекс* (относительно примитивного корня g). Если a дано, то среди чисел 0, 1, ..., $p-1$ единственным способом определяется α. Но брать α среди этих чисел необязательно. Если α' есть какое-нибудь другое число, для которого $a \equiv g^{\alpha'}$, можем свести α' к числу упомянутого ряда, прибавляя или вычитая подходящее кратное $p-1$; это не изменит значения $g^{\alpha'}$, так как $g^{p-1} \equiv 1$. Приведенное значение α' должно быть равно α, поэтому $\alpha' \equiv \alpha \pmod{p-1}$.

Если $p = 19$ и $g = 2$, то индексы чисел 1, ..., 18 таковы:

Число:	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Индекс:	18	1	13	2	16	14	6	3	8	17	12	15	5	7	11	4	10	9

Чтобы составить такую таблицу, мы сопоставляем индекс 1 первообразному корню (здесь 2), индекс 2 его квадрату (здесь 4) и так далее, вычисляя степени первообразного корня по модулю p (здесь 19). Таблицу индексов для всех простых, меньших 1000, в 1839 году опубликовал Якоби (под названием Canon Arithmeticus).

С помощью индексов действие умножения по mod p можно свести к действию сложения так же, как, используя логарифмы, можно свести обычное умножение к сложению (если ограничиться умножением положительных чисел). Если a и b — данные числа, α и β — их индексы, то $a \equiv g^\alpha$, $b \equiv g^\beta$, откуда $ab \equiv g^{\alpha+\beta}$ (все сравнения берутся по модулю p). Следовательно, индекс произведения ab равен $\alpha + \beta$ или отличается от него на кратное $p-1$. Таким образом, чтобы перемножить два числа, находят в таблице их индексы, складывают эти индексы, затем переносят результат в ряд $0, 1, \ldots, p-1$, вычитая из него, если понадобится, кратное $p-1$, и, наконец, находят в таблице число с вычисленным индексом. Например, чтобы найти значение $10 \cdot 12 \pmod{19}$, мы из написанной выше таблицы берем соответствующие индексы: 17 и 15; сумма этих индексов, равная 32, приводится к 14 после вычитания $18 = p-1$; число с индексом 14 равно 6 — это и есть ответ. Тем же способом можно выполнять и деление по (mod p); для этого надо заменить сложение индексов вычитанием.

Использование индексов дает возможность исследовать структуру k-степенных вычетов и невычетов по mod p. Мы хотим выяснить, разрешимо сравнение

$$x^k \equiv a \pmod{p} \tag{6}$$

или нет. Если индексом x служит ξ, то индекс x^k будет равен $k\xi$ или отличается от $k\xi$ на кратное $p-1$. Поэтому записанное выше сравнение эквивалентно следующему:

$$k\xi \equiv \alpha \pmod{p-1}, \tag{7}$$

где α — индекс a. Это — линейное сравнение от неизвестного ξ по модулю $p-1$.

Если k взаимно просто с $p-1$, ситуация не сложна: линейное сравнение (7) однозначно разрешимо относительно ξ, поэтому

сравнение (6) имеет единственное решение для x. В этом случае каждое число является k-степенным вычетом и однозначно представимо в виде k-й степени. Другими словами, если k взаимно просто с $p-1$, то числа

$$1^k,\ 2^k,\ 3^k,\ \ldots,\ (p-1)^k$$

сравнимы с числами $1, 2, \ldots, p-1$, взятыми в каком-то другом порядке. Например, если $p = 19$ и $k = 5$, числа $1^5, 2^5, \ldots, 18^5$ сравнимы по $\bmod 19$ с 1, 13, 15, 17, 9, 5, 11, 12, 16, 3, 7, 8, 14, 10, 2, 4, 6, 18.

Совершенно другая ситуация возникает, если k и $p-1$ имеют общий множитель. Рассмотрим сначала один частный случай, скажем, $p = 19$ и $k = 3$. Сравнение (7) в этом случае имеет вид

$$3\xi \equiv \alpha \pmod{18}.$$

Это сравнение, очевидно, неразрешимо, если α не делится на 3. Если α делится на 3, скажем $\alpha \equiv 3\beta$, последнее сравнение принимает вид $\xi \equiv \beta \pmod 6$. Это дает одно значение для ξ по модулю 6, но три значения по модулю 18 (число 18 служит модулем, по которому определяется ξ), именно: β, $\beta+6$, $\beta+12$, если β — одно из решений. Таким образом, если α делится на 3, то число a сравнимо с тремя различными кубами. Взяв таблицу индексов по модулю 19, мы увидим, что числа, индексы которых делятся на 3, равны 1, 7, 8, 11, 12, 18. Если a — одно из этих чисел, то сравнение $x^3 \equiv a \pmod p$ имеет точно три решения. Эти числа являются кубическими вычетами по $\bmod 19$, а оставшиеся 12 чисел — кубическими невычетами.

Общий случай исследуется аналогично. Пусть K — наибольший общий делитель k и $p-1$. Сравнение (7) неразрешимо относительно ξ, если α не делится на K, так как k и модуль делятся на K. С другой стороны, если α делится на K, сравнение (7) разрешимо относительно ξ и имеет ровно K решений. Таким образом, *k-степенными вычетами по* $\bmod p$ *являются как раз те числа, индексы которых делятся на K, наибольший общий делитель k и $p-1$*. Если a есть k-степенной вычет, сравнение (6) имеет ровно K решений.

Число k-степенных вычетов равно $\frac{p-1}{K}$, так как индексами могут быть числа $1, 2, \ldots, p-1$ и в точности $\frac{1}{K}$ часть этих чисел делится на K.

Простейшим является случай $k = 2$; в этом случае мы имеем дело с квадратичными вычетами и невычетами. Если предположить, что $p > 2$, то $p-1$ четно и наибольший общий делитель 2 и $p-1$ равен 2. В этом случае заключение таково: *квадратичные вычеты суть числа с четными индексами, а квадратичные невычеты — числа с нечетными индексами. Число тех и других одинаково и равно* $\frac{1}{2}(p-1)$. Если a — какой-нибудь квадратичный вычет, наша теория устанавливает, что сравнение $x^2 \equiv a \pmod{p}$ имеет точно два решения. Ясно, что если $x \equiv x_1$ — одно из решений, то $x \equiv -x_1$ — другое решение.

Если $p = 19$, то числа 1, 4, 5, 6, 7, 9, 11, 16, 17 являются квадратичными вычетами, а числа 2, 3, 8, 10, 12, 13, 14, 15, 18 — квадратичными невычетами.

3. Квадратичные вычеты. В оставшейся части этой главы мы ограничимся теорией квадратичных вычетов и невычетов, которая может быть развита существенно дальше, чем общая теория k-степенных вычетов. Будем далее предполагать, что p — простое число, не равное 2.

Как мы уже видели, половина из чисел $1, 2, \ldots, p-1$ — квадратичные вычеты, другая половина — квадратичные невычеты. Квадратичные вычеты сравнимы с числами $1^2, 2^2, \ldots, \left(\frac{1}{2}(p-1)\right)^2$, ибо оставшиеся числа от $\frac{1}{2}(p+1)$ до $p-1$ после возведения в квадрат дают тот же результат:

$$(p-x)^2 \equiv x^2 \pmod{p}.$$

Квадратичные вычеты и невычеты обладают простым мультипликативным свойством. Произведение двух вычетов или двух невычетов является вычетом, а произведение вычета и невычета является невычетом. Это сразу следует из того, что вычеты имеют четные индексы, а невычеты — нечетные индексы: сумма двух четных или двух нечетных индексов четна, сумма же четного и нечетного индекса нечетна. Таким образом, например, в таблице квадратичных вычетов и невычетов для простого числа 19 в конце п. 2 произведение любых двух чисел, взятых из одной строки, сравнимо с некоторым числом из первой строки, а произведение двух чисел, взятых из разных строк, сравнимо с числом из второй строки.

Несомненно, что именно это мультипликативное свойство и навело Лежандра на мысль ввести символ квадратичного характера числа a по простому модулю p. Символ Лежандра определяется так:

$$\left(\frac{a}{p}\right) = \begin{cases} 1, & \text{если } a \text{ — квадратичный вычет по mod } p, \\ -1, & \text{если } a \text{ — квадратичный невычет по mod } p. \end{cases}$$

Иногда мы будем использовать также обозначение $(a|p)$. Другая форма этого определения такова: $(a|p) = (-1)^\alpha$, где α равно индексу a. Отмеченное выше мультипликативное свойство принимает вид

$$\left(\frac{ab}{p}\right) = \left(\frac{a}{p}\right)\left(\frac{b}{p}\right).$$

Любое число a (не сравнимое с 0) удовлетворяет сравнению Ферма $a^{p-1} \equiv 1 \pmod p$. Так как $p-1$ — четно, это сравнение раскладывается на множители; полагая $p-1 = 2P$, можно сказать, что для каждого числа a либо $a^P \equiv 1 \pmod p$, либо $a^P \equiv -1 \pmod p$. Эйлер, вероятно, первым заметил, что эти две возможности имеют место в зависимости от того, является a квадратичным вычетом или a есть квадратичный невычет. В нашем рассмотрении это доказывается тотчас же. Если α есть индекс a, то $a^P \equiv q^{\alpha P} \pmod p$. Если α четно, то αP кратно $p-1$ и $g^{\alpha P} \equiv 1 \pmod p$. Если α нечетно, то $\alpha P = \frac{1}{2}(p-1)\alpha$ не кратно $p-1$ и $g^{\alpha P}$ не сравнимо с 1, а следовательно, сравнимо с -1. Этот факт носит название *критерия Эйлера* для квадратичного характера a. В терминах символа Лежандра он принимает вид

$$\left(\frac{a}{p}\right) \equiv a^P \pmod p, \tag{8}$$

где $P = \frac{1}{2}(p-1)$.

Критерий Эйлера сам по себе не приносит большой пользы при исследовании свойств квадратичных вычетов и невычетов, но из него тотчас же вытекает правило вычисления квадратичного характера от -1. Величина $(-1)^P$ будет равна 1 или -1 в зависимости от того, четно P или нет, т. е. в зависимости от того, имеет p вид $4k+1$ или $4k+3$. Отсюда следует, что -1 *есть квадратичный вычет для простых вида* $4k+1$ *и квадратичный невычет для простых вида* $4k+3$. Это означает, что

для простых вида $4k+1$ ряды квадратичных вычетов и невычетов симметричны, так как значения характера от $p-a$ и от a одинаковы. Действительно, $p-a \equiv -a$ и $(-a|p) = (-1|p)(a|p) = (a|p)$. С другой стороны, если p имеет вид $4k+3$, характер $p-a$ противоположен по знаку характеру a, что можно было заметить в случае $p=19$ (в конце п. 2).

Тот факт, что сравнение $x^2+1 \equiv 0 \pmod{p}$ разрешимо для простых вида $4k+1$ и неразрешимо для простых вида $4k+3$, был известен Ферма. После нескольких безуспешных попыток доказать этот факт Эйлер в 1749 году нашел доказательство; свой критерий он установил в 1755 году.

Лагранж в 1773 году заметил, что если рассматриваемое сравнение разрешимо, то существует очень простой способ точно указать его решения. При $p=4k+1$ теорема Вильсона (II, 5) показывает, что $1 \cdot 2 \cdot 3 \cdot \ldots \cdot 4k \equiv -1 \pmod{p}$.

Но

$$\begin{aligned} 4k &\equiv -1 \pmod{p}, \\ 4k-1 &\equiv -2 \pmod{p}, \\ &\ldots\ldots\ldots\ldots \\ 2k+1 &\equiv -2k \pmod{p}. \end{aligned}$$

Подставляя эти значения, мы получаем

$$(1 \cdot 2 \cdot 3 \cdot \ldots \cdot 2k)^2 \equiv -1 \pmod{p},$$

так как число введенных отрицательных знаков равно $2k$ и является четным. Поэтому решениями сравнения $x^2 \equiv -1 \pmod{p}$ являются числа $x \equiv \pm(2k)!$, где $p=4k+1$. Например, если $p=13$, так что $k=3$, то решениями служат $x \equiv \pm 6! \equiv \pm 720 \equiv \pm 5 \pmod{13}$. Эта конструкция, конечно, бесполезна для вычислений, но всегда интересно в дополнение к доказательству существования иметь и точную конструкцию.

4. Лемма Гаусса. Более глубокие свойства квадратичных вычетов и невычетов, в особенности свойства, связанные с законом взаимности (п. 5), были открыты эмпирически и впервые доказывались непрямыми и запутанными методами. Не ранее 1808 года (через семь лет после опубликования своих «*Исследований*») Гаусс открыл одну простую лемму, дающую ключ к ясному и элементарному доказательству закона взаимности.

Лемма Гаусса дает правило для вычисления квадратичного характера числа a (не сравнимого с 0) по отношению к простому p. Как всегда, мы предполагаем $p > 2$ и берем $P = \frac{1}{2}(p-1)$. Правило Гаусса предписывает образовать числа

$$a,\ 2a,\ 3a,\ \ldots,\ Pa \tag{9}$$

и добиться, вычитая соответствующее кратное p, чтобы каждое из них лежало между $-\frac{1}{2}p$ и $\frac{1}{2}p$. Пусть ν — число отрицательных в образованном ряду чисел. Тогда $(a|p) = (-1)^\nu$, т. е. *a является квадратичным вычетом, если ν четно, и квадратичным невычетом, если ν нечетно.* Доказательство этого факта совсем просто. Правило предписывает заменить каждое из чисел в ряду (9) сравнимым с ним числом из ряда ±1, ±2, ..., $\pm P$, что мы, очевидно, можем сделать. Если мы сделаем это, то ни одно число из ряда 1, 2, ..., P не встретится более одного раза ни с плюсом, ни с минусом. Действительно, если бы какое-нибудь число встретилось дважды с одинаковым знаком, то в ряду (9) нашлись бы два числа, сравнимые друг с другом по mod p, чего нет; если же какое-нибудь число встретилось бы с противоположными знаками, то сумма двух чисел из ряда (9) была бы сравнима с нулем по mod p, чего также не может быть. Поэтому получающийся ряд состоит из чисел ±1, ±2, ..., $\pm P$, причем каждому приписан *определенный знак.* Перемножая два ряда, получаем

$$(a)(2a)(3a)\ldots(Pa) \equiv (\pm1)(\pm2)\ldots(\pm P) \pmod{p}.$$

После сокращения на 2, 3, ..., P отсюда следует, что

$$a^P \equiv (\pm1)(\pm1)\ldots(\pm1) = (-1)^\nu,$$

где ν — число знаков минус. В силу критерия Эйлера отсюда немедленно получается нужный результат. Чтобы проиллюстрировать лемму Гаусса численно, возьмем $p = 19$ и $a = 5$. Здесь $P = 9$, и мы должны заменить числа 5, 10, 15, ..., 45 по mod 19 так, чтобы они лежали между -9 и 9 включительно. Получаются числа 5, -9, -4, 1, 6, -8, -3, 2, 7. Как и в общем случае, этими числами являются числа от 1 до 9, каждое со своим знаком. Число отрицательных знаков равно 4, и так как 4 четно, 5 является квадратичным вычетом по mod 19, или, символически: $(5|19) = 1$.

С помощью леммы Гаусса можно дать простое правило для нахождения квадратичного характера числа 2. Когда $a = 2$, ряд чисел в (9) таков: 2, 4, 6, ..., $2P$ и $2P = p-1$. Мы должны определить, сколько чисел в этом ряду после попадания в интервал между $-\frac{1}{2}p$ и $\frac{1}{2}p$ станут отрицательными. Так как рассматриваемые числа расположены между 0 и p, то отрицательными становятся числа, большие $\frac{1}{2}p$. Поэтому мы должны найти, сколько чисел вида $2x$ удовлетворяет неравенству $\frac{1}{2}p < 2x < p$; другими словами, сколько найдется таких целых x, для которых $\frac{1}{4}p < 2x < \frac{1}{2}p$. Положим $p = 8k + r$, где r равно 1, 3, 5 или 7. В этих обозначениях наше условие имеет вид $2k+\frac{1}{4}r < x < 4k+\frac{1}{2}r$; мы хотим узнать, четно или нет количество чисел x, удовлетворяющих этому условию. Четность количества этих чисел не нарушится, если вычесть из обеих частей неравенства четные числа $2k$ и $4k$. Поэтому достаточно рассматривать неравенство $\frac{1}{4}r < x < \frac{1}{2}r$. Это неравенство не имеет решения, если r равно 1; имеет одно решение при r, равном 3 или 5; и два решения, если r равно 7. Следовательно, 2 является квадратичным вычетом в первом и последнем случаях и невычетом в двух других случаях. Таким образом, имеет место следующее правило: 2 *есть квадратичный вычет для простых вида* $8k \pm 1$ *и квадратичный невычет для простых вида* $8k \pm 3$. Этот факт был известен Ферма, но доказали его впервые, причем очень сложным способом, Эйлер и Лагранж.

Поучительно получить с помощью леммы Гаусса еще одно правило такого же типа, ибо этот метод будет использован в следующем пункте для доказательства закона взаимности. Найдем, для каких простых 3 является вычетом и для каких — невычетом. Числа 3, 6, 9, ..., $3P$ меньше $\frac{3}{2}p$, следовательно, отрицательными после приведения к интервалу между $-\frac{1}{2}p$ и $\frac{1}{2}p$ будут только те из них, которые лежат между $\frac{1}{2}p$ и p. Найдем число целых x, для которых $\frac{1}{2}p < 3x < p$ или $\frac{1}{6}p < x < \frac{1}{3}p$. Положим $p = 12k+r$, где r равно 1, 5, 7 или 11 (для простого, кроме случая, когда p равно 2 или 3, что исключается, имеются только эти возможности). Тогда неравенство принимает вид $2k + \frac{1}{6}r < x < 4k + \frac{1}{3}r$. Мы снова можем не обращать внимания на четные числа $2k$ и $4k$ и свести все к неравенству $\frac{1}{6}r < x < \frac{1}{3}r$. Оно не имеет решения, если r равно 1; имеет одно решение, если r равно

5 или 7; и два решения, если r равно 11. Следовательно, 3 *есть квадратичный вычет для простых вида* $12k\pm1$ *и квадратичный невычет для простых вида* $12k \pm 5$.

5. Закон взаимности. Мы только что доказали, что квадратичный характер от 2 по mod p зависит только от остатка r, если p представлено в виде $8k+r$, и что квадратичный характер от 3 по mod p зависит только от остатка r', если p представлено в виде $12k + r'$. Более того, в первом случае значение характера одинаково для r и для $8 - r$, а в последнем его значение одинаково для r' и для $12 - r'$.

На основании глубокого численного исследования Эйлер пришел к выводу, что аналогичное утверждение имеет место и в общем случае, но доказать этого он не сумел. Пусть a — какое-нибудь натуральное число; представим p в виде $4ka + r$, где $0 < r < 4a$. Эйлер предположил тогда, что *квадратичный характер по* mod p *имеет одно и то же значение для всех простых* p*, для которых* r *имеет одно и то же значение; более того, характер один и тот же для* r *и* $4a - r$. Этот результат эквивалентен квадратичному закону взаимности, который будет сформулирован в конце этого пункта.

Лежандр дал неполное доказательство закона взаимности; первым полным доказательством (очень сложным) было доказательство Гаусса, который открыл этот закон в возрасте девятнадцати лет.

Гипотезу Эйлера можно доказать с помощью леммы Гаусса, следуя по тому же пути, что и в случаях $a = 2$, 3. Мы должны посмотреть, сколько чисел ряда a, $2a$, ..., Pa, где $P = \frac{1}{2}(p - 1)$, лежит между $\frac{1}{2}p$ и p, между $\frac{3}{2}p$ и $2p$ и так далее. Так как Pa, наибольшее кратное a, меньше $\frac{1}{2}pa$, то последний подлежащий рассмотрению интервал — это интервал от $\left(b - \frac{1}{2}\right)p$ до bp, где b равно $\frac{1}{2}a$, если a четно, и $\frac{1}{2}(a-1)$, если a нечетно. Таким образом, нужно подсчитать, сколько чисел, кратных a, лежит в интервалах $\left(\frac{1}{2}p, p\right)$, $\left(\frac{3}{2}p, 2p\right)$, ..., $\left(\left(b - \frac{1}{2}\right)p, bp\right)$. Ни одно из встречающихся здесь чисел не кратно a, так что вопрос о том, нужно ли учитывать концы интервалов, не возникает.

Разделив все числа, о которых идет речь, на a, получаем, что искомое число равно полному числу целых точек во всех

интервалах

$$\left(\frac{p}{2a}, \frac{p}{a}\right), \left(\frac{3p}{2a}, \frac{2p}{a}\right), \ldots, \left(\frac{(2b-1)p}{2a}, \frac{bp}{a}\right).$$

Положим теперь $p = 4ak + r$. Так как все знаменатели равны либо a, либо $2a$, то без всякого вычисления видно, что замена p на $4ak + r$ равносильна замене p на r с точностью до некоторых четных чисел, добавляющихся к конечным точкам интервалов. Как и раньше, на эти четные числа мы можем не обращать внимания. Следовательно, если ν — полное число целых точек в интервалах

$$\left(\frac{r}{2a}, \frac{r}{a}\right), \left(\frac{3r}{2a}, \frac{2r}{a}\right), \ldots, \left(\frac{(2b-1)r}{2a}, \frac{br}{a}\right), \tag{10}$$

то a является квадратичным вычетом или невычетом по $\bmod p$ в зависимости от того, четно или нечетно ν. Число ν зависит только от r и одинаково для всех p, дающих при делении на $4a$ остаток r.

Это доказывает основную часть гипотезы Эйлера. Посмотрим теперь, что произойдет при замене r на $4a - r$. Эта замена преобразует ряд интервалов (10) в ряд

$$\left(2 - \frac{r}{2a}, 4 - \frac{r}{a}\right), \left(6 - \frac{3r}{2a}, 3 - \frac{2r}{a}\right), \ldots,$$
$$\left(4b - 2 - \frac{(2b-1)r}{2a}, 4b - \frac{br}{a}\right). \tag{11}$$

Обозначим через ν' число целых точек в этих интервалах; мы должны доказать, что ν и ν' имеют одинаковую четность. Простое рассмотрение показывает, что интервал $\left(2 - \frac{r}{2a}, 4 - \frac{r}{a}\right)$ и интервал $\left(\frac{r}{2a}, \frac{r}{a}\right)$ одинаковы с точки зрения четности числа целых, содержащихся в них. Действительно, если вычесть концы первого интервала из 4, то этот интервал заменится на интервал $\left(\frac{r}{a}, 2 + \frac{r}{2a}\right)$. Вместе со вторым интервалом $\left(\frac{r}{2a}, \frac{r}{a}\right)$ они образуют интервал длины 2, а такой интервал содержит точно два целых числа. Аналогичные рассмотрения применимы и к другим интервалам в рядах (10) и (11); следовательно, число $\nu + nu'$ четно, что и устанавливает нужный результат.

Квадратичный закон взаимности был впервые ясно сформулирован Лежандром в 1785 году. Он относится к двум различным простым p и q и выражает квадратичный характер p по $\bmod q$ в терминах квадратичного характера q по $\bmod p$. Этот закон состоит в том, что указанные *характеры совпадают в случае, если хоть одно из чисел p и q не представило в виде $4k+3$, и противоположны, если и p, и q имеют вид $4k+3$.* Это можно выразить символически формулой

$$\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = (-1)^{\frac{p-1}{2}\,\frac{q-1}{2}}. \tag{12}$$

Показатель при -1 в правой части четен, если хоть одно из чисел p и q имеет вид $4k+1$, и нечетен, когда оба числа имеют вид $4k+3$. Мы выведем закон взаимности из только что установленных результатов о значении квадратичного характера фиксированного числа a по различным простым модулям.

Предположим сначала, что $p \equiv q \pmod 4$. Не нарушая общности, можно считать, что $p > q$; положим $p - q = 4a$. Тогда мы имеем

$$\left(\frac{p}{q}\right) = \left(\frac{4a+q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right).$$

Аналогично

$$\left(\frac{q}{p}\right) = \left(\frac{p-4a}{p}\right) = \left(\frac{-4a}{p}\right) = \left(\frac{-1}{p}\right)\left(\frac{a}{p}\right).$$

Но $\left(\frac{a}{p}\right)$ и $\left(\frac{a}{q}\right)$ одинаковы, ибо p и q имеют один и тот же остаток при делении на $4a$. Значит, $\left(\frac{p}{q}\right)\left(\frac{q}{p}\right) = \left(\frac{-1}{q}\right)$, а это есть 1, если p и q вида $4k+1$, и -1, если они имеют вид $4k+3$.

Предположим теперь, что $p \not\equiv q \pmod 4$; тогда $p \equiv -q \pmod 4$. Положим $p + q = 4a$. Тогда, так же как и раньше, мы получим

$$\left(\frac{p}{q}\right) = \left(\frac{4a-q}{q}\right) = \left(\frac{4a}{q}\right) = \left(\frac{a}{q}\right).$$

Аналогично

$$\left(\frac{q}{p}\right) = \left(\frac{a}{p}\right).$$

Но $\left(\frac{a}{p}\right)$ и $\left(\frac{a}{q}\right)$ равны, ибо числа p и q дают при делении на $4a$ один и тот же остаток. Этим доказательство закона взаимности заканчивается.

Квадратичный закон взаимности — одна из наиболее известных теорем теории чисел. Он обнаруживает простое и в то же время замечательное взаимоотношение между разрешимостью сравнений $x^2 \equiv q \pmod{p}$ и $x^2 \equiv p \pmod{q}$, взаимоотношение, отнюдь не очевидное заранее. Стремление установить, что кроется за этим законом, было важным фактором в работе многих математиков и привело к далеко идущим открытиям. Первое точное доказательство, данное Гауссом в его «*Исследованиях*», проводилось индукцией по двум простым p и q; такое доказательство сложно и мало удовлетворительно. Гаусс дал семь доказательств, основанных на совершенно различных методах и устанавливающих связь между законом взаимности и различными другими арифметическими теориями.

Закон взаимности дает возможность в каждом конкретном случае вычислять значения $(a|p)$, не прибегая к рассмотрению вопроса о разрешимости сравнений. Вычислим, например, $(34|97)$. Первый шаг состоит в разложении 34 на $2 \cdot 17$. Так как 97 — простое вида $8k+1$, то мы имеем $(2|97) = 1$, поэтому $(34|97) = (17|97)$. Так как 17 и 97 — простые, не являющиеся одновременно числами вида $4k+3$, то, по закону взаимности, $(17|97) = (97|17)$ или $(12|17)$, так как $97 \equiv 12 \pmod{17}$. Далее $(12|17) = (3|17) = (17|3)$, опять-таки благодаря закону взаимности. Так как $17 \equiv -1 \pmod{3}$, то значение символа равно $(-1|3)$ или -1.

Для кубических или высших степенных вычетов такого простого закона, как квадратичный закон взаимности, нет. Но можно кратко упомянуть об одном результате Гаусса, связанном с вычетами четвертой степени. Сначала мы должны напомнить, что, согласно результатам п. 1, теория вычетов четвертой степени интересна лишь для простых вида $4n+1$, так как если p имеет вид $4n+3$, то наибольший общий делитель 4 и $p-1$ равен 2, т. е. в обозначениях п. 1 $k = 2$, а потому в этом случае вычеты четвертой степени совпадают с квадратичными вычетами. Если же p имеет вид $4n+1$, то половина квадратичных вычетов суть вы-

четы четвертой степени (таковы те из них, индекс которых делится на 4), а другая половина и все квадратичные невычеты суть невычеты четвертой степени. Результат Гаусса состоит в том, что 2 есть вычет четвертой степени по mod p тогда и только тогда, когда простое p представимо в виде $x^2 + 64y^2$. Надо отметить, что простое p вида $4n + 1$ всегда представимо, как $a^2 + b^2$ (это мы докажем в главе V); очевидно, одно из чисел a и b должно быть нечетным, а другое — четным. Таким образом, условие Гаусса состоит в том, что четное из чисел a и b должно делиться на 8. Например, 2 является вычетом четвертой степени по mod 73, так как $73 = 3^2 + 64$.

6. Распределение квадратичных вычетов. Вернемся к вопросам, связанным с распределением квадратичных вычетов и невычетов по единственному простому модулю p. Мы знаем, что половина из чисел $1, 2, \ldots, p-1$ — квадратичные вычеты, другая же половина — квадратичные невычеты. Уже первые наблюдения показывают, что если p — большое простое число, то вычеты и невычеты распределяются довольно случайно. Это распределение подчинено, конечно, известным законам, например мультипликативному закону, и тому факту, что всякий точный квадрат всегда является квадратичным вычетом.

Можно ставить разные вопросы, проверяющие случайный характер этого распределения. Можно спросить, например, как распределены вычеты и невычеты в подынтервале интервала от 0 до p. Пусть α и β — две фиксированные правильные дроби; верно ли, что при большом p примерно половина чисел между αp и βp — квадратичные вычеты? Если это так, то можно сказать, что квадратичные вычеты равномерно распределены. Это предложение действительно верно, но, кажется, не известно никакого элементарного доказательства этого факта.

Более простой вопрос, ответ на который дал Гаусс, связан с характерами последовательных чисел. Пусть n и $n+1$ — два последовательных числа в ряду $1, 2, \ldots, p-1$; спрашивается, как часто они имеют предписанные значения характеров? Возможные значения характеров для пары чисел таковы: BB, BH, HB, HH *). Если квадратичные вычеты и невычеты распределены

*) B — вычет, H — невычет. (*Прим. перев.*)

случайно, то следует ожидать, что каждый из четырех указанных типов встречается примерно одинаково часто. Обозначим через (BB) и так далее число пар n, $n+1$ с предписанными значениями характеров. Ясно, что $(BB)+(BH)$ равно числу пар, в которых n — квадратичный вычет. Здесь n принимает значения 1, 2, ..., $p-2$. Полное число квадратичных вычетов среди 1, 2, ..., $p-1$ равно $\frac{1}{2}(p-1)$, а характер $p-1$, или -1, равен $(-1)^{(p-1)/2}$. Поэтому

$$(BB)+(BH)=\tfrac{1}{2}(p-2-\epsilon), \tag{13}$$

где $\epsilon=(-1)^{(p-1)/2}$. Аналогично

$$(HB)+(HH)=\tfrac{1}{2}(p-2+\epsilon), \tag{14}$$

$$(BB)+(HB)=\tfrac{1}{2}(p-1)-1, \tag{15}$$

$$(BH)+(HH)=\tfrac{1}{2}(p-1). \tag{16}$$

Для четырех неизвестных имеются четыре соотношения, но они не независимы, ибо при сложении первых двух мы получаем тот же результат, что и при сложении двух последних. Поэтому, чтобы определить эти четыре неизвестных, нам нужно еще одно соотношение.

Рассмотрим произведение символов Лежандра $\left(\frac{n}{p}\right)$ и $\left(\frac{n+1}{p}\right)$. Оно равно $+1$ в случаях BB и HH и -1 в случаях BH и HB. Значит, $(BB)+(HH)-(BH)-(HB)$ равно сумме всех символов Лежандра $\left(\frac{n(n+1)}{p}\right)$, где n принимает значения $1, 2, \ldots, p-2$. Любое целое n из этого ряда имеет обратное по $\bmod p$, которое мы обозначим через m. Но $n(n+1)\equiv n^2(1+m) \pmod p$, откуда $\left(\frac{n(n+1)}{p}\right)=\left(\frac{1+m}{p}\right)$. В то время как n принимает значения $1, 2, \ldots, p-2$, т. е. все значения от 1 до $p-1$, кроме $p-1$, его обратное m также принимает все значения от 1 до $p-1$, кроме $p-1$. Поэтому $1+m$ принимает все значения от 2 до $p-1$. Сумма символов Лежандра этих чисел есть

$$\left(\frac{2}{p}\right)+\left(\frac{3}{p}\right)+\ldots+\left(\frac{p-1}{p}\right).$$

Но $\left(\frac{1}{p}\right)+\left(\frac{2}{p}\right)+\left(\frac{3}{p}\right)+\ldots+\left(\frac{p-1}{p}\right)=0$, так как вычетов столько же, сколько и невычетов. Поэтому интересующая нас сумма равна $-\left(\frac{1}{p}\right)$ или -1. Таким образом,

$$(BB)+(HH)-(BH)-(HB)=-1. \tag{17}$$

Это соотношение, скомбинированное с помощью почленного сложения и вычитания с полученными ранее соотношениями, дает нам значения (BB) и др. Если сложить (17) с (13) и (14), получится $(BB)+(HH)=\frac{1}{2}(p-3)$. С другой стороны, вычитание (14) из (15) дает $(BB)-(HH)=-\frac{1}{2}(1+\epsilon)$. Отсюда (BB) и др. лежат между $\frac{1}{4}(p-5)$ и $\frac{1}{4}(p+1)$. Поэтому утверждение о том, что они все примерно равны $\frac{1}{4}p$ для больших p, выполняется довольно хорошо.

Важным шагом в доказательстве была оценка суммы символов Лежандра $\left(\frac{n(n+1)}{p}\right)$. Если мы условимся считать $\left(\frac{0}{p}\right)=0$, то можно вместо 1, 2, ..., $p-2$ допустить для n все значения 0, 1, 2, ..., $p-1$, не изменив суммы. Поэтому результат можно выразить в форме

$$\sum\left(\frac{n(n+1)}{p}\right)=-1, \tag{18}$$

где символ Σ обозначает суммирование по полной системе вычетов по mod p. Можно показать, что этот результат имеет место и в более общем случае для любой суммы $\sum\left(\frac{n^2+bn+c}{p}\right)$ с квадратным полиномом, старший коэффициент которого равен 1, хотя и не только что использованным методом. Очевидное исключение представляет, конечно, случай, в котором полином является точным квадратом. Для полиномов более высоких степеней подобные вопросы глубоко исследовались в течение примерно последних двадцати лет. В 1934 году Хассе (H. Hasse) показал с помощью очень трудных и глубоких методов, что значение любой кубической суммы $\sum\left(\frac{an^3+bn^2+cn+d}{p}\right)$ лежит

между $-2\sqrt{p}$ и $2\sqrt{p}$ *). В дальнейшем А. Вейль (A. Weil) обобщил этот результат. Теорема А. Вейля имеет важные далеко идущие следствия.

Замечания к главе III. п. 1. Имеется другое доказательство существования первообразного корня, принадлежащее Гауссу. Но я предпочел доказательство Лежандра как более конструктивное.

В согласии с теоремами Ферма и Эйлера (II, 3) число называют первообразным корнем по модулю m, если порядок его равен $\varphi(m)$. Гаусс доказал, что первообразные корни существуют для модулей 2, 4, p^n, $2p^n$, где p — простое, большее 2, n — любое целое, и только для этих модулей.

п. 2. Таблица индексов для простых, меньших 97, имеется в книге Успенского и Хислета ([14]).

п. 3. Можно вывести мультипликативное свойство и критерий Эйлера непосредственно из определения квадратичного вычета, не используя индексов, но такие доказательства менее наглядны.

п. 5. Применяя это рассмотрение к закону взаимности, я следую Шольцу (см. ([19])).

п. 6. Факт равной распределенности вычетов и невычетов следует из одного важного неравенства, открытого Полиа в 1917 году и независимо от него Виноградовым в 1918 году. Это неравенство устанавливает, что сумма символов Лежандра $(n|p)$ (где n пробегает любую совокупность последовательных чисел n) по абсолютной величине меньше $C\sqrt{p}\log p$, где C — некоторая постоянная. Так как $\sqrt{p}\log p$ мало по сравнению с p при больших p, то отсюда следует, что в любом интервале от αp до βp, где p — большое, а α и β — фиксированные числа, вычетов почти столько же, сколько и невычетов. О более глубоких результатах, связанных с распределением квадратичных вычетов и невычетов, см. работу Берджэсса (D. Burgess, Mathematika, **4** (1957), 106—112) **).

---

*) Элементарное доказательство теоремы Хассе дал Ю. И. Манин, см. ([5], гл. 10). (*Прим. перев.*)

**) Имеется перевод в журнале «Математика» (сб. переводов), 2:6, 1958, стр. 3—9. (*Прим. перев.*)

ГЛАВА IV

НЕПРЕРЫВНЫЕ ДРОБИ

1. Введение. В (I, 6) мы описали алгоритм Евклида для нахождения наибольшего общего делителя двух данных чисел. Имеется еще один способ описания этого алгоритма, в результате которого отношение двух чисел представляется в виде непрерывной дроби. Этот способ станет ясным из следующего численного примера.

Применим алгоритм Евклида к числам 67 и 24. Последовательные шаги алгоритма таковы:

$$\begin{aligned} 67 &= 2\cdot 24 + 19,\\ 24 &= 1\cdot 19 + \ 5,\\ 19 &= 3\cdot \ 5 + \ 4,\\ 5 &= 1\cdot \ 4 + \ 1. \end{aligned}$$

Последний остаток, равный 1, как известно, указывает на то, что числа 67 и 24 взаимно просты. Представим теперь каждое из этих уравнений в виде дроби:

$$\frac{67}{24} = 2 + \frac{19}{24}, \quad \frac{24}{19} = 1 + \frac{5}{19}, \quad \frac{19}{5} = 3 + \frac{4}{5}, \quad \frac{5}{4} = 1 + \frac{1}{4}.$$

Последняя дробь каждого из этих уравнений обратна первой дроби следующего уравнения. Мы можем поэтому, исключив все промежуточные дроби, представить исходную дробь $\frac{67}{24}$ в виде

$$2 + \cfrac{1}{1 + \cfrac{1}{3 + \cfrac{1}{1 + \cfrac{1}{4}}}}.$$

Такое выражение называется *непрерывной дробью*. Для удобства печати принимается следующая форма записи:

$$2+\frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{1+}\,\frac{1}{4}.$$

Числа 2, 1, 3, 1, 4 называются *элементами* непрерывной дроби, или *неполными частными*; они служат неполными частными в последовательных шагах алгоритма Евклида, применяемого к числителю и знаменателю исходной дроби. *Полные частные* — это сами числа $\frac{67}{24}$, $\frac{24}{19}$, $\frac{19}{5}$, $\frac{5}{4}$. Каждое из них разлагается в непрерывную дробь, которая получается из вышенаписанной дроби отбрасыванием нескольких первых членов, например:

$$\frac{24}{19}=1+\frac{1}{3+}\,\frac{1}{1+}\,\frac{1}{4},\qquad \frac{19}{5}=3+\frac{1}{1+}\,\frac{1}{4}.$$

Из упомянутого примера и из известных нам фактов об алгоритме Евклида следует, что каждое рациональное число $\frac{a}{b}$ можно представить в виде непрерывной дроби

$$\frac{a}{b}=q+\frac{1}{r+}\,\frac{1}{s+}\,\ldots\frac{1}{w},$$

где элементы q, r, s, …, w являются натуральными числами. Последний элемент, w в предыдущей записи, должен быть больше 1, так как это последнее частное в алгоритме Евклида.

Легко доказать, что имеется только одно представление данного рационального числа непрерывной дробью. Действительно, предположим, что

$$\frac{a}{b}=q+\frac{1}{r+}\,\frac{1}{s+}\,\ldots=q'+\frac{1}{r'+}\,\frac{1}{s'+}\,\ldots,$$

где q', r', s' … — натуральные числа, последнее из которых больше 1. Дроби, которые прибавляются в левой части к q, а в правой части к q', меньше 1. Поэтому числа q и q' равны целой части рационального числа $\frac{a}{b}$ и, следовательно, $q=q'$. Вычитая q и q' и переходя к обратной дроби, получаем $r+\frac{1}{s+}\,\ldots=r'+$ $+\frac{1}{s'+}\,\ldots$ Аналогичное рассуждение показывает, что $r=r'$ и так далее.

Прежде чем двигаться дальше, читателю, не знакомому с непрерывными дробями, следует попрактиковаться в разложе-

нии рациональных чисел. Вот примеры:

$$\frac{17}{11} = 1 + \frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{5}, \qquad \frac{11}{31} = \frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{2}.$$

Если рациональное число, как во втором примере, меньше 1, то первое неполное частное равно 0, и мы его опускаем.

2. Общая непрерывная дробь. Непрерывные дроби оказываются очень полезными в теории чисел; используя их, часто можно дать точную конструкцию для решения задачи, в то время как другими методами доказывается лишь существование такого решения.

Представим общую непрерывную дробь в виде

$$q_0 + \frac{1}{q_1+}\,\frac{1}{q_2+}\,\dots\,\frac{1}{q_n}. \tag{1}$$

Прежде чем исследовать арифметические свойства непрерывных дробей, нужно установить некоторые чисто алгебраические соотношения. Эти соотношения не зависят от природы элементов $q_0, q_1, \dots, q_n$. Мы будем поэтому обращаться с этими элементами как с переменными, не обязанными быть натуральными числами.

Если шаг за шагом вычислять значение непрерывной дроби (1), то, в конце концов, для нее, очевидно, получится выражение, являющееся отношением двух сумм, составленных из различных произведений чисел $q_0, q_1, \dots, q_n$. Если n равно 1, мы имеем

$$q_0 + \frac{1}{q_1} = \frac{q_0q_1 + 1}{q_1}.$$

Если n равно 2, получаем

$$q_0 + \frac{1}{q_1+}\,\frac{1}{q_2} = q_0 + \frac{q_2}{q_1q_2 + 1} = \frac{q_0q_1q_2 + q_0 + q_2}{q_1q_2 + 1}.$$

Значение $q_1 + \dfrac{1}{q_2}$ мы получили из предыдущего вычисления, подставив q_1 и q_2 вместо q_0 и q_1. Аналогично при $n = 3$ имеем

$$q_0 + \frac{1}{q_1+}\,\frac{1}{q_2+}\,\frac{1}{q_3} = q_0 + \frac{q_2q_3 + 1}{q_1q_2q_3 + q_1 + q_3} =$$

$$= \frac{q_0q_1q_2q_3 + q_0q_1 + q_0q_3 + q_2q_3 + 1}{q_1q_2q_3 + q_1 + q_3}. \tag{2}$$

Здесь мы снова использовали результат предшествующего шага.

Ясно, что, продолжая таким образом, мы можем найти значение любой непрерывной дроби. Будем обозначать числитель непрерывной дроби (1), вычисленный таким способом, через $[q_0, q_1, \ldots, q_n]$. Таким образом,

$$[q_0] = q_0, \quad [q_0, q_1] = q_0q_1 + 1, \quad [q_0, q_1, q_2] = q_0q_1q_2 + q_0 + q_2,$$
$$[q_0, q_1, q_2, q_3] = q_0q_1q_2q_3 + q_0q_1 + q_0q_3 + q_2q_3 + 1, \ \ldots$$

Мы видели, что в рассмотренных случаях выражение, получаемое для знаменателя непрерывной дроби, равно $[q_1, q_2, \ldots, q_n]$. Это верно и в общем случае. Действительно, если посмотреть на третий шаг (довольно типичный) в (2), то мы увидим, что знаменателем результата становится числитель $q_1 + \dfrac{1}{q_1+}\,\dfrac{1}{q_2}$, поэтому знаменатель равен $[q_1, q_2, q_3]$.

Следовательно, общая непрерывная дробь имеет вид

$$q_0 + \frac{1}{q_1+}\,\cdots\frac{1}{q_n} = \frac{[q_0, q_1, \ldots, q_n]}{[q_1, q_2, \ldots, q_n]}. \tag{3}$$

Из вычисления в (2) ясно, как строится функция $[q_0, q_1, q_2, q_3]$ из $[q_1, q_2, q_3]$ и $[q_2, q_3]$. Это вычисление дает

$$[q_0, q_1, q_2, q_3] = q_0[q_1, q_2, q_3] + [q_1, q_2].$$

Такое же соотношение имеет место и в общем случае:

$$[q_0, q_1, \ldots, q_n] = q_0[q_1, \ldots, q_n] + [q_1, q_2, \ldots, q_n]. \tag{4}$$

Это *рекуррентное соотношение* шаг за шагом определяет функцию «квадратные скобки». Формула (4) применима, начиная с $n = 2$. Если при $n = 1$ второй скобке в равенстве (4) приписать значение 1, то это соотношение можно будет применять и при $n = 1$. Действительно, в этом случае получаем верное равенство

$$[q_0, q_1] = q_0[q_1] + 1 = q_0q_1 + 1.$$

В качестве иллюстрации это правило можно применить к последнему примеру, рассмотренному в конце п. 1. Мы имеем

$$[4, 2] = 4 \cdot 2 + 1 = 9,$$
$$[1, 4, 2] = 1 \cdot [4, 2] + [2] = 9 + 2 = 11,$$
$$[2, 1, 4, 2] = 2 \cdot [1, 4, 2] + [4, 2] = 2 \cdot 11 + 9 = 31.$$

Таким образом, $2+\dfrac{1}{1+}\,\dfrac{1}{4+}\,\dfrac{1}{2}=\dfrac{[2,1,4,2]}{[1,4,2]}=\dfrac{31}{11}$.

Сделаем одно замечание. Мы видели, что общую непрерывную дробь можно выразить в виде (3), где две квадратные скобки представляют собой некоторые суммы произведений переменных q_0, q_1, ..., q_n. Мы не доказали, что в этом представлении числитель и знаменатель нельзя ни на что сократить. В первом случае, если числитель и знаменатель — полиномы от переменных q_0, q_1, ..., q_n, можно доказать, что эти полиномы неприводимы, т. е. не разлагаются на множители, являющиеся полиномами. Во втором случае, если q_0, q_1, ..., q_n — целые числа, то числитель и знаменатель также целые числа и можно доказать, что эти числа всегда взаимно просты. Второе из этих утверждений будет доказано в п. 4. Первый факт доказывается еще проще, но с точки зрения теории чисел не представляет интереса.

3. Правило Эйлера. Мы видели, что $[q_0, q_1, \ldots, q_n]$ является суммой некоторых произведений, образованных из элементов q_0, q_1, ..., q_n. Каковы же эти произведения? Эйлер впервые ответил на этот вопрос, указав общее правило для вычисления значения непрерывной дроби. *Сначала берется произведение всех элементов, затем всевозможные произведения, которые можно получить, опустив какуюнибудь пару последовательных элементов. Затем берутся произведения, получающиеся отбрасыванием любых двух пар последовательных элементов, и так далее.* Сумма этих произведений равна $[q_0, q_1, \ldots, q_n]$. Ясно, что если $n+1$ четно, то на последнем шаге отбрасыванием всех элементов мы получим пустое произведение. Его значение принимается по определению равным 1.

Пример применения правила Эйлера:

$$[q_0, q_1, q_2, q_3] = q_0q_1q_2q_3 + q_2q_3 + q_0q_3 + q_0q_1 + 1.$$

Мы взяли сначала произведение всех элементов, затем произведение, получающееся, если опустить пару q_0, q_1, затем произведение, получающееся, если опустить пару q_1, q_2, затем то, что получится, если опустить пару q_2, q_3, и, наконец, пустое произ-

ведение, которое получается, если опустить обе пары q_0, q_1 и q_2, q_3.

Другой пример, в котором на один элемент больше:

$$[q_0, q_1, q_2, q_3, q_4] = q_0q_1q_2q_3q_4 + \\ + q_2q_3q_4 + q_0q_3q_4 + q_0q_1q_4 + q_0q_1q_2 + q_0 + q_2 + q_4.$$

Во второй строке мы написали все произведения, получающиеся, если опустить одну пару, а затем — то, что получается, если опустить две различные пары, например, опуская q_0, q_1 и q_2, q_3, получаем q_4.

Замечая, что правило верно для нескольких первых квадратно-скобочных функций, докажем его в общем случае по индукции, применив рекуррентное соотношение (4). В предположении, что правило имеет место для квадратных скобок в правой части (4), мы должны доказать его для квадратной скобки в левой части (4). Выражение $[q_2, \ldots, q_n]$ равно сумме произведений, получающихся из q_0, q_1, ..., q_n, в которых опускается пара q_0, q_1. В то же время $q_0[q_1, \ldots, q_n]$ равно в точности сумме тех произведений, в которых не опускается пара q_0, q_1; действительно, все произведения должны содержать множитель q_0; с другой стороны, если из всех этих произведений удалить q_0, останутся всевозможные произведения, получающиеся из q_1, ..., q_n отбрасыванием каких-либо пар последовательных элементов. Таким образом, мы получим сумму предписываемых произведений из элементов q_0, q_1, ..., q_n, так что доказываемое правило имеет место для $[q_0, q_1, \ldots, q_n]$.

Тем самым правило доказано индукцией по числу переменных.

Из правила Эйлера сразу же следует, что *величина* $[q_0, q_1, \ldots, q_n]$ *не изменится, если записать все элементы в обратном порядке* $[q_0, q_1, \ldots, q_n] = [q_n, q_{n-1}, \ldots, q_0]$. Например, $[2, 4, 1, 2] = [2, 1, 4, 2]$. Отсюда следует, что, кроме рекуррентного соотношения (4), имеется аналогичное соотношение, выражающее квадратную скобку $[q_0, q_1, \ldots, q_n]$ через квадратные скобки, в которых опущены последний элемент и два последних элемента. Это соотношение имеет вид

$$[q_0, q_1, \ldots, q_n] = q_n[q_0, \ldots, q_{n-1}] + [q_0, q_1, \ldots, q_{n-2}]. \qquad (5)$$

Последнее соотношение эквивалентно (4), ибо, если написать элементы в противоположном порядке, оно принимает вид

$$[q_n, q_{n-1}, \ldots, q_0] = q_n[q_{n-1}, \ldots, q_0] + [q_{n-2}, \ldots, q_0],$$

а это — соотношение (4), переписанное в новых обозначениях.

Рекуррентное соотношение (5) в большинстве случаев более удобно, чем (4). Для нас привычнее добавлять элементы, на которые оканчивается непрерывная дробь, чем те элементы, с которых она начинается; соотношение (5) дает возможность исследовать, что происходит в этом случае.

4. Подходящие данной непрерывной дроби. Пусть

$$q_0 + \frac{1}{q_1+} \ldots \frac{1}{q_n} \tag{6}$$

— какая-нибудь непрерывная дробь. В этом пункте мы будем предполагать, что элементы q_0, q_1, …, q_n являются натуральными числами. Различные непрерывные дроби, получающиеся, если оборвать дробь (6) ранее, чем на q_n: q_0, $q_0+\frac{1}{q_1}$, $q_0+\frac{1}{q_1+}\frac{1}{q_2}$, … называются *подходящими* дробями к данной непрерывной дроби. Причина, по которой принято это название, выяснится позднее. Значение подходящей дроби, которая получится, если оборвать исходную дробь на q_m, равно $q_0 + \frac{1}{q_1+} \ldots \frac{1}{q_m} = \frac{[q_0, \ldots, q_m]}{[q_1, \ldots, q_m]}$. Для упрощения обозначений положим

$$A_m = [q_0, \ldots, q_m], \qquad B_m = [q_1, \ldots, q_m], \tag{7}$$

так что записанная выше подходящая дробь равна $\frac{A_m}{B_m}$. Первая подходящая дробь равна $\frac{A_0}{B_0} = \frac{q_0}{1}$. Последняя равна $\frac{A_n}{B_n}$, т. е. значению самой непрерывной дроби. Числа A_0, B_0, A_1, B_1, …, будучи суммами произведений, образованных из q_i по правилу Эйлера, являются натуральными числами.

Рекуррентное соотношение (5) в новых обозначениях принимает простую форму

$$A_m = q_m A_{m-1} + A_{m-2}. \tag{8}$$

То же рекуррентное соотношение (если отбросить в нем q_0) показывает, что

$$B_m = q_m B_{m-1} + B_{m-2}. \tag{9}$$

Таким образом, числители и знаменатели подходящих дробей получаются по одинаковым общим правилам. Эти правила очень удобны для вычислений. Написав первые две подходящие дроби, мы получим последующие подходящие, применяя (8) и (9). Например, непрерывная дробь для $\frac{42}{31}$ имеет вид $1+\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{2}$. Две первые подходящие дроби равны, очевидно, $\frac{1}{1}$ и $\frac{3}{2}$. Так как следующее неполное частное равно 1, то следующая подходящая дробь равна $\frac{3+1}{2+1} = \frac{4}{3}$. Следующее неполное частное равно 4, поэтому следующая подходящая дробь равна $\frac{4\cdot 4+3}{4\cdot 3+2} = \frac{19}{14}$. Последнее неполное частное равно 2, значит, последняя подходящая дробь есть $\frac{2\cdot 19+4}{2\cdot 14+3} = \frac{42}{31}$.

Между двумя последовательными подходящими дробями имеется очень важное простое соотношение. Вот оно:

$$A_m B_{m-1} - B_m A_{m-1} = (-1)^{m-1}. \tag{10}$$

Например, если m равно 1, то $A_0 = q_0$, $B_0 = 1$, $A_1 = q_0 q_1 + 1$, $B_1 = q_1$, откуда

$$A_1 B_0 - B_0 A_1 = (q_0 q_1 + 1) - q_0 q_1 = 1. \tag{11}$$

Чтобы доказать (10) в общем случае, подставим A_m и B_m из рекуррентных соотношений (8) и (9). Это дает

$$\begin{aligned} A_m B_{m-1} - B_m A_{m-1} &= (q_m A_{m-1} + A_{m-2}) B_{m-1} - \\ &- (q_m B_{m-1} + B_{m-2}) A_{m-1} = -(A_{m-1} B_{m-2} - B_{m-1} A_{m-2}). \end{aligned}$$

Следовательно, выражение в левой части (10), скажем Δ_m, обладает свойством $\Delta_m = -\Delta_{m-1}$, откуда $\Delta_m = -\Delta_{m-1} = \Delta_{m-2} = \ldots = \pm\Delta_1$, в конце берется знак плюс, если m нечетно, и знак минус, если m четно, поэтому вместо него можно поставить $(-1)^{m-1}$. Так как в силу (11) $\Delta_1 = 1$, отсюда следует общий результат.

Из равенства (10) немедленно следует, что A_m и B_m *взаимно просты*: любой их общий множитель должен быть делителем

1. Таким образом, дробь $\frac{A_m}{B_m}$, представляющая общую подходящую дробь, записана с наименьшими из возможных числителем и знаменателем. В частности, взяв m равным n, видим, что это верно и для прежней формулы (3), выражающей значение общей непрерывной дроби. Таким образом, мы доказали предложение, высказанное в конце п. 2.

Если рациональнее число $\frac{a}{b}$ разложить в непрерывную дробь, то подходящие этой непрерывной дроби образуют последовательность рациональных чисел, последнее из которых равно самой дроби $\frac{a}{b}$. Каковы соотношения между величиной этих чисел и величиной $\frac{a}{b}$? Легко установить, что *подходящие дроби поочередно меньше или больше, чем значение* $\frac{a}{b}$. Чтобы доказать это, перепишем соотношение (10) в виде

$$\frac{A_m}{B_m} - \frac{A_{m-1}}{B_{m-1}} = \frac{(-1)^{m-1}}{B_{m-1}B_m}. \tag{12}$$

Мы видим, что разность в левой части положительна при нечетном m и отрицательна при четном m. Кроме того, так как числа B_0, B_1, B_2, ... образуют монотонно возрастающую последовательность, разность в (12) монотонно убывает при возрастании m. Таким образом, $\frac{A_1}{B_1}$ больше, чем $\frac{A_0}{B_0}$; $\frac{A_2}{B_2}$ меньше, чем $\frac{A_1}{B_1}$, но больше, чем $\frac{A_0}{B_0}$; $\frac{A_3}{B_3}$ больше, чем $\frac{A_2}{B_2}$, но меньше, чем $\frac{A_1}{B_1}$, ..., и, наконец, последняя подходящая дробь $\frac{A_n}{B_n} = \frac{a}{b}$. Отсюда следует, что все четные подходящие дроби $\frac{A_0}{B_0}$, $\frac{A_2}{B_2}$ ... меньше, чем $\frac{a}{b}$, а все нечетные — больше, чем $\frac{a}{b}$.

Можно доказать, что *каждая подходящая дробь ближе к окончательному значению* $\frac{a}{b}$, *чем предшествующие подходящие.* Доказательство несложно, но мы его здесь опускаем.

Другой интересный факт состоит в том, что подходящие дроби являются «наилучшими из возможных» приближений к $\frac{a}{b}$ дробями фиксированной сложности. Мы измеряем сложность

дроби величиной ее знаменателя, так что любая дробь, расположенная к $\frac{a}{b}$ ближе, чем подходящая дробь $\frac{A_m}{B_m}$, имеет знаменатель, больший B_m.

Чтобы проиллюстрировать эти свойства подходящих дробей, рассмотрим непрерывную дробь для $\frac{42}{31}$. Последовательными подходящими являются $\frac{1}{1}, \frac{3}{2}, \frac{4}{3}, \frac{19}{14}, \frac{42}{31}$. В представлении десятичными дробями эти числа имеют вид

$$1;\ 1,5;\ 1,333\ldots;\ 1,3571\ldots;\ 1,3548\ldots;$$

мы видим, что они поочередно меньше или больше окончательного числа и монотонно приближаются к нему.

5. Уравнение $ax - by = 1$. В (I, 8) было доказано, что если a и b — произвольные взаимно простые натуральные числа, то можно найти натуральные числа x и y, удовлетворяющие уравнению $ax - by = 1$. Процесс разложения $\frac{a}{b}$ в непрерывную дробь дает точную конструкцию для чисел x и y. Предположим, что имеется непрерывная дробь

$$\frac{a}{b} = q_0 + \frac{1}{q_1+}\ldots\frac{1}{q_n}.$$

Последняя подходящая дробь $\frac{A_n}{B_n}$ равна самой дроби $\frac{a}{b}$. Предшествующая ей подходящая дробь $\frac{A_{n-1}}{B_{n-1}}$ удовлетворяет равенству $A_nB_{n-1} - B_nA_{n-1} = (-1)^{n-1}$ или $aB_{n-1} - bA_{n-1} = (-1)^{n-1}$ (ввиду (10) из предыдущего пункта). Следовательно, если взять $= B_{n-1}$, $y = A_{n-1}$, мы получим решение в натуральных числах уравнения $ax - by = (-1)^{n-1}$. Если n нечетно, то это — рассматриваемое уравнение. Если n четно, так что $(-1)^{n-1} = -1$, мы все же можем решить уравнение с $+1$ одним из следующих двух способов (по существу, оба способа одинаковы). Один способ — взять $x = b - B_{n-1}$ и $y = a - A_{n-1}$, тогда

$$ax - by = a(b - B_{n-1}) - b(a - A_{n-1}) = -aB_{n-1} + bA_{n-1} = 1.$$

Другой способ — изменить непрерывную дробь, представив по-

следний элемент q_n в виде $(q_n - 1) + \frac{1}{1}$. Новая непрерывная дробь имеет на один член больше, чем старая, и ее предпоследняя подходящая дробь является решением уравнения, в котором справа стоит $+1$. На самом деле, здесь получается то же решение, что и в первом случае.

В качестве простого числового примера найдем натуральные числа x и y, удовлетворяющие уравнению

$$61x - 48y = 1.$$

Непрерывная дробь для $\frac{61}{48}$ такова:

$$1 + \frac{1}{3+}\,\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{4}.$$

Подходящими дробями для нее будут $\frac{1}{1}$, $\frac{4}{3}$, $\frac{5}{4}$, $\frac{14}{11}$, $\frac{61}{48}$. Так как в этом случае n равно 4, числа $x = 11$ и $y = 14$ удовлетворяют уравнению $61x - 48y = -1$. Чтобы решить нужное уравнение, возьмем $x = 48 - 11 = 37$, $y = 61 - 14 = 47$. Или видоизменим непрерывную дробь:

$$1 + \frac{1}{3+}\,\frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{3+}\,\frac{1}{1}.$$

Теперь подходящими дробями служат

$$\frac{1}{1},\ \frac{4}{3},\ \frac{5}{4},\ \frac{14}{11},\ \frac{47}{37},\ \frac{61}{48},$$

предпоследняя подходящая дробь, $\frac{47}{37}$, дает решение.

Легко заметить, что эта конструкция дает наименьшее решение уравнения: так получается решение, для которого x меньше b и y меньше a. Если обозначить это решение через x_0, y_0, то общее решение будет задаваться формулой $x = x_0 + bt$, $y = y_0 + at$, где t — любое положительное целое число или нуль. Если t не равно нулю, то x больше b, а y больше a.

6. Бесконечные непрерывные дроби. До сих пор мы рассматривали разложение *рациональных* чисел в непрерывную дробь. Можно разложить в непрерывную дробь и *иррациональное* число, но в этом случае разложение будет бесконечным.

Пусть α — какое-нибудь иррациональное число. Обозначим через q_0 целую часть α, т. е. наибольшее целое число, не превосходящее α. Тогда $\alpha = q_0 + \alpha'$, где α' — дробная часть α, удовлетворяющая неравенству $0 < \alpha' < 1$. Положим $\alpha' = \dfrac{1}{\alpha_1}$, тогда $\alpha = q_0 + \dfrac{1}{\alpha_1}$ и $\alpha_1 > 1$. Ясно, что α_1 также иррационально, так как если бы оно было рациональным, то рациональным было бы и само α. Далее, повторяем эту операцию с α_1, представив его в виде $\alpha_1 = q_1 + \dfrac{1}{\alpha_2}$, где $\alpha_2 > 1$. Этот процесс можно продолжать без конца. Получив иррациональное число $\alpha_n > 1$, представим его в виде $\alpha_n = q_n + \dfrac{1}{\alpha_{n+1}}$, где $\alpha_{n+1} > 1$, q_n — натуральное число. Используя полученные равенства, находим

$$\alpha = q_0 + \frac{1}{q_1+} \ \ldots \frac{1}{q_n+} \frac{1}{\alpha_{n+1}}. \tag{13}$$

Числа $q_1, \ldots, q_n$ — натуральные; целое число q_0 может быть положительным, отрицательным или нулем. Если $\alpha > 1$, q_0 — положительное число и все элементы дроби являются натуральными числами. Числа $q_0, q_1, \ldots$ называются, как и раньше, *элементами*, или *неполными частными*, непрерывной дроби; *полное* частное, соответствующее q_n, равно α_n, или (что то же) $q_n + \dfrac{1}{\alpha_{n+1}}$. Описанный процесс никогда не окончится, потому что каждое полное частное $\alpha_1, \alpha_2, \ldots$ является иррациональным числом.

Подходящие к непрерывной дроби: $\dfrac{A_0}{B_0} = q_0$, $\dfrac{A_1}{B_1} = q_0 + \dfrac{1}{q_1}$, $\dfrac{A_2}{B_2} = q_0 + \dfrac{1}{q_1+} \dfrac{1}{q_2}$, $\ldots$ образуют бесконечную последовательность рациональных чисел. Рекуррентные соотношения (8) и (9) выполняются и для этой последовательности, так как рассматриваемые дроби являются подходящими дробями к конечной непрерывной дроби (13) и можно применить результаты, доказанные ранее. Мы видим теперь, что вначале важно было не ограничиваться рассмотрением непрерывных дробей с натуральными элементами. Сделав так, мы не смогли бы применить полученные результаты к непрерывной дроби (13), так как в нее входит иррациональное число α_{n+1}.

Равенство (13) дает возможность выразить α через полное частное α_{n+1} и две подходящие дроби $\frac{A_n}{B_n}$ и $\frac{A_{n-1}}{B_{n-1}}$. Действительно, используя прежние обозначения, из (13) можно легко найти, что

$$\alpha = \frac{[q_0, q_1, \ \ldots, \ q_n, \alpha_{n+1}]}{[q_1, q_2, \ \ldots, \ q_n, \alpha_{n+1}]}.$$

Далее, в силу (5)

$$[q_0, \ q_1, \ \ldots, \ q_n, \ \alpha_{n+1}] = \\ = \alpha_{n+1}[q_0, q_1, \ \ldots, \ q_n] + [q_0, q_1, \ \ldots, \ q_{n-1}] = \alpha_{n+1}A_n + A_{n-1}.$$

Аналогично знаменатель α равен $\alpha_{n+1}B_n + B_{n+1}$, откуда

$$\alpha = \frac{\alpha_{n+1}A_n + A_{n-1}}{\alpha_{n+1}B_n + B_{n-1}}. \tag{14}$$

В дальнейшем на протяжении этой главы формула (14) будет наиболее полезна.

Установив, что формула (13) имеет место для сколь угодно больших n, хочется написать

$$\alpha = q_0 + \frac{1}{q_1+} \ \frac{1}{q_2+} \ \ldots \tag{15}$$

Но до этого необходимо точно представить себе, что может означать такая запись. По виду этого утверждения из него следует, что как-то производится бесконечное число операций сложения и деления, указанных в правой части, посредством которых вырабатывается некоторое число; утверждается, что это число равно α. Единственным способом осмыслить значение результата бесконечного числа операций является использование понятия предела. Если мы сможем доказать, что последовательность подходящих дробей $\frac{A_0}{B_0}$, $\frac{A_1}{B_1}$, $\frac{A_2}{B_2}$, ..., где $\frac{A_n}{B_n} = q_0 + \\ + \frac{1}{q_1+} \ldots \frac{1}{q_n}$, имеет некоторый предел при бесконечном возрастании n, то правую часть (15) можно будет истолковать как значение этого предела. Если этот предел равен α, то равенство (15) действительно имеет место.

Нетрудно доказать, что $\frac{A_n}{B_n}$ стремится к α при неограничен-

ном возрастании n. Равенство (14) дает

$$\alpha - \frac{A_n}{B_n} = \frac{\alpha_{n+1}A_n + A_{n-1}}{\alpha_{n+1}B_n + B_{n-1}} - \frac{A_n}{B_n} = \\ = \frac{A_{n-1}B_n - A_nB_{n-1}}{B_n(\alpha_{n+1}B_n + B_{n-1})} = \frac{\pm 1}{B_n(\alpha_{n+1}B_n + B_{n-1})}$$

(ввиду (10)). Так как $\alpha_{n+1} > q_{n+1}$, имеем

$$\left|\alpha - \frac{A_n}{B_n}\right| < \frac{1}{B_nB_{n+1}}. \tag{16}$$

Числа B_0, B_1, B_2, ... образуют строго возрастающую последовательность натуральных чисел; значит, B_n неограниченно возрастает вместе с n, и (16) доказывает, что $\dfrac{A_n}{B_n}$ имеет своим пределом при неограниченном возрастании n число α. Это свойство оправдывает слова «подходящая дробь»: предел последовательности $\dfrac{A_n}{B_n}$ равен (при неограниченном возрастании n) исходному числу α.

В связи с представлением иррационального числа бесконечной непрерывной дробью возникает еще один вопрос. В какой степени неполные частные q_0, q_1, q_2, ... определялись числом α? Предположим, что мы выбрали произвольную бесконечную последовательность целых чисел q_0, q_1, q_2, ..., в которой все числа, кроме, быть может, первого, являются натуральными. Можно ли приписать какое-нибудь значение непрерывной дроби

$$q_0 + \frac{1}{q_1+}\,\frac{1}{q_2+}\,\ldots\,?$$

Если можно, то будет ли получившееся число иррациональным и связана ли эта непрерывная дробь с дробью, получающейся применением нашего первоначального процесса разложения к упомянутому числу? До тех пор, пока мы не выяснили этих вопросов, наши знания о непрерывных дробях неполны. В действительности ответы на эти вопросы очень просты. Если образовать непрерывную дробь из любой бесконечной последовательности натуральных чисел q_1, q_2, ... и произвольного целого числа q_0, то соответствующая последовательность подходящих дробей будет иметь предел. Вероятно, простейшее доказатель-

ство состоит в рассмотрении последовательности четных подходящих дробей $\frac{A_0}{B_0}$, $\frac{A_2}{B_2}$, ... Это — возрастающая ограниченная сверху последовательность (все ее члены меньше, например, $\frac{A_1}{B_1}$). Поэтому в силу одного из основных предложений о пределах эта последовательность имеет предел. Аналогично последовательность, образованная нечетными подходящими дробями, имеет предел. Эти пределы равны, так как предел разности двух последовательных подходящих равен 0. Таким образом, мы можем приписать некоторое значение любой бесконечной непрерывной дроби. Обозначив это значение через α, можно убедиться, что непрерывная дробь α, полученная методом, описанным в начале этого пункта, совпадает с непрерывной дробью, которой мы приписали значение α. Действительно, значение бесконечной непрерывной дроби

$$\frac{1}{q_1+}\,\frac{1}{q_2+}\cdots$$

лежит между 0 и 1; поэтому q_0 должно совпадать с целой частью α. Если положить $\alpha = q_0 + \frac{1}{\alpha_1}$, то q_1 будет равно целой части α_1 и так далее. Другими словами, непрерывная дробь *единственна.* В частности, число, определяемое бесконечной непрерывной дробью, должно быть иррациональным, так как непрерывная дробь, порождаемая рациональным числом, конечна.

Мы видим теперь, что бесконечные непрерывные дроби могут служить не только для *представления* данных иррациональных чисел, но и для *конструирования* произвольных иррациональных чисел. Один из способов описания возникающей ситуации — сказать, что процесс образования непрерывных дробей устанавливает взаимно однозначное соответствие между (I) всеми иррациональными числами, бóльшими 1, и (II) всеми бесконечными последовательностями q_0, q_1, q_2, ... натуральных чисел.

7. Диофантовы приближения. Процесс образования непрерывных дробей порождает бесконечную последовательность рациональных приближений к данному иррациональному чис-

лу α; эту последовательность образуют подходящие дроби. Некоторую информацию о близости этих приближений к α дает неравенство (16). Из него, в частности, следует, что если $\frac{x}{y}$ — одна из подходящих дробей α, то

$$\left|\alpha - \frac{x}{y}\right| < \frac{1}{y^2}\,. \tag{17}$$

Мы получаем здесь простейший резулыат из теории диофантовых приближений — области математики, занимающейся вопросом о приближении иррациональных чисел рациональными.

Более детальное рассмотрение показывает, что для бесконечного числа рациональных приближений выполняются несколько лучшие неравенства. Прежде всего можно доказать, что из двух последовательных подходящих дробей по крайней мере одна удовлетворяет неравенству

$$\left|\alpha - \frac{x}{y}\right| < \frac{1}{2y^2}\,,$$

так что это неравенство выполняется для бесконечного числа рациональных приближений. Еще немного более сильное неравенство, именно

$$\left|\alpha - \frac{x}{y}\right| < \frac{1}{\sqrt{5}y^2}\,, \tag{18}$$

выполняется для одной из каждых *трех* последовательных подходящих дробей. Таким образом, любое иррациональное число a имеет бесконечно много рациональных приближений, удовлетворяющих неравенству (18), — результат, впервые полученный в 1891 году Гурвицем (Hurwitz). Но дальше этого уже пойти нельзя. Существуют иррациональные числа, для которых всякое более точное неравенство, скажем

$$\left|\alpha - \frac{x}{y}\right| < \frac{1}{ky^2}\,, \tag{19}$$

где $k > \sqrt{5}$, имеет лишь конечное число решений в целых числах x и y. Простейший пример такого числа — число, задаваемое непрерывной дробью вида

$$\theta = 1 + \frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{1+}\,\ldots$$

Это число обладает тем свойством, что любое неравенство типа (19) с θ вместо α имеет только конечное число решений. Значение θ легко найти из уравнения

$$\theta = 1 + \frac{1}{\theta},$$

или

$$\theta^2 - \theta - 1 = 0.$$

Решая это квадратное уравнение, найдем $\theta = \frac{1}{2}(1 + \sqrt{5})$ (отрицательный корень надо отбросить).

Доказательства только что упомянутых результатов не очень трудны, но за этими доказательствами мы все же вынуждены отослать читателя к литературе, указанной в *Замечаниях*.

8. Квадратичные иррациональности. Простейшими и наиболее известными иррациональными числами являются квадратичные иррациональности, т. е. числа, которые получаются в результате решений квадратных уравнений с целыми коэффициентами. В частности, квадратный корень из любого натурального числа N, если N не есть точный квадрат, — квадратичная иррациональность, так как этот корень служит решением уравнения $x^2 - N = 0$. Непрерывные дроби для квадратичных иррациональностей обладают замечательными свойствами. Мы теперь займемся исследованием этих свойств.

Начнем с нескольких численных примеров. В качестве первого очень простого примера рассмотрим $\sqrt{2}$. Так как целая часть $\sqrt{2}$ равна 1, то первый элемент q_0 непрерывной дроби также равен 1, и на первом шагу разложения мы получаем

$$\sqrt{2} = 1 + \frac{1}{\alpha_1}.$$

Здесь

$$\alpha_1 = \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1.$$

Целая часть α_1 равна 2, поэтому следующий шаг дает

$$\alpha_1 = 2 + \frac{1}{\alpha_2}.$$

Здесь

$$\alpha_2 = \frac{1}{\alpha_1 - 2} = \frac{1}{\sqrt{2} - 1} = \sqrt{2} + 1.$$

Так как α_2 совпадает с α_1, то дальнейшие вычисления излишни: все последующие шаги будут совпадать с последним описанным шагом. Следующие элементы непрерывной дроби будут равны 2, и мы получим

$$\sqrt{2} = 1 + \frac{1}{2+}\,\frac{1}{2+}\,\frac{1}{2+}\,\ldots$$

Еще несколько примеров:

$$\sqrt{3} = 1 + \frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\,\ldots,$$

$$\sqrt{5} = 2 + \frac{1}{4+}\,\frac{1}{4+}\,\ldots,$$

$$\sqrt{6} = 2 + \frac{1}{2+}\,\frac{1}{4+}\,\frac{1}{2+}\,\frac{1}{4+}\,\ldots$$

В качестве немного более сложного примера рассмотрим число

$$\alpha = \frac{24 - \sqrt{15}}{17}.$$

Так как $\sqrt{15}$ лежит между 3 и 4, целая часть α равна 1. Первый шаг дает

$$\alpha = 1 + \frac{1}{\alpha_1}.$$

Здесь

$$\alpha_1 = \frac{1}{\alpha - 1} = \frac{17}{7 - \sqrt{15}} = \frac{7 + \sqrt{15}}{2}.$$

Целая часть α_1 равна 5, поэтому

$$\alpha_1 = 5 + \frac{1}{\alpha_2},$$

где

$$\alpha_2 = \frac{1}{\alpha_1 - 5} = \frac{2}{\sqrt{15} - 3} = \frac{\sqrt{15} + 3}{3}.$$

Целая часть α_2 равна 2, так что

$$\alpha_2 = 2 + \frac{1}{\alpha_3},$$

где

$$\alpha_3 = \frac{1}{\alpha_2 - 2} = \frac{3}{\sqrt{15} - 3} = \frac{\sqrt{15} + 3}{2}.$$

Целая часть α_3 равна 3, поэтому

$$\alpha_3 = 3 + \frac{1}{\alpha_4},$$

где

$$\alpha_4 = \frac{1}{\alpha_3 - 3} = \frac{2}{\sqrt{15} - 3} = \frac{\sqrt{15} + 3}{3}.$$

Так как $\alpha_4 = \alpha_2$, последние два шага будут повторяться и дальше, так что непрерывная дробь имеет вид

$$\frac{24 - \sqrt{15}}{17} = 1 + \frac{1}{5+}\,\frac{1}{2+}\,\frac{1}{3+}\,\frac{1}{2+}\,\frac{1}{3+}\,\ldots$$

Мы можем записать это короче:

$$1, 5, \overline{2, 3},$$

где черта выделяет период, повторяющийся бесконечное число раз. В этой краткой записи предыдущие примеры принимают вид

$$\sqrt{2} = 1, \overline{2}; \quad \sqrt{3} = 1, \overline{1, 2}; \quad \sqrt{5} = 2, \overline{4}; \quad \sqrt{6} = 2, \overline{2, 4}.$$

В каждом из этих случаев существует полное частное α_n, равное некоторому предшествующему полному частному α_m. Начиная с такого полного частного, непрерывная дробь становится *периодической*. Элементы от q_m до q_{n-1} все время повторяются. Общую теорему о том, что *любая квадратичная иррациональность разлагается в периодическую* (*начиная с некоторого места*) *дробь*, первым доказал Лагранж в 1770 году, но сам факт был известен и более ранним математикам. Мы докажем эту теорему в п. 10; предварительно, в п. 9, мы рассмотрим чисто периодические непрерывные дроби.

Таблица непрерывных дробей для $\sqrt{N}$ при $N = 2, 3, \ldots, 50$ (за исключением точных квадратов) дана на стр. 110. Для простоты над периодом, состоящим из всех элементов, кроме первого, черта опускается. Можно заметить, что эти непрерывные дроби имеют некоторые общие черты, причина этого выяснится в следующем пункте.

Метод, которым мы пользовались в предыдущих примерах, для удобства вычислений можно упростить, рассматривая только встречающиеся *целые числа* и оформляя работу более компактно.

9. Чисто периодические непрерывные дроби. В рассмотренных примерах непрерывные дроби не были чисто периодическими: период дроби начинался не с первого элемента. Однако легко привести примеры и чисто периодических непрерывных дробей; скажем, прибавив 1 к непрерывной дроби для $\sqrt{2}$, получаем чисто периодическую непрерывную дробь

$$\sqrt{2}+1 = 2 + \frac{1}{2+}\,\frac{1}{2+}\,\frac{1}{2+}\,\ldots,$$

аналогично

$$\sqrt{6}+2 = 4 + \frac{1}{2+}\,\frac{1}{4+}\,\frac{1}{2+}\,\ldots$$

Числа, представляемые чисто периодическими непрерывными дробями, являются квадратичными иррациональностями частного вида; выясним, чем характеризуются эти числа.

Начнем с конкретного примера. Рассмотрим какую-нибудь чисто периодическую непрерывную дробь, скажем

$$\alpha = 4 + \frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{4+}\,\frac{1}{1+}\,\frac{1}{3+}\,\ldots$$

Определение α можно записать в виде

$$\alpha = 4 + \frac{1}{1+}\,\frac{1}{3+}\,\frac{1}{\alpha}\,. \tag{20}$$

Это соотношение приводит к квадратному уравнению для α. Чтобы найти это уравнение, заметим, что оно является частным случаем соотношения (13) с $\alpha_{n+1} = \alpha$. Следовательно, мы можем воспользоваться формулой (14), из которой вытекает, что

$$\alpha = \frac{19\alpha + 5}{4\alpha + 1}\,, \tag{21}$$

ибо $\frac{19}{4}$ и $\frac{5}{1}$ — подходящие дроби, предшествующие элементу $\frac{1}{\alpha}$ в (20). Таким образом, квадратное уравнение для α имеет вид

$$4\alpha^2 - 18\alpha - 5 = 0. \tag{22}$$

Наряду с α удобно рассматривать также число β, период которого равен периоду α, записанному в обратном порядке:

$$\beta = 3 + \frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{3+}\,\frac{1}{1+}\,\frac{1}{4+}\,\ldots$$

Соотношение, аналогичное (2), имеет вид

$$\beta = 3 + \frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{\beta}\,.$$

По общей формуле (14) получим

$$\beta = \frac{19\beta + 4}{5\beta + 1} \tag{23}$$

(здесь подходящие дроби равны $\frac{19}{5}$ и $\frac{4}{1}$). Следовательно, β удовлетворяет квадратному уравнению

$$5\beta^2 - 18\beta - 4 = 0. \tag{24}$$

Ясно, что это уравнение тесно связано с уравнением (22), которому удовлетворяет α. В самом деле, уравнение (22) можно преобразовать в уравнение (24), положив $\alpha = -\frac{1}{\beta}$. Следовательно, число $-\frac{1}{\beta}$ является одним из корней квадратного уравнения (22). Оно не может равняться α, так как α и β положительны, а $-\frac{1}{\beta}$ отрицательно. Значит, $-\frac{1}{\beta}$ — *второй* корень уравнения (22). Этот второй корень называют *алгебраически сопряженным* с α, или просто сопряженным с α. Обозначая число, сопряженное с α, через α', получим $\alpha' = -\frac{1}{\beta}$.

В действительности приведенное рассуждение является вполне общим. В случае чисто периодической непрерывной дроби, скажем

$$\alpha = q_0 + \frac{1}{q_1+}\,\ldots\,\frac{1}{q_n+}\,\frac{1}{\alpha}\,,$$

уравнение, соответствующее (21), имеет вид

$$\alpha = \frac{A_n\alpha + A_{n-1}}{B_n\alpha + B_{n-1}}\,.$$

Если β — число, полученное обращением периода α, то соответствующее уравнение для β имеет вид

$$\frac{A_n\beta + B_n}{A_{n-1}\beta + B_{n-1}} = \beta.$$

Это следует из того, что значение $[q_0, q_1, \ldots, q_n]$ не изменяется,

если элементы q_0, q_1, ..., q_n берутся в обратном порядке (п. 3). Квадратные уравнения для α и для β связаны между собой так же, как и раньше, поэтому число $-\frac{1}{\beta}$ сопряжено с α. Так как β больше 1, число $-\frac{1}{\beta}$ лежит между -1 и 0. Таким образом, *любая чисто периодическая непрерывная дробь равна некоторой квадратичной иррациональности α; число α больше 1, а сопряженное с α число лежит между -1 и 0 и равно $-\frac{1}{\beta}$, где β определяется непрерывной дробью с периодом, равным периоду непрерывной дроби для α, записанному в обратном порядке.*

Замечательно, что это простое свойство полностью характеризует числа, представляемые чисто периодическими непрерывными дробями; как мы сейчас докажем, любое квадратично иррациональное число, удовлетворяющее этому условию, разлагается в чисто периодическую непрерывную дробь. Это впервые точно доказал Галуа (Galois) в 1828 году, хотя неявно результат содержался и в более ранней работе Лагранжа (Lagrange).

Будем называть квадратично иррациональное число α *приведенным*, если $\alpha > 1$ и если сопряженное к α, обозначаемое α', удовлетворяет условию $-1 < \alpha' < 0$. Наша задача — доказать, что непрерывная дробь для такого α — чисто периодическая. Это доказательство, конечно, является более трудным, чем доказательство предыдущего результата, в котором мы исходили из непрерывной дроби; более того, это доказательство нельзя удовлетворительно иллюстрировать на численном примере.

Начнем с исследования вида приведенной квадратичной иррациональности. Мы знаем, что α удовлетворяет некоторому квадратному уравнению $a\alpha^2 + b\alpha + c = 0$ с целыми a, b, c. Решив это уравнение, можно выразить α в виде

$$\alpha = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{P \pm \sqrt{D}}{Q},$$

где P и Q — целые числа, а D — положительное целое число, не являющееся точным квадратом. Можно считать, что перед $\sqrt{D}$ стоит знак плюс, так как знак минус можно было бы заменить на плюс, изменив знаки у чисел P и Q. Итак,

$$\alpha = \frac{P + \sqrt{D}}{Q}. \tag{25}$$

Тогда сопряженное к α число α', которое является вторым корнем квадратного уравнения, задается формулой

$$\alpha' = \frac{P - \sqrt{D}}{Q}.$$

Заметим, что $\dfrac{P^2 - D}{Q} = \dfrac{b^2 - (b^2 - 4ac)}{2a} = 2c$, следовательно, $P^2 - D$ кратно Q.

Так как α предполагается *приведенным*, мы имеем $\alpha > 1$ и $-1 < \alpha' < 0$. Это значит, что

(I) $\alpha - \alpha' > 0$, т. е. $\dfrac{\sqrt{D}}{Q} > 0$, откуда $Q > 0$;

(II) $\alpha + \alpha' > 0$, т. е. $\dfrac{P}{Q} > 0$, откуда $P > 0$;

(III) $\alpha' < 0$, т. е. $P < \sqrt{D}$;

(IV) $\alpha > 1$, т. е. $Q < P + \sqrt{D} < 2\sqrt{D}$.

Таким образом, приведенная квадратичная иррациональность представима в виде (25) с натуральными P и Q, которые удовлетворяют*) условиям

$$P < \sqrt{D}, \qquad Q < 2\sqrt{D}; \tag{26}$$

кроме того, $P^2 - D$ кратно Q.

Будем разлагать α в непрерывную дробь. Первый шаг в процессе разложения — представить α в виде

$$\alpha = q_0 + \frac{1}{\alpha_1}, \tag{27}$$

где q_0 — целая часть α и $\alpha_1 > 1$. Легко видеть, что α_1 снова является приведенной квадратичной иррациональностью; действительно, равенство (27) показывает, что сопряженные к α и α_1 связаны аналогичным соотношением: $\alpha' = q_0 + \dfrac{1}{\alpha_1'}$. Поэтому $\alpha_1' = -\dfrac{1}{q_0 - \alpha'}$; так как α' отрицательно и q_0 — натуральное число, мы имеем $q_0 - \alpha' > 1$, значит, α_1' лежит между -1 и 0.

*) Не каждое α, удовлетворяющее этим условиям, является приведенным, так как из этих условий не следует, что $\alpha' > -1$.

Аналогично все последующие полные частные α_2, α_3, ... — приведенные квадратичные иррациональности.

Из выражения для α_1 следует, что

$$\frac{1}{\alpha_1} = \alpha - q_0 = \frac{P + \sqrt{D}}{Q} - q_0 = \frac{P - Qq_0 + \sqrt{D}}{Q}.$$

Положим $P_1 = -P + Qq_0$. Тогда

$$\alpha_1 = \frac{Q}{-P_1 + \sqrt{D}} = \frac{P_1 + \sqrt{D}}{Q_1},$$

где Q_1 определяется равенством

$$D - P_1^2 = QQ_1. \tag{28}$$

Заметим, что Q_1 — целое, так как $P^2 - D$ кратно Q и $P_1 \equiv -P \pmod{Q}$. Мы имеем

$$\alpha_1 = \frac{P_1 + \sqrt{D}}{Q_1}, \tag{29}$$

и так как α_1 — приведенное, целые P_1 и Q_1 положительны и удовлетворяют условиям (26). Более того, $P_1^2 - D$ кратно Q_1 в силу равенства (28).

Теперь можно рассмотреть вопрос о том, как процесс разложения в непрерывную дробь продолжается дальше. На следующем шагу мы начнем с α_1, вместо α, но процесс будет тем же самым. Вообще, каждое полное частное имеет вид

$$\alpha_n = \frac{P_n + \sqrt{D}}{Q_n},$$

где P_n и Q_n суть натуральные числа, удовлетворяющие (26) и обладающие тем свойством, что $P_n^2 - D$ кратно Q_n. В силу (26) для P_n и Q_n имеется лишь конечное число возможностей, так что, в конце концов, мы придем к некоторой паре значений, уже встречавшейся ранее. Тем самым мы придем к некоторому уже встречавшемуся полному частному; не позднее, чем с этого места, начинается период непрерывной дроби.

Нужно еще доказать, что непрерывная дробь будет *чисто* периодической, т. е. что ее период начинается с первого элемента. Для этого мы покажем, что если $\alpha_n = \alpha_m$, то $\alpha_{n-1} = \alpha_{m-1}$; тем самым мы сможем вернуться к началу непрерывной дроби. Доказательство основано на том, что неполные частные q_n мож-

но сопоставить не только полным частным α_n, но и (аналогичным способом) их сопряженным. Соотношение между любым полным частным и следующим за ним полным частным имеет вид $\alpha_n = q_n + \dfrac{1}{\alpha_{n+1}}$. То же соотношение связывает и их сопряженные, так что $\alpha'_n = q_n + \dfrac{1}{\alpha'_{n+1}}$. Но каждое сопряженное лежит между -1 и 0; введем для $-\dfrac{1}{\alpha'_{n+1}}$ символ β_n. Тогда каждое из чисел β_n больше 1. Последнее соотношение принимает вид

$$-\frac{1}{\beta_n} = q_n - \beta_{n+1},$$

или

$$\beta_{n+1} = q_n + \frac{1}{\beta_n}.$$

Из этого соотношения следует, что q_n, являясь целой частью α_n, может в то же время интерпретироваться и как целая часть β_{n+1}.

Пусть теперь α_n и α_m — два равных полных частных и $m < n$. Тогда их сопряженные α'_n и α'_m также равны и потому $\beta_n = \beta_m$. В силу только что доказанного q_{n-1} есть целая часть β_n, а q_{m-1} — целая часть β_m. Следовательно, $q_{n-1} = q_{m-1}$. Но

$$\alpha_{n-1} = q_{n-1} + \frac{1}{\alpha_n}, \quad \alpha_{m-1} = q_{m-1} + \frac{1}{\alpha_m},$$

значит,

$$\alpha_{n-1} = \alpha_{m-1}.$$

Повторив рассуждение, получим $\alpha_{n-2} = \alpha_{m-2}$ и так далее до тех пор, пока мы не установим, что α_{n-m} таково же, как и само α. Полагая $n - m = r$, находим

$$\alpha = q_0 + \frac{1}{q_1+} \cdots \frac{1}{q_{r-1}+} \frac{1}{\alpha},$$

а это показывает, что α разлагается в чисто периодическую непрерывную дробь. Таким образом, мы установили основной результат этого пункта: чисто периодические непрерывные дроби являются приведенными квадратичными иррациональностями, и обратно, всякая приведенная квадратичная иррациональность разлагается в чисто периодическую непрерывную дробь.

Теперь можно объяснить специфический вид непрерывной дроби для $\sqrt{N}$ (N — натуральное число, не являющееся точным квадратом), который мы наблюдали в таблице. Непрерывная дробь для $\sqrt{N}$ не может быть, конечно, чисто периодической, так как число, сопряженное к $\sqrt{N}$, равно $-\sqrt{N}$ и не лежит между -1 и 0. Рассмотрим число $\sqrt{N}+q_0$, где q_0 — целая часть $\sqrt{N}$. Сопряженным к этому числу является число $-\sqrt{N}+q_0$, лежащее между -1 и 0. Значит, непрерывная дробь для $\sqrt{N}+q_0$ чисто периодическая, а так как эта дробь, очевидно, начинается с $2q_0$, она имеет вид

$$\sqrt{N}+q_0 = 2q_0 + \frac{1}{q_1+}\cdots\frac{1}{q_n+}\,\frac{1}{2q_0+}\cdots \tag{30}$$

Согласно уже установленному в этом пункте, непрерывная дробь

$$q_n + \frac{1}{q_{n-1}+}\cdots\frac{1}{q_1+}\,\frac{1}{2q_0+}\,\frac{1}{q_n+}\cdots,$$

получаемая из α обращением периода, равна $-\dfrac{1}{\alpha'}$, где $\alpha=\sqrt{N}+$ $+\,q_0$. Далее $\alpha' = -\sqrt{N}+q_0$, значит,

$$-\frac{1}{\alpha'} = \frac{1}{\sqrt{N}-q_0} = q_1 + \frac{1}{q_2+}\cdots\frac{1}{q_n+}\,\frac{1}{2q_0+}\cdots$$

ввиду (30). Сравнивая последние две непрерывные дроби (и вспоминая, что разложение числа в непрерывную дробь однозначно), мы видим, что $q_n = q_1$, $q_{n-1} = q_2$, ... Следовательно, *непрерывная дробь для* $\sqrt{N}$ *всегда имеет вид*

$$q_0, \overline{q_1, q_2, \ldots, q_2, q_1, 2q_0}.$$

Период этой дроби начинается сразу же после первого элемента q_0 *и состоит из симметричной части* $q_1, q_2, \ldots, q_2, q_1$, *после которой следует* $2q_0$. Симметричная часть может иметь или не иметь центрального элемента; например, в $\sqrt{54} = 7, \overline{2,1,6,1,2,14}$ имеется центральный элемент, в то время как в $\sqrt{53} = 7, \overline{3,1,1,3,14}$ его нет.

10. Теорема Лагранжа. Мы можем доказать теперь общую теорему Лагранжа о том, что *всякая* квадратичная иррациональность разлагается в непрерывную дробь, которая, начи-

ная с некоторого места, будет периодической. Достаточно показать, что, разлагая любую квадратичную иррациональность α в непрерывную дробь, мы, в конце концов, достигнем полного частного α_n, являющегося приведенной квадратичной иррациональностью; тогда, начиная с этого места, непрерывная дробь будет периодической.

Соотношение между α и каким-либо его полным частным дается известной формулой (14):

$$\alpha = \frac{\alpha_{n+1}A_n + A_{n-1}}{\alpha_{n+1}B_n + B_{n-1}}.$$

Так как α и α_{n+1} — квадратичные иррациональности, а A_n, B_n, A_{n-1}, B_{n-1} — целые (даже натуральные) числа, то такое же соотношение имеется и между α' и α'_{n+1}. Выражая из него α'_{n+1} через α', получаем

$$\alpha'_{n+1} = -\frac{B_{n-1}\alpha' - A_{n-1}}{B_n\alpha' - A_n} = -\frac{B_{n-1}}{B_n}\left(\frac{\alpha' - A_{n-1}/B_{n-1}}{\alpha' - A_n/B_n}\right).$$

Какие сведения о величине α'_{n+1} при больших n дает это соотношение? И $\dfrac{A_n}{B_n}$, и $\dfrac{A_{n-1}}{B_{n-1}}$ стремятся при бесконечном возрастании n к α, следовательно, стоящая в скобках дробь стремится к 1. Числа B_{n-1} и B_n положительны, поэтому α'_{n+1} начиная с некоторого места, становится отрицательным. Числа $\dfrac{A_n}{B_n}$ через одно больше или меньше α, поэтому стоящая в скобках дробь один раз несколько меньше, а в следующий раз несколько больше 1. Если выбрать значение n, для которого эта дробь немного меньше 1, и заметить, что $B_{n-1} < B_n$, то получится, что α'_{n+1} лежит между -1 и 0. Для такого значения n число α_{n+1} является приведенной квадратичной иррациональностью. Следовательно, начиная с этого шага, непрерывная дробь будет чисто периодической. Тем самым теорема Лагранжа доказана.

Лишь для немногих иррациональных чисел, не являющихся квадратичными иррациональностями, известны какие-нибудь закономерности, которым подчинены их непрерывные дроби. Одним из таких примеров является число $\dfrac{e-1}{e+1}$, где $e=2{,}71828\ldots$ есть основание натуральных логарифмов. Непрерывная дробь

в этом случае имеет вид

$$\frac{e-1}{e+1} = \frac{1}{2+}\,\frac{1}{6+}\,\frac{1}{10+}\,\frac{1}{14+}\,\ldots,$$

ее элементы образуют арифметическую прогрессию. Более того, если k — любое положительное число, то

$$\frac{e^{2/k}-1}{e^{2/k}+1} = \frac{1}{k+}\,\frac{1}{3k+}\,\frac{1}{5k+}\,\frac{1}{7k+}\,\ldots$$

Эти результаты были получены Эйлером в 1737 году. Непрерывная дробь для e несколько более сложна:

$$e = 2 + \frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{6+}\,\ldots;$$

здесь числа 2, 4, 6, ... встречаются после каждой пары 1, 1. Этот факт также установил Эйлер.

Очень мало известно о непрерывных дробях для алгебраических чисел, не являющихся квадратичными иррациональностями. Нам неизвестно, например, ограничены или нет элементы непрерывной дроби для $\sqrt[3]{2}$:

$$\sqrt[3]{2} = 1 + \frac{1}{3+}\,\frac{1}{5+}\,\frac{1}{1+}\,\frac{1}{1+}\,\frac{1}{4+}\,\frac{1}{1+}\,\ldots,$$

и не видно никакого метода, с помощью которого можно было бы взяться за такую задачу. Известны некоторые результаты о диофантовых приближениях алгебраических чисел (VII, 5), из которых следует, что элементы непрерывных дробей для таких чисел не могут расти быстрее, чем с некоторой скоростью. Но результаты, найденные этим способом, вероятно, далеки от действительно имеющих место.

11. Уравнение Пелля (Pell). Это уравнение

$$x^2 - Ny^2 = 1 \quad \text{или} \quad x^2 = Ny^2 + 1, \tag{31}$$

где N — натуральное число, не являющееся точным квадратом. Уравнение (31) не представляет интереса, если N — точный квадрат, так как разность двух точных квадратов не может быть равна 1, если исключить случай $1^2 - 0^2$. Замечательно, что уравнение Пелля всегда имеет решение с натуральными x и y; в действительности оно имеет даже бесконечно много таких

решений.

Разрозненные упоминания о частных случаях уравнения Пелля встречаются на протяжении всей истории математики. Наиболее любопытна так называемая задача о скоте (Cattle problem) Архимеда, опубликованная Лессингом (Lessing) в 1773 году по рукописи из библиотеки Вольфенбюттеля (Wolfenbüttel). Утверждают, что эту задачу Архимед предложил Эратосфену; большинство экспертов, исследовавших этот вопрос, пришли к заключению, что задачу действительно поставил Архимед. Эта задача содержит восемь неизвестных (число животных различных видов), удовлетворяющих семи линейным уравнениям и двум условиям, в которых требуется, чтобы некоторые числа были точными квадратами. После элементарных алгебраических преобразований задача сводится к решению уравнения

$$t^2 - 4\,729\,494\,u^2 = 1.$$

В наименьшем решении этого уравнения, найденном Амтором (Amthor) в 1880 году, u содержит 41 цифру. Наименьшее решение исходной задачи, полученное отсюда, состоит из чисел с сотнями тысяч знаков. Античные математики, конечно, не могли решить этой задачи, но уже сам факт ее постановки говорит о том, что они обладали какими-то не сохранившимися сведениями об уравнении Пелля.

В более поздние времена первый систематический метод решения уравнения Пелля дал лорд Браункер *) в 1657 году. Это был, по существу, метод разложения $\sqrt{N}$ в непрерывную дробь, который будет объяснен ниже. Примерно в то же время де Бесси (Frenicle de Bessy) в не сохранившейся работе составил таблицы решений уравнения (31) для всех N вплоть до 150 и предложил Браункеру решить уравнение

$$x^2 - 313y^2 = 1.$$

Браункер дал решение (в котором x состоит из шестнадцати цифр), отметив, что нашел это решение своим методом за час

*) Вильям Браункер (William Brouncker (1620—1684)) в качестве второго виконта Браункера получил в наследство от своего отца замок Лионс (Lyons) в Ирландии в 1667 году. Читатели *«Дневника»* помнят, что Пепи (Pepys) был низкого мнения о его моральном облике. Однако математические исследования лорда Браункера делают ему честь.

или два. Валлис (Wallis), излагая метод Браункера, и Ферма, комментируя работу Валлиса, утверждали, что могут доказать разрешимость уравнения Пелля. Кажется, Ферма первый ясно высказал утверждение о том, что это уравнение имеет бесконечно много решений. Первым опубликованным доказательством было доказательство Лагранжа, появившееся около 1766 года. Эйлер по недоразумению приписал этому уравнению имя Пелля; Эйлер думал, что предложенный Валлисом метод решения этого уравнения принадлежит Пеллю (John Pell) — английскому математику того же времени.

Решение уравнения Пелля легко может быть получено в терминах непрерывной дроби для $\sqrt{N}$. Мы видели в п. 9, что

$$\sqrt{N} = q_0 + \frac{1}{q_1+} \dots \frac{1}{q_n+} \frac{1}{2q_0+} \frac{1}{q_1+} \dots$$

(Мы видели также, что $q_n = q_1$ и так далее, но здесь это для нас неважно.) Пусть теперь $\dfrac{A_{n-1}}{B_{n-1}}$ и $\dfrac{A_n}{B_n}$ — две подходящие дроби, стоящие непосредственно перед элементом $2q_0$, именно:

$$\frac{A_{n-1}}{B_{n-1}} = q_0 + \frac{1}{q_1+} \dots \frac{1}{q_{n-1}}; \quad \frac{A_n}{B_n} = q_0 + \frac{1}{q_1+} \dots \frac{1}{q_{n-1}+} \frac{1}{q_n}.$$

По формуле (14) имеем

$$\sqrt{N} = \frac{\alpha_{n+1}A_n + A_{n-1}}{\alpha_{n+1}B_n + B_{n-1}},$$

где α_{n+1} — полное частное, стоящее за q_n, т. е.

$$\alpha_{n+1} = 2q_0 + \frac{1}{q_1+} \dots = \sqrt{N} + q_0.$$

Подставляя это значение для α_{n+1} и производя умножение, получаем

$$\sqrt{N}(\sqrt{N} + q_0)B_n + \sqrt{N}B_{n-1} = (\sqrt{N} + q_0)A_n + A_{n-1}.$$

Так как $\sqrt{N}$ иррационально, а все остальные числа — целые, то из этого уравнения следуют равенства

$$NB_n = q_0A_n + A_{n-1},$$
$$q_0B_n + B_{n-1} = A_n.$$

Их можно рассматривать как выражения A_{n-1} и B_{n-1} в терми-

нах A_n и B_n:

$$A_{n-1} = NB_n - q_0 A_n, \quad B_{n-1} = A_n - q_0 B_n.$$

Подставив это в (10), получим

$$A_n(A_n - q_0 B_n) - B_n(NB_n - q_0 A_n) = (-1)^{n-1}$$

или

$$A_n^2 - NB_n^2 = (-1)^{n-1}. \tag{32}$$

Значит, $x = A_n$ и $y = B_n$ являются решением уравнения

$$x^2 - Ny^2 = (-1)^{n-1}.$$

Если n нечетно, то мы получим решение уравнения Пелля. Если n четно, заметим, что приведенное рассуждение можно применить к двум подходящим дробям, стоящим в конце следующего периода. Так как элемент q_n, который встретится во втором случае, равен q_{2n+1}, если нумеровать элементы с начала, то в (32) следует заменить n на $2n+1$; это дает

$$A_{2n+1}^2 - NB_{2n+1}^2 = (-1)^{2n} = 1,$$

так что (в любом случае) уравнение (31) разрешимо в натуральных числах.

Проиллюстрируем это построение двумя численными примерами, одним — для нечетного n, другим — для четного n. Возьмем сначала $N = 21$. Непрерывная дробь (см. табл. I) такова:

$$\sqrt{21} = 4, \overline{1, 1, 2, 1, 1, 8}$$

и $n = 5$. Подходящие равны

$$\frac{4}{1}, \frac{5}{1}, \frac{9}{2}, \frac{23}{5}, \frac{32}{7}, \frac{55}{12}, \ldots$$

и $x = 55$, $y = 12$ — решение уравнения

$$x^2 - 21y^2 = 1.$$

Возьмем далее $N = 29$. Непрерывная дробь имеет вид

$$\sqrt{29} = 5, \overline{2, 1, 1, 2, 10};$$

$n = 4$. Подходящие равны

$$\frac{5}{1}, \frac{11}{2}, \frac{16}{3}, \frac{27}{5}, \frac{70}{13}, \ldots$$

и $x = 70$, $y = 13$ дает решение уравнения

Таблица I

№	Непрерывная дробь для $\sqrt{N}$	x	y	$x^2 - Ny^2$
2	1; 2	1	1	−1
3	1; 1, 2	2	1	+1
5	2; 4	2	1	−1
6	2; 2, 4	5	2	+1
7	2; 1, 1, 1, 4	8	3	+1
8	2; 1, 4	3	1	+1
10	3; 6	3	1	−1
11	3; 3, 6	10	3	+1
12	3; 2, 6	7	2	+1
13	3; 1, 1, 1, 1, 6	18	5	−1
14	3; 1, 2, 1, 6	15	4	+1
15	3; 1, 6	4	1	+1
17	4; 8	4	1	−1
18	4; 4, 8	17	4	+1
19	4; 2, 1, 3, 1, 2, 8	170	39	+1
20	4; 2, 8	9	2	+1
21	4; 1, 1, 2, 1, 1, 8	55	12	+1
22	4; 1, 2, 4, 2, 1, 8	197	42	+1
23	4; 1, 3, 1, 8	24	5	+1
24	4; 1, 8	5	1	+1
26	5; 10	5	1	−1
27	5; 5, 10	26	5	+1
28	5; 3, 2, 3, 10	127	24	+1
29	5; 2, 1, 1, 2, 10	70	13	−1
30	5; 2, 10	11	2	+1
31	5; 1, 1, 3, 5, 3, 1, 1, 10	1520	273	+1
32	5; 1, 1, 1, 10	17	3	+1
33	5; 1, 2, 1, 10	23	4	+1
34	5; 1, 4, 1, 10	35	6	+1
35	5; 1, 10	6	1	+1
37	6; 12	6	1	−1
38	6; 6, 12	37	6	+1
39	6; 4, 12	25	4	+1
40	6; 3, 12	19	3	+1
41	6; 2, 2, 12	32	5	−1
42	6; 2, 12	13	2	+1
43	6; 1, 1, 3, 1, 5, 1, 3, 1, 1, 12	3482	531	+1
44	6; 1, 1, 1, 2, 1, 1, 1, 12	199	30	+1
45	6; 1, 2, 2, 2, 1, 12	161	24	+1
46	6; 1, 3, 1, 1, 2, 6, 2, 1, 1, 3, 1, 12	24335	3588	+1
47	6; 1, 5, 1, 12	48	7	+1
48	6; 1, 12	7	1	+1
50	7; 14	7	1	−1

$$x^2 - 29y^2 = -1.$$

Чтобы получить решение уравнения с 1 вместо -1, продолжим ряд подходящих до $\frac{A_9}{B_9}$ (здесь $2n+1=9$). Имеем $\frac{A_4}{B_4} = \frac{70}{13}$, следующие подходящие дроби имеют вид:

$$\frac{727}{135}, \frac{1524}{283}, \frac{2251}{418}, \frac{3775}{701}, \frac{9801}{1820}.$$

Отсюда найдем, что $x = 9801$, $y = 1820$ — решение уравнения

$$x^2 - 29y^2 = 1.$$

Можно доказать, что описанный выше процесс приводит всегда к наименьшему решению. Наименьшие решения уравнения $x^2 - Ny^2 = \pm 1$ вплоть до $N = 50$ приведены в табл. I.

Использованные в этом пункте методы позволяют установить еще некоторые факты, относящиеся к уравнению Пелля. Во-первых, это уравнение имеет бесконечно много решений, и все его решения получаются из подходящих дробей, соответствующих элементам q_n в конце каждого периода. Если n нечетно, т. е. у непрерывной дроби есть средний элемент (как в примере с $\sqrt{21}$), все эти решения суть решения уравнения с $+1$. Если же n четно, т. е. если среднего элемента нет (как в примере с $\sqrt{29}$), то выбранные так подходящие дроби попеременно дают решения уравнений с -1 и $+1$.

Все решения могут быть получены из первого решения и прямым вычислением, без дальнейшего разложения в непрерывную дробь. Если x_0, y_0 — наименьшее решение уравнения $x^2 - Ny^2 = \pm 1$, задаваемое подходящей дробью $\frac{A_n}{B_n}$, то общее решение этого уравнения получается по формуле

$$x + y\sqrt{N} = (x_0 + y_0\sqrt{N})^r,$$

где $r = 1, 2, 3, \ldots$ Так, в примере с $\sqrt{29}$ находим

$$9801 + 1820\sqrt{29} = (70 + 13\sqrt{29})^2.$$

Различие случаев четного и нечетного n приводит к задачам, полное решение которых до сих пор неизвестно. Например, нет способа полностью охарактеризовать числа N, для которых n четно. Если уравнение $x^2 - Ny^2 = -1$ разрешимо, то разрешимо и сравнение $x^2 + 1 \equiv 0 \pmod{N}$. Отсюда следует, что в этом

случае N не может делиться ни на 4, ни на простое вида $4k+3$ (III, 3). Как мы увидим позже (VI, 5), такое N представимо в виде u^2+v^2 с взаимно простыми u и v. Это условие является необходимым условием разрешимости уравнения $x^2 - Ny^2 = -1$, но это условие не достаточное; например, для числа $N = 34$ указанное условие выполнено, однако уравнение $x^2 - Ny^2 = -1$ неразрешимо.

Решения более общего уравнения

$$x^2 - Ny^2 = \pm M,$$

где M — целое положительное, меньшее чем $\sqrt{N}$, тесно связаны с разложением $\sqrt{N}$ в непрерывную дробь. Можно доказать, что *каждое решение любого из таких уравнений порождается некоторой подходящей дробью в непрерывной дроби для* $\sqrt{N}$.

12. Геометрическая интерпретация непрерывных дробей. Замечательную геометрическую интерпретацию непрерывной дроби иррационального числа предложил в 1895 году Клейн (F. Klein). Пусть α — иррациональное число, предполагаемое для простоты положительным. Рассмотрим всевозможные точки плоскости с целыми положительными координатами и предположим, что в этих точках расставлены колышки. Прямая $y = \alpha x$ ни через один из этих колышков не пройдет. Представим себе, что вдоль этой прямой натянута веревка, один конец которой закреплен в бесконечно удаленной точке этой прямой. Если другой конец веревки, расположенный в начале координат, отвести в сторону, веревка зацепится за некоторые колышки; если его отвести в другую сторону, то веревка зацепится за какие-то другие колышки. Колышки (лежащие под прямой) расположены в точках с координатами (B_0, A_0), (B_2, A_2) ... соответствующих подходящим дробям, меньшим α. Другой ряд колышков (над прямой) состоит из точек с координатами (B_1, A_1), (B_3, A_3), ... соответствующих подходящим дробям, большим α. Каждое из положений веревки образует ломаную линию, приближающуюся к прямой $y = \alpha x$.

Диаграмма иллюстрирует случай

$$\alpha = \sqrt{3} = 1 + \frac{1}{1+}\,\frac{1}{2+}\,\frac{1}{1+}\,\frac{1}{2+}\,\ldots$$

Подходящие дроби здесь равны

$$\frac{1}{1}, \frac{2}{1}, \frac{5}{3}, \frac{7}{4}, \frac{19}{11}, \frac{26}{15}, \ldots$$

Колышки под прямой расположены в точках (1, 1), (3, 5), (11, 19), ..., а колышки над прямой — в точках (1, 2), (4, 7), (15, 26), ...

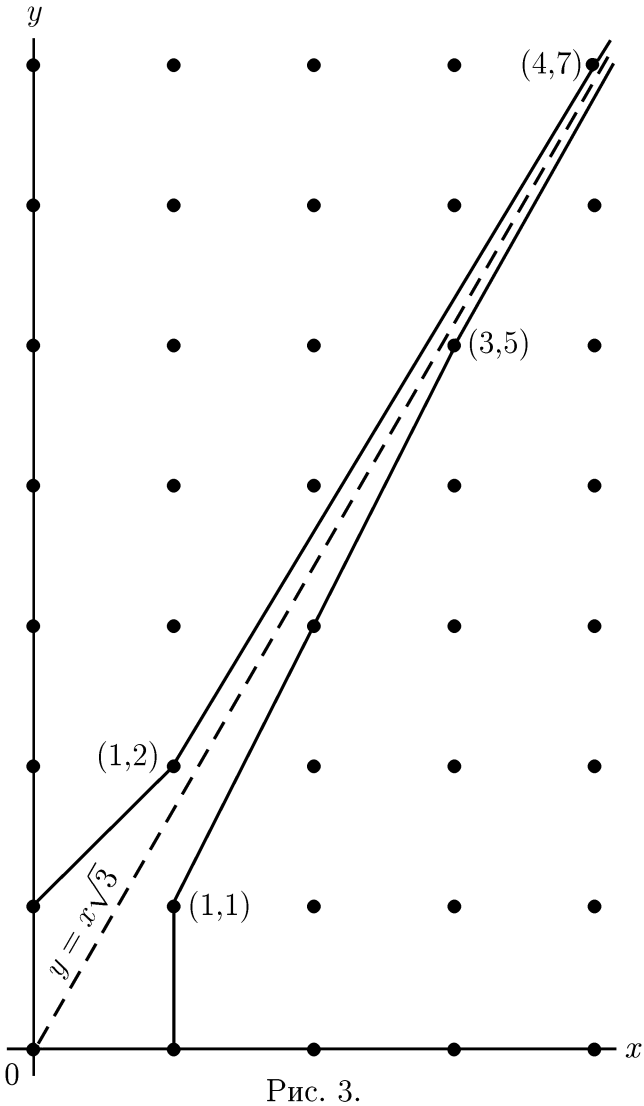

Рис. 3.

Большинство элементарных теорем о непрерывных дробях имеют простую геометрическую интерпретацию. Обозначим через P_n точку $(A_n,\ B_n)$. Рекуррентные соотношения (8) и (9) по-

казывают, что вектор от P_{n-2} до P_n (P_{n-2} и P_n — две последовательные вершины одной из ломаных) равен целому кратному вектора от начала O до P_{n-1}. Соотношение (10) можно интерпретировать как утверждение о том, что площадь треугольника $P_{n-1}P_n$ при любом n равна $\frac{1}{2}$. Это утверждение можно вывести непосредственно из описанной выше конструкции с веревкой: в самом деле, очевидно, что в треугольнике $OP_{n-1}P_n$ нет ни одной целой точки, помимо вершин; кроме того, можно легко доказать, что площадь любого треугольника, обладающего этим свойством, равна $\frac{1}{2}$.

Замечания к главе IV. Лучшее из имеющихся на английском языке изложений теории непрерывных дробей содержится в книге Кристаля (см. Chrystal, Algebra, vol. II, chs. 32—34). На русском языке эта теория излагается в книге А. Я. Хинчина «*Цепные дроби*», Физматгиз, 1961 *). Типичной работой по этому предмету является работа Перрона (Perron, Die Lehre von den Kettenbrüchen, Teubner, 1929). Доказательства различных результатов, приведенных в этой главе без доказательств, можно найти у Кристаля или Перрона. О диофантовых приближениях читатель может справиться в книгах Перрона, Нивена или Касселса (см. Perron, Irrationalzahlen, Göschens Lehrbücherei, vol. I, 1947; Niven, Irrational Numbers, Carus Math. Monographs, № 11, 1956; Cassels, Introduction to Diophantine Approximation (Cambridge Math. Tracts, № 45, 1957)) и в книге Хинчина*).

п. 1—6. Практически вся эта теория принадлежит Эйлеру.

п. 7. См. ([9], ch. 11); Perron, § 8 или А. Я. Хинчин, гл. II*).

п. 8. Ссылки на таблицы можно найти у Перрона, стр. 100 или в ([7], vol. II, ch. 12). Об удобных методах вычисления непрерывных дробей для квадратичных иррациональностей, см ([7], p. 372).

п. 10. Вывод непрерывной дроби для e и пр. см. у Перрона (Perron, §§ 31 и 64) или в заметке Дэвиса (C. S. Davis, J. London Math. Soc., **20** (1945), 194—198).

п. 11. Относительно задачи о скоте см. Thomas Heath, Diophantus of Alexandria (Cambridge, 1910), pp. 121—124 и ([7], vol. II, pp. 342—345).

п. 12. См. F. Klein, Ausgewahlte Kapitel der Zahlentheorie (Teubner, 1907), pp. 17—25. Идея, кажется, принадлежит Смиту (см. H. J. S. Smith, Collected Math. Papers, vol. 2, pp. 146—147).

---

*) Ссылки на книгу А. Я. Хинчина добавлены мной. (*Прим. перев.*)

ГЛАВА V

СУММЫ КВАДРАТОВ

1. Числа, представимые в виде суммы двух квадратов. Вопрос о представимости чисел в виде суммы двух квадратов — очень старый вопрос; он рассматривается еще в «*Арифметике*» Диофанта (около 250 года нашей эры), но точный смысл утверждений Диофанта неясен. Правильный ответ на этот вопрос впервые дали немецкий математик Жирар (Albert Girard) в 1625 году и немного позднее Ферма. Возможно, что у Ферма имелось доказательство его результата, но первым из известных нам доказательств является доказательство Эйлера, опубликованное в 1749 году.

Легко установить, что некоторые числа не представимы в виде суммы двух квадратов. Во-первых, квадрат любого четного числа сравним с 0 по mod 4, а квадрат любого нечетного числа сравним с 1 по mod 4. Отсюда следует, что сумма любых двух квадратов сравнима с $0+0$, или с $0+1$, или с $1+1$ по mod 4, т. е. с 0, 1 или 2 по mod 4. Таким образом, ни одно число вида $4k+3$ не представимо суммой двух квадратов.

Но мы можем пойти дальше. Пусть число N имеет простой множитель q вида $4k+3$; уравнение $x^2+y^2=N$ влечет сравнение $x^2 \equiv -y^2 \pmod{N}$, а так как -1 является квадратичным невычетом по модулю q, это сравнение разрешимо лишь при $x \equiv 0 \pmod{q}$ и $y \equiv 0 \pmod{q}$. Значит, x и y делятся на q, следовательно, N делится на q^2 и уравнение $x^2+y^2=N$ можно сократить на q^2. Если $N=q^2N_1$ и N_1 еще делится на q, то такое же рассуждение показывает, что N_1 делится на q^2 и так далее; таким образом мы находим, что точная степень q, делящая N, должна быть четной. Итак, *всякое число, представимое в виде суммы двух квадратов, содержит в своем разложении лишь четные степени простых вида* $4k+3$. Прежнее условие состо-

яло в том, что N не должно иметь вид $4k+3$. Указанное теперь условие сильнее, так как всякое число вида $4k+3$ содержит хотя бы один простой множитель вида $4k+3$ в нечетной степени.

Если отбросить числа, которые, согласно этому условию, заведомо не представили суммой двух квадратов, то ряд оставшихся чисел начинается с 1, 2, 4, 5, 8, 9, 10, 13, 16, 17, 18, 20, ...; читатель путем последовательных проб может убедиться, что каждое из этих чисел представляется в виде суммы двух целых квадратов. Это верно и в общем случае: необходимое и достаточное условие представимости числа N в виде суммы двух квадратов состоит в том, что любой простой множитель N вида $4k+3$ должен входить в N в четной степени.

Докажем теперь это утверждение. Важную роль в доказательстве играет тождество, представляющее произведение двух сумм квадратов в виде суммы двух квадратов. Это тождество

$$(a^2+b^2)(c^2+d^2)=(ac+bd)^2+(ad-bc)^2 \qquad (1)$$

принадлежит Леонардо из города Пиза (Pisa) (его называют также Фибоначчи); он приводит это тождество в своей «*Книге Абак*» в 1202 году.

Любое число, удовлетворяющее указанным выше условиям, состоит из множителей, равных 2, простых вида $4k+1$ и квадратов простых вида $4k+3$. Повторное применение тождества (1) показывает, что если каждый из таких сомножителей представляется в виде суммы двух квадратов, то и само число также представимо в виде суммы двух квадратов. Число 2 представимо в виде 1^2+1^2; если q есть простое вида $4k+3$, то (q^2 представляется в виде q^2+0^2. Остается доказать, что *любое простое вида* $4k+1$ *представимо в виде* x^2+y^2, этот результат мы установим в следующем пункте. Из него будет следовать, что вышеуказанное условие необходимо и достаточно для того, чтобы рассматриваемое число представлялось в виде суммы двух квадратов.

В нашем изложении допускаются представления x^2+y^2, в которых x и y могут иметь общие множители (например, $q^2=$ $=q^2+0^2$); но это не очень существенно: требование взаимной простоты x и y приводит к результату, мало отличающемуся от предыдущего. Подробнее об этом будет говориться в (VI, 5),

где развита теория, включающая рассматриваемый здесь вопрос как частный случай.

2. Простые вида $4k+1$. Мы приведем здесь классическое доказательство того, что каждое простое p вида $4k+1$ представимо как сумма двух квадратов; это доказательство принадлежит, по существу, Эйлеру. Оно состоит из двух шагов. Первый шаг заключается в установлении того, что некоторое кратное p представимо в виде z^2+1; на втором шагу мы установим, что p представимо в виде x^2+y^2.

Первый шаг эквивалентен доказательству разрешимости сравнения

$$z^2+1 \equiv 0 \pmod{p}$$

для любого простого p вида $4k+1$. Мы уже знаем, что это так, из критерия Эйлера для квадратичных вычетов и невычетов (см. III, 3).

Второй шаг доказательства начинается с уже установленного факта, из которого следует, что существует такое натуральное m, для которого

$$mp = z^2+1.$$

Мы можем, конечно, предполагать, что z лежит между $-\frac{1}{2}p$ и $\frac{1}{2}p$ (этого можно добиться, вычитая из z подходящее кратное p). Предположив это, получаем

$$m = \frac{z^2+1}{p} < \frac{\frac{1}{4}p^2+1}{p} < p.$$

Отсюда, в частности, следует, что существуют такие целые x и y, для которых выполняется равенство

$$mp = x^2+y^2, \tag{2}$$

где m — натуральное число, меньшее p. Покажем теперь, что если $m>1$, то найдется такое натуральное m', меньшее m, для которого уравнение $m'p = x^2+y^2$ все еще разрешимо в целых x и y. Отсюда (если повторить это рассуждение несколько раз) будет следовать, что разрешимо уравнение $p = x^2+y^2$, в котором $m=1$.

Рассуждения проводятся следующим образом. Определим два целых числа u и v, лежащих между $-\frac{1}{2}m$ и $\frac{1}{2}m$ (включи-

тельно, если m четное), которые сравнимы соответственно с x и y по модулю m:

$$u \equiv x \pmod{m}; \quad v \equiv y \pmod{m}. \tag{3}$$

Тогда $u^2 + v^2 \equiv x^2 + y^2 \equiv 0 \pmod{m}$, так что

$$mr = u^2 + v^2 \tag{4}$$

для некоторого целого r. Заметим, что r не может быть равно 0, так как тогда u и v были бы равны 0, а потому x и y были бы кратны m; а это противоречит равенству (2), ибо отсюда и из (2) следует, что простое число p делится на m. Число r удовлетворяет неравенству

$$r = \frac{u^2 + v^2}{m} < \frac{\frac{1}{4}m^2 + \frac{1}{4}m^2}{m} < m.$$

Перемножим равенства (2) и (4), применяя тождество (1). Это дает

$$m^2 rp = (x^2 + y^2)(u^2 + v^2) = (xu + yv)^2 + (xv - yu)^2. \tag{5}$$

Важно отметить, что каждое из чисел $xu + yv$ и $xv - yu$ делится на m. Действительно, ввиду сравнения (3) $xu + yv \equiv$ $\equiv x^2 + y^2 \equiv 0 \pmod{m}$ и $xv - yu \equiv xy - yx \equiv 0 \pmod{m}$. Следовательно, (5) можно разделить на m^2, что дает

$$rp = X^2 + Y^2$$

с некоторыми целыми X и Y. Тем самым мы доказали, что существует такое натуральное число r, меньшее m, для которого rp представимо в виде суммы двух квадратов.

Как мы уже говорили, этого достаточно, чтобы доказать представимость rp в виде суммы двух квадратов. Проиллюстрируем это доказательство на численном примере. Возьмем $p = 277$ — это простое вида $4k + 1$. Мы знаем, что сравнение $1 + z^2 \equiv 0 \pmod{277}$ разрешимо и его решение легко найти подбором или с помощью таблицы индексов. Число $z = 60$ является решением этого сравнения:

$$60^2 + 1 = 3601 = 277 \cdot 13.$$

Исходным пунктом доказательства, как и в общем случае, служит равенство $13 \cdot 277 = 60^2 + 1^2$. Следуя плану доказательства, приводим числа 60 и 1 по модулю 13; получим числа -5 и 1.

Аналогом равенства (4) служит равенство

$$13 \cdot 2 = (-5)^2 + 1^2.$$

Следующий шаг состоит в перемножении двух равенств и применении тождества (1). Получаем $13^2 \cdot 2 \cdot 277 = (60^2+1^2)((-5)^2+ + 1^2) = (60 \cdot (-5) + 1 \cdot 1)^2 + (60 \cdot 1 - 1 \cdot (-5))^2 = (-299)^2 + 65^2$. Числа справа, как это и должно быть, делятся на 13, сокращение приводит к равенству

$$2 \cdot 277 = (-23)^2 + 5^2.$$

Далее этот процесс повторяется. Числа -23 и 5, приведенные по модулю 2, дают 1, соответствующее уравнение имеет вид

$$2 \cdot 1 = 1^2 + 1^2.$$

Перемножая его с предыдущим равенством и применяя тождество (1), получаем

$$2^2 \cdot 1 \cdot 277 = (-23+5)^2 + (-23-5)^2 = (-18)^2 + (-28)^2.$$

И, наконец,

$$277 = 9^2 + 14^2.$$

В связи с доказанной общей теоремой нужно заметить что представление p в виде $x^2 + y^2$ *единственно* (если исключить очевидные возможности замены x на y и изменения их знаков). Ферма, обратив внимание на этот факт, назвал его «фундаментальной теоремой о прямоугольных треугольниках»: отсюда следует, что существует ровно один прямоугольный треугольник, гипотенуза которого равна $\sqrt{p}$, а катеты измеряются натуральными числами.

Доказать единственность представления не трудно. Предположим, что имеет место равенство

$$p = x^2 + y^2 = X^2 + Y^2. \tag{6}$$

Мы знаем, что сравнение $z^2 + 1 \equiv 0 \pmod p$ имеет ровно два решения: $z \equiv \pm h \pmod p$. Значит,

$$x \equiv \pm hy \pmod p \quad \text{и} \quad X \equiv \pm hY \pmod p.$$

Так как знаки чисел x, y, X, Y несущественны, можно считать, что выполнено

$$x \equiv hy \pmod p, \qquad X \equiv hY \pmod p. \tag{7}$$

Перемножим равенства (6) и применим тождество (1). Тогда получим

$$p^2 = (x^2 + y^2)(X^2 + Y^2) = (xX + yY)^2 + (xY - yX)^2.$$

Далее $xY - yX \equiv 0 \pmod{p}$ в силу (7). Значит, оба числа справа делятся на p и равенство можно разделить на p^2. Это дает представление 1 в виде суммы двух квадратов, а такое представление единственно: $1 = (\pm 1)^2 + 0^2$. Таким образом, в предыдущем равенстве одно из чисел $xX + yY$ и $xY - yX$ должно быть равно 0. Если $xY - yX = 0$, то, так как x, y и X, Y взаимно просты, либо $x = X$ и $y = Y$, либо $x = -X$ и $y = -Y$. Аналогично, если $xX + yY = 0$, то либо $x = Y$ и $y = -X$, либо $x = -Y$ и $y = X$. В каждом из этих случаев оба представления в (6), по существу, одинаковы.

3. Конструкция для x и y. Любое простое p вида $4k+1$ однозначно представимо как $x^2 + y^2$; математики пытались найти выражения для x и y в терминах p, так как конструкция обычно приносит большее удовлетворение, чем чистое доказательство существования, хотя граница между ними не всегда отчетливо проводится. Известны четыре конструкции для x и y, они принадлежат Лежандру (1808), Гауссу (1825), Серре (1848) и Якобсталю (1906); мы изложим их, не входя в детали доказательств. Представляет интерес разнообразие методов, используемых в этих конструкциях.

Конструкция Лежандра основана на разложении $\sqrt{p}$ в непрерывную дробь. Это разложение имеет вид (IV, 9)

$$\sqrt{p} = q_0 + \frac{1}{q_1+}\,\frac{1}{q_2+}\,\cdots\,\frac{1}{q_2+}\,\frac{1}{q_1+}\,\frac{1}{2q_0+}\,\cdots,$$

период состоит из симметричных частей q_1, q_2, ..., q_2, q_1 за которыми следует $2q_0$. В такой форме это применимо не только к простому вида $4k + 1$, но и к любому числу, не являющемуся точным квадратом. Напомним (см. IV, 11), что если в симметричной части нет центрального члена, то разрешимо уравнение $x^2 - py^2 = -1$. Верно и обратное, хотя это и не было доказано в (IV, 11). Лежандр совершенно элементарным способом доказал, что если p — простое вида $4k + 1$, то уравнение $x^2 - py^2 = -1$ разрешимо. Следовательно, в силу только что сформулирован-

ной обратной теоремы центральный элемент в разложении $\sqrt{p}$ отсутствует и период имеет вид

$$q_1,\ q_2,\ \ldots,\ q_m,\ q_m,\ \ldots,\ q_2,\ q_1,\ 2q_0.$$

Пусть теперь α — полное частное, начинающееся в середине периода, именно:

$$\alpha = \alpha_m = q_m + \frac{1}{q_{m-1}+}\ \cdots\ \frac{1}{q_1+}\ \frac{1}{2q_0+}\ \frac{1}{q_1+}\ \cdots$$

Это чисто периодическая непрерывная дробь, период которой состоит из $q_m,\ q_{m-1},\ \ldots,\ q_1,\ 2q_0,\ q_1,\ \ldots,\ q_m$. Так как этот период симметричен, мы, как в (IV, 9), имеем $\alpha' = -\frac{1}{\alpha}$, где α' обозначает число, сопряженное с α. Представим теперь α в виде

$$\alpha = \frac{P + \sqrt{p}}{Q}$$

с целыми P и Q. Уравнение $\alpha\alpha' = -1$ дает

$$\frac{P + \sqrt{p}}{Q}\ \frac{P - \sqrt{p}}{Q} = -1,$$

или

$$p = P^2 + Q^2.$$

Это — конструкция Лежандра. В качестве примера рассмотрим случай $p = 29$. Разложение $\sqrt{29}$ в непрерывную дробь протекает так:

$$\begin{aligned}
\sqrt{29} &= 5 + \frac{1}{\alpha_1},\\
\alpha_1 &= \frac{1}{4}(5 + \sqrt{29}) = 2 + \frac{1}{\alpha_2},\\
\alpha_2 &= \frac{1}{5}(3 + \sqrt{29}) = 1 + \frac{1}{\alpha_3},\\
\alpha_3 &= \frac{1}{5}(2 + \sqrt{29}) = 1 + \frac{1}{\alpha_4},\\
\alpha_4 &= \frac{1}{4}(3 + \sqrt{29}) = 2 + \frac{1}{\alpha_5},\\
\alpha_5 &= 5 + \sqrt{29}.
\end{aligned}$$

Непрерывная дробь для $\sqrt{29}$ поэтому равна $5,\ \overline{2,\ 1,\ 1,\ 2,\ 10}$. Нужное нам полное частное — число $\alpha = \alpha_3$, отсюда $P = 2$ и $Q = 5$, и $29 = 2^2 + 5^2$.

Вторая конструкция была предложена Гауссом; по форме

эта конструкция наиболее проста (хотя обосновать ее и не так просто). Если $p = 4k + 1$, положим

$$x \equiv \frac{(2k)!}{2(k!)^2} \pmod{p}, \quad y \equiv (2k)!\, x \pmod{p},$$

где x и y выбраны между $-\frac{1}{2}p$ и $\frac{1}{2}p$. Тогда $p = x^2 + y^2$. Доказательство было дано Коши (Cauchy) и Якобшталем (Jacobsthal), оба доказательства не очень просты. Чтобы проиллюстрировать конструкцию, положим снова $p = 29$. Тогда

$$x \equiv \frac{14!}{2 \cdot 7!^2} = 1716 \equiv 5 \pmod{29},$$
$$y \equiv 14! \cdot x \equiv 14! \cdot 5 \equiv 2 \pmod{29}.$$

Эта конструкция, несмотря на свою элементарность, не очень удобна для вычислений.

Третья конструкция — конструкция Серре (Serret). Она, подобно построению Лежандра, использует непрерывные дроби, но здесь раскладывается в непрерывную дробь рациональное число. Разложим $\frac{p}{h}$ в непрерывную дробь (h удовлетворяет сравнению $h^2 + 1 \equiv 0 \pmod{p}$ и $0 < h < \frac{1}{2}p$). Можно доказать, что эта непрерывная дробь имеет вид

$$\frac{p}{h} = q_0 + \frac{1}{q_1+} \cdots \frac{1}{q_m+} \frac{1}{q_m+} \frac{1}{q_1+} \cdots \frac{1}{q_0}, \tag{8}$$

так что последовательность элементов непрерывной дроби симметрична и центральный член отсутствует. Используя обозначения главы IV, положим

$$x = [q_0, q_1, \ldots, q_m], \qquad y = [q_0, q_1, \ldots, q_{m-1}].$$

Тогда

$$p = x^2 + y^2.$$

Например, если $p = 29$, то $h = 12$, так как

$$12^2 + 1 = 145 = 5 \cdot 29.$$

Непрерывная дробь

$$\frac{29}{12} = 2 + \frac{1}{2+} \frac{1}{2+} \frac{1}{2}.$$

Отсюда

$$x = [2, 2] = 5, \quad y = [2] = 2.$$

То же построение в несколько иной форме было предложено в 1855 году Смитом (H. J. S. Smith). Он хотел дать простое и непосредственное доказательство того, что всякое простое число вида $4k+1$ представимо как сумма двух квадратов. Не пользуясь сравнениями, Смит доказал, что существует такое h, что $0 < h < \frac{1}{2}p$ и непрерывная дробь $\frac{p}{h}$ имеет вид (8). Определяя x и y, как и раньше, он доказал, следуя Серре, что $p = x^2 + y^2$.

Перейдем теперь к конструкции Якобшталя (Jacobsthal). Она основана на построении, аналогичном рассмотренному в (III, 6) в связи с распределением квадратичных вычетов. Рассмотрим следующую сумму символов Лежандра:

$$S(a) = \sum_n \left(\frac{n(n^2-a)}{p}\right),$$

где a — какое-нибудь число, не сравнимое с 0 по mod p, а суммирование распространяется на некоторую полную систему вычетов, например на числа $n = 0, 1, 2, \ldots, p-1$. Легко доказать, что $|S(a)|$ имеет лишь два возможных значения: одно, когда a — квадратичный вычет, другое, когда a — квадратичный невычет. Более того, оба этих значения четны, ибо слагаемое с $n = 0$ равно 0, а слагаемые, соответствующие n и $-n$, одинаковы, так как $(-1|p) = 1$. Положим

$$x = \tfrac{1}{2}|S(R)|, \qquad y = \tfrac{1}{2}|S(N)|,$$

где R — какой-нибудь квадратичный вычет, а N — какой-то квадратичный невычет. Тогда

$$p = x^2 + y^2.$$

Доказательство этого не очень трудно и основано главным образом на умелом применении равенства (18) главы III.

В качестве примера возьмем опять $p = 29$. Положим $R = 1$, а $N = 2$ (2 — квадратичный невычет по модулю 29). Значения $n(n^2-1)$ по mod 29 состоят из 0 и чисел

$$6,\ -5,\ 2,\ 4,\ 7,\ -12,\ 11,\ -5,\ 4,\ -14,\ 5,\ 9,\ 4,$$

каждое из которых встречается дважды. Сумма соответствующих им символов Лежандра равна 5, так что $x = 5$. Значения $n(n^2-2)$ по mod 29 состоят из 0 и чисел

$$-1,\ 4,\ -8,\ -2,\ -1,\ 1,\ 10,\ 3,\ -14,\ -6,\ 4,-7,\ -4,\ -10,$$

взятых по два раза каждое. Сумма значений их символов Лежандра равна 2, откуда $y = 2$.

4. Представление четырьмя квадратами. Жирар (Girard) и Ферма заметили, что *любое натуральное число представляется как сумма четырех квадратов целых чисел.* Учитывая, что в таком представлении некоторые слагаемые могут равняться нулю, эту теорему можно перефразировать так: каждое натуральное число представляется как сумма не более четырех квадратов натуральных чисел. Некоторые историки считают, что этот факт был известен уже Диофанту из Александрии: он не указал необходимых условий представимости числа суммой четырех квадратов, отметив, однако, что суммой двух или трех квадратов могут быть представлены лишь числа некоторого типа.

Эйлер многократно пытался доказать эту теорему, но безуспешно. Его постигла неудача, быть может, из-за того, что он пытался представить каждое число в виде суммы двух чисел, каждое из которых представляется суммой двух квадратов. Таким путем доказать эту теорему нелегко. Первое доказательство было дано в 1740 году Лагранжем. Лагранж отметил, что им были использованы работы Эйлера.

Доказательство Лагранжа аналогично доказательству теоремы о сумме двух квадратов, рассмотренному в пп. 1 и 2, и лишь немного сложнее его. В этом случае также имеется тождество, выражающее произведение двух сумм четырех квадратов в виде суммы четырех квадратов. Это тождество (принадлежащее Эйлеру) имеет вид

$$\begin{aligned}(a^2+b^2+c^2+d^2)(A^2+B^2+C^2+D^2) &= \\ &= (aA+bB+cC+dD)^2+(aB-bA-cD+dC)^2+ \\ &+(aC+bD-cA-dB)^2+(aD-bC+cB-dA)^2.\end{aligned} \tag{9}$$

Благодаря этому тождеству достаточно доказать, что каждое простое число представимо как сумма четырех квадратов; тогда представимость составных чисел будет следовать после повторного применения этого тождества. Так как мы знаем, что простое число 2 и все простые вида $4k+1$ представляются в виде суммы двух квадратов, то остается лишь доказать, что

любое простое вида $4k+3$ *представимо в виде суммы четырех квадратов.*

Как и в п. 2, доказательство распадается на два шага. Первый шаг состоит в доказательстве того, что некоторое кратное mp числа p, где $0 < m < p$, представимо в виде суммы четырех квадратов. На следующем шагу отсюда выводится, что само p представимо в таком виде. Первый шаг доказательства будет завершен, если мы установим разрешимость сравнения

$$x^2 + y^2 + 1 \equiv 0 \pmod{p}.$$

Действительно, можно выбрать решение с x и y, численно меньшими $\frac{1}{2}p$, и мы получим

$$mp = x^2 + y^2 + 1^2 + 0^2$$

с

$$m < \frac{\frac{1}{4}p^2 + \frac{1}{4}p^2 + 1}{p} < p.$$

Эйлер дал простое доказательство, устанавливающее разрешимость (10) без всяких вычислений. Перепишем сравнение (10) в виде

$$x^2 + 1 \equiv -y^2 \pmod{p}.$$

Любой квадратичный невычет сравним с числом вида $-y^2$, так как -1 — квадратичный невычет для любого простого вида $4k+3$ (III, 3). Таким образом, чтобы решить вышеупомянутое сравнение, достаточно найти квадратичный вычет R и квадратичный невычет N такие, что $R + 1 = N$. Если в качестве N взять первый квадратичный невычет в ряду 1, 2, 3, ..., то это условие, очевидно, будет выполняться, откуда и следует разрешимость сравнения.

Заметим, между прочим, что разрешимость сравнения (10) есть частный случай теоремы Шевалле (II, 8). Мы видели, что сравнение

$$x^2 + y^2 + z^2 \equiv 0 \pmod{p}$$

разрешимо с x, y, z, не сравнимыми с 0. Предполагая $z \neq 0$ и определяя X и Y так, чтобы $x \equiv Xz$, $y \equiv Yz$, получим $X^2 + Y^2 + 1 \equiv 0$.

Перейдем теперь ко второму шагу доказательства, начинающемуся с представления mp в виде

$$mp = a^2 + b^2 + c^2 + d^2, \tag{11}$$

где $0 < m < p$. Мы будем доказывать, так же как и в п. 2, что если $m > 1$, то найдется r, лежащее в промежутке $0 < r < m$ и обладающее тем же свойством, что и m. Отсюда, повторяя это рассуждение, можно получить, что 1 обладает этим свойством и, значит, p представимо в виде суммы четырех квадратов.

Начнем с приведения a, b, c, d по модулю m: именно, определим числа A, B, C, D так, чтобы они были сравнимы соответственно с a, b, c, d по модулю m и удовлетворяли условиям

$$-\tfrac{1}{2}m < A \leqslant \tfrac{1}{2}m, \quad -\tfrac{1}{2}m < B \leqslant \tfrac{1}{2}m,$$
$$-\tfrac{1}{2}m < C \leqslant \tfrac{1}{2}m, \quad -\tfrac{1}{2}m < D \leqslant \tfrac{1}{2}m.$$

Существует такое число r, что

$$mr = A^2 + B^2 + C^2 + D^2. \tag{12}$$

Число r не равно нулю, так как иначе все числа A, B, C, D равнялись бы нулю и все a, b, c, d были бы кратны m. Из (11) мы получили бы, что mp делится на m^2 или p делится на m, а это невозможно, так как p простое, а m больше 1, но меньше p.

Далее мы имеем

$$r = \frac{A^2 + B^2 + C^2 + D^2}{m} \leqslant \frac{\frac{1}{4}m^2 + \frac{1}{4}m^2 + \frac{1}{4}m^2 + \frac{1}{4}m^2}{m} = m.$$

Но этого недостаточно; нам нужно, чтобы r было строго меньше m. Возможность $r = m$ осуществится, только если все A, B, C, D равны $\frac{1}{2}m$. В таком случае m четно и все A, B, C, D сравнимы с $\frac{1}{2}m$ по модулю m. Но тогда $a^2 \equiv \frac{1}{4}m^2 \pmod{m^2}$; аналогичные сравнения имеют место и для b, c, d. Из (11) следует, что $mp \equiv 0 \pmod{m^2}$, а это, как мы уже видели, невозможно. Следовательно, число r в (12) удовлетворяет неравенству $0 < r < m$.

Продолжая доказательство, перемножим равенства (11) и (12) и применим тождество (9). Получим

$$m^2rp = x^2 + y^2 + z^2 + w^2, \tag{13}$$

где через x, y, z, w обозначены четыре выражения в правой части (9). Все эти выражения представляют собой числа, делящиеся на m. Действительно,

$$x = aA + bB + cC + dD \equiv a^2 + b^2 + c^2 + d^2 \equiv 0 \pmod{m},$$
$$y = aB - bA - cD + dC \equiv ab - ba - cd + dc \equiv 0 \pmod{m};$$

аналогично исследуются z и w. Мы можем сократить на m^2 обе части равенства (13); тогда получится представление rp в виде суммы четырех квадратов. Этим доказательство теоремы заканчивается.

Приведенное доказательство теоремы Лагранжа о сумме четырех квадратов немного проще, чем его первоначальное доказательство, и по существу совпадает с доказательством, данным позднее Эйлером. Хотя детали доказательства можно видоизменять, я не знаю никакого другого простого и элементарного доказательства, принципиально отличающегося от этого.

5. Представление тремя квадратами. Это гораздо более трудный вопрос. Одна из трудностей в том, что здесь тождества типа (1) или (9) не имеют места. Действительно, легко видеть, что произведение двух чисел, каждое из которых есть сумма трех квадратов, не обязано само быть суммой трех квадратов. Например, $3 = 1^2 + 1^2 + 1^2$ и $5 = 2^2 + 1^2 + 0^2$, в то же время 15 не представимо в виде суммы трех квадратов.

Как и в п. 1, мы можем установить, что некоторые числа не представимы в виде суммы трех квадратов. Любой квадрат сравним с 0, 1 или 4 по модулю 8. Следовательно, сумма трех квадратов не может быть сравнима с 7 по mod 8: 7 нельзя представить как сумму трех чисел, каждое из которых равно 0, 1 или 4. Значит, числа вида $8k + 7$ не представляются в виде суммы трех квадратов. Кроме того, число, делящееся на 4, скажем $4m$, представимо в виде суммы трех квадратов, только если так представимо само m. Действительно, любой квадрат сравним с 0 или 1 по mod 4, так что сумма трех квадратов делится на 4 только, если все квадраты четны. Таким образом, числа вида $4(8k + 7)$, $16(8k + 7)$ и так далее в виде суммы трех квадратов не представляются. Вообще ни одно число вида $4^l(8k + 7)$ не является суммой трех квадратов.

Оказывается, что каждое число иного вида представимо в виде суммы трех квадратов. Первое доказательство этого факта дал Лежандр, однако по ходу доказательства он предполагал, что любая арифметическая прогрессия a, $a+b$, $a+2b$, ..., где a и b взаимно просты, содержит бесконечно много простых чисел. Этот факт впервые в 1837 году (т. е. через сорок лет после работы Лежандра) доказал Дирихле. Гаусс в своих «*Арифметических исследованиях*» (Disquisitiones Arithmeticae) дал полное доказательство теоремы о сумме трех квадратов; его доказательство основано на довольно трудных результатах развитой им теории квадратичных форм. Были даны и другие доказательства, но ни одно из них не является одновременно простым и элементарным.

Замечания к главе V. п. 1. Читатель, знакомый с комплексными числами, узнает в тождестве (1) равенство $|\alpha\beta|^2 = |\alpha|^2|\beta|^2$, где $\alpha = a + ib$, $\beta = c + id$. Числа вида $a + ib$ с целыми a и b называются *целыми числами Гаусса*; представить число n в виде суммы двух квадратов — значит найти целое число Гаусса $a+ib$ с нормой a^2+b^2, равной n. На языке целых чисел Гаусса эта теория принимает более изящную форму.

п. 3. Для ссылок, см. ([7], vol. II, ch. 6 и vol. III, ch. 2). Указанные конструкции, вообще говоря, не дают для x и y положительных значений, хотя при $p = 29$ мы получили положительные x и y.

п. 4. Тождество (9) так же относится к кватернионам, как тождество (1) относится к комплексным числам (см. замечание к п. 1). Гурвиц рассмотрел вопрос о представлении четырьмя квадратами с помощью кватернионов; см. об этом ([9], ch. 20).

п. 5. Доказательство теоремы о трех квадратах, основанное на теореме Дирихле о простых в арифметических прогрессиях, дано в книге Ландау ([18], pp. 114—121).

Число представлений. Недостаток места не дает возможности рассказать об известных формулах для количества представлений числа n в виде суммы двух и четырех квадратов. В этих формулах подсчитывается число представлений целыми числами (положительными, отрицательными и нулем), причем два не тождественных представления считаются различными. Правило подсчета числа представлений (предложенное Лежандром) состоит в следующем. Подсчитаем число делителей n вида $4x+1$ и число делителей n вида

$4x + 3$. Если количества этих делителей обозначить соответственно через D_1 и D_3, то количество представлений числа n будет равно $4(D_1 - D_3)$. Для четырех квадратов правило подсчета числа представлений нашел Якоби; он вывел это правило из одного тождества, связывающего два бесконечных ряда. Если n нечетно, то число представлений n суммой четырех квадратов равно $8\sigma(n)$. Если n четно, положим $n = 2^r n'$, где n' нечетно; тогда число представлений n будет равно $24\sigma(n')$. Здесь, как и в (I, 5), $\sigma(n)$ обозначает сумму делителей n. Доказательство этих результатов можно найти, например, в книге ([9], chs. 16, 20). Количество представлений числа в виде суммы трех квадратов является гораздо более сложной функцией; его можно выразить через число классов квадратичных форм (VI, 9).

ГЛАВА VI

КВАДРАТИЧНЫЕ ФОРМЫ

1. Введение. В главе V мы нашли необходимое и достаточное условие представимости числа в виде суммы двух квадратов, условие, связанное с простыми множителями этого числа. Эйлеру и другим математикам восемнадцатого столетия удалось найти необходимые и достаточные условия представимости числа в виде $x^2 + 2y^2$ или $x^2 + 3y^2$; такие условия выражаются в терминах простых делителей числа. Эти математики, естественно, попытались, получить аналогичные результаты для общей квадратичной формы. Под квадратичной формой будем понимать выражение

$$ax^2 + bxy + cy^2,$$

являющееся однородным полиномом второй степени от своих переменных и имеющее целые коэффициенты a, b, c. Мы ограничимся рассмотрением форм от двух переменных или *бинарных* форм, хотя имеется также теория квадратичных форм от трех переменных (*тернарные* формы) или от любого числа переменных.

Теория квадратичных форм впервые была развита Лагранжем в 1773 году, ему принадлежат здесь многие основные идеи. Эта теория была упрощена и расширена Лежандром, а затем Гауссом, который ввел много новых понятий, использовав их для доказательства трудных и глубоких теорем, ускользавших от Лагранжа и Лежандра.

Классической задачей теории квадратичных форм является *проблема представления*: если дана квадратичная форма, то какие числа она представляет? Простой ответ можно дать лишь для некоторых форм частного вида, как $x^2 + y^2$ или $x^2 + 2y^2$, или $x^2 + 3y^2$; но в общем случае ответ на этот вопрос далеко не прост. Теорией дается простой ответ на несколько другой

вопрос — вопрос о представлении числа не данной конкретной формой, а хотя бы одной формой из некоторого класса форм.

Общие идеи теории, возникающие из понятия эквивалентности (п. 2), важны и в других более трудных и более развитых разделах теории чисел. Изучение квадратичных форм естественно вводит в круг этих идей и дает возможность познакомиться с ними в контексте, где они воспринимаются наиболее легко.

2. Эквивалентные формы. Фундаментальным понятием, связанным с квадратичными формами (а также с другими формами), является понятие эквивалентности. Сразу же видно, что две формы $2x^2 + 3y^2$ и $3x^2 + 2y^2$, по существу, одинаковы; одна получается из другой перестановкой переменных. Не столь ясно, что форма $2x^2 + 4xy + 5y^2$, по существу, совпадает с предыдущими формами. Эту форму можно представить в виде

$$2(x + y)^2 + 3y^2.$$

Когда переменные x и y принимают целые значения, $x + y$ и y принимают целые значения, и наоборот. Ясно, что всякое свойство достаточно общей природы, которым обладает форма $2x^2 + 3y^2$, является также свойством формы $2(x+y)^2+3y^2$, и обратно. В частности, это верно для свойств, связанных с представимостью чисел этими формами: если известны представления числа одной из форм, то сразу же можно найти представления этого числа и другой формой. Рассматриваемые формы связаны очень *простой подстановкой:* если положить $x = X + Y$ и $y = Y$, то

$$2x^2 + 3y^2 = 2X^2 + 4XY + 5Y^2.$$

Эта подстановка обладает следующим свойством: если x и y принимают целые значения, то X и Y также принимают целые значения, и обратно.

Рассмотрим теперь общий вопрос о том, какие подстановки вида

$$x = pX + qY, \quad y = rX + sY \tag{1}$$

обладают этим свойством, т. е. устанавливают одно однозначное соответствие между всеми парами целых x, y и всеми

парами целых X, Y? Мы *заранее* не накладываем никакого ограничения на коэффициенты p, q, r, s, хотя, конечно, ясно, что они должны быть целыми, так как значения $x = p$, $y = r$ соответствуют значениям $X = 1$, $Y = 0$, а значения $x = q$, $y = s$ соответствуют значениям $X = 0$, $Y = 1$. Если все четыре коэффициента целые, то целым значениям X и Y отвечают целые значения x и y.

Мы хотим, чтобы было верно и обратное. Для выяснения, когда это будет выполняться, нужно выразить X и Y через x и y. Если умножить первое равенство на s, а второе на q и вычесть одно из другого, то получится

$$sx - qy = (ps - qr)X,$$

аналогично

$$-rx + py = (ps - qr)Y.$$

Число $ps - qr$ не может равняться 0, так как иначе $sx - qy$ и $-rx + py$ также всегда равнялись бы 0 и переменные x и y не были бы независимы. Взяв $\Delta = ps - qr$ и разделив оба равенства на Δ, получим уравнения, выражающие X и Y через x и y:

$$X = \frac{s}{\Delta}x - \frac{q}{\Delta}y, \quad Y = -\frac{r}{\Delta}x + \frac{p}{\Delta}y. \tag{2}$$

Здесь все четыре коэффициента также должны быть целыми. Это будет, конечно, так, если $\Delta = \pm 1$. И обратно, все коэффициенты будут целыми только при $\Delta = \pm 1$; действительно, если все четыре коэффициента целые, то целым будет также и число

$$\frac{p}{\Delta}\frac{s}{\Delta} - \frac{q}{\Delta}\frac{r}{\Delta},$$

равное $\frac{1}{\Delta}$, а это возможно лишь при $\Delta = \pm 1$. Таким образом, все *коэффициенты p, q, r, s подстановки должны быть целыми и число $ps - qr$ должно равняться* ± 1; тогда и только тогда подстановка обладает требуемым свойством, т. е. сопоставляет целым парам x, y целые пары X, Y и наоборот.

Выражение $ps - qr$ называется *определителем* подстановки. Чтобы не осложнять дальнейшую теорию, будем использовать подстановки только с определителем 1 и не будем использовать

подстановок с определителем -1. Подстановка вида (1) с целыми коэффициентами и определителем, равным 1, называется *унимодулярной* подстановкой.

Две формы, связанные унимодулярной подстановкой, называются *эквивалентными*. Например, мы видели, что форму $2x^2 + 3y^2$ можно преобразовать в форму $2X^2 + 4XY + 5Y^2$ подстановкой

$$x = X + Y, \quad y = Y;$$

эта подстановка унимодулярна, значит, указанные формы эквивалентны. Не фиксируя буквы для обозначения переменных и изменяя их при каждой подстановке, удобно обозначать квадратичную форму $ax^2 + bxy + cy^2$ через (a, b, c) и символически выражать эквивалентность двух форм так:

$$(2, 0, 3) \sim (2, 4, 5).$$

Исходный пример $(2, 0, 3) \sim (3, 0, 2)$ требует пояснения. Подстановка, переставляющая переменные, т. е. подстановка $x = Y$, $y = X$, согласно предыдущему определению не является унимодулярной, ибо ее определитель равен -1. Вместо нее, однако, можно использовать подстановку $x = Y$, $y = -X$, являющуюся унимодулярной и преобразующую $(2, 0, 3)$ в $(3, 0, 2)$. Примененная к общей форме, эта подстановка дает

$$(a, b, c) \sim (c, -b, a). \tag{3}$$

Используя термин «эквивалентность», мы предполагаем, что отношение эквивалентности двух форм обладает некоторыми простыми свойствами; если это не так, то использование такого термина может ввести в заблуждение. Эти свойства таковы: (I) любая форма эквивалентна самой себе; (II) если одна форма эквивалентна другой, то и вторая форма эквивалентна первой; (III) две формы, эквивалентные третьей, эквивалентны между собой. Все эти факты сразу же следуют из определения эквивалентности форм. Во-первых, каждая форма эквивалентна самой себе: эту эквивалентность осуществляет *тождественная подстановка* $x = X$, $y = Y$. Во-вторых, если одна форма переводится в другую подстановкой (1), то вторая форма преобразуется в первую обратной подстановкой (2), где $\Delta = 1$. Наконец, третье утверждение следует из того факта, что две унимодулярные

подстановки, примененные одна за другой, можно заменить одной унимодулярной подстановкой.

Действительно, если сначала применяется подстановка

$$x = pX + qY, \quad y = rX + sY,$$

а затем подстановка

$$X = P\xi + Q\eta, \quad Y = R\xi + S\eta,$$

то эти подстановки можно заменить одной подстановкой

$$\begin{aligned} x &= p(P\xi + Q\eta) + q(R\xi + S\eta), \\ y &= r(P\xi + Q\eta) + s(R\xi + S\eta). \end{aligned}$$

Результирующая подстановка имеет целые коэффициенты, и ее определитель равен

$$\begin{aligned} (pP + qR)(rQ + sS) - (pQ + qS)(rP + sR) &= \\ = (ps - qr)(PS - QR) &= 1. \end{aligned}$$

Очевидно (как мы это уже отмечали в одном частном случае), что задача представления решается одинаково для эквивалентных форм. Аналогичное замечание можно сделать и в связи с задачей *собственного* представления. Говорят, что число n собственно представимо формой (a, b, c), если $n = ax^2 + bxy + cy^2$, где x и y — *взаимно простые* целые числа. Унимодулярная подстановка преобразует взаимно простые пары x, y во взаимно простые пары X, Y и обратно; действительно, если X и Y имеют общий множитель, то x и y имеют тот же общий множитель. Следовательно, если две формы эквивалентны, то существует унимодулярная подстановка, переводящая собственные представления числа первой формой в собственные представления этого числа второй формы.

3. Дискриминант. Дискриминантом квадратичной формы (a, b, c) называется число $b^2 - 4ac$. Так, дискриминант формы $(2, 0, 3)$ равен -24; дискриминант формы $(2, 4, 5)$ также равен $4^2 - 4 \cdot 2 \cdot 5 = -24$.

Заметим, что дискриминанты эквивалентных форм равны. Это проще всего доказать непосредственным вычислением. Действительно, применим подстановку (1) к форме $ax^2 + bxy + cy^2$;

получим форму $AX^2 + BXY + CY^2$, где

$$\begin{cases} A = ap^2 + bpr + cr^2; \\ B = 2apq + b(ps + qr) + 2crs; \\ C = aq^2 + bqs + cs^2. \end{cases} \tag{4}$$

Можно проверить, что

$$B^2 - AC = (b^2 - ac)(ps - qr)^2. \tag{5}$$

Так как $ps - qr = 1$, то формы (a, b, c) и (A, B, C) имеют один и тот же дискриминант. Тождество (5), конечно, не зависит от природы коэффициентов p, q, r, s в подстановке. Это чисто алгебраическое соотношение, так что мы встречаемся здесь с частным случаем весьма общей ситуации. Функция коэффициентов алгебраической формы, например $b^2 - 4ac$ в нашем случае, не меняющаяся при преобразовании формы подстановкой с единичным определителем, называется *алгебраическим инвариантом* формы. Дискриминант бинарной квадратичной формы — простейший пример такого инварианта.

Эквивалентные формы имеют одинаковый дискриминант; однако формы одного дискриминанта могут не быть эквивалентными. Например, формы $(1, 0, 6)$ и $(2, 0, 3)$ имеют одинаковый дискриминант -24, но не являются эквивалентными. Чтобы убедиться в этом, достаточно заметить, что форма $x^2 + 6y^2$ представляет число 1 при $x = 1$ и $y = 0$, в то время как форма $2x^2 + 3y^2$ ни при каких целых x и y не принимает значение 1.

Дискриминант d квадратичной формы — целое число, положительное, отрицательное или нуль. Не каждое целое число может служить дискриминантом формы: $b^2 - 4ac \equiv b^2 \pmod 4$, любой квадрат сравним с 0 или 1 по mod 4. Значит, d должно быть сравнимо с 0 или 1 по mod 4. Возможные значения дискриминанта таковы:

$$\ldots,\ -11,\ -8,\ -7,\ -4,\ -3,\ 0,\ 1,\ 4,\ 5,\ 8,\ 9,\ \ldots$$

Каждое из таких чисел служит дискриминантом по крайней мере одной формы. В самом деле, пусть d — любое число, сравнимое с 0 или 1 по mod 4, мы можем удовлетворить равенству $b^2 - 4ac = d$, полагая $a = 1$ и взяв b равным 0 или 1, в

зависимости от того, сравнимо d с 0 или с 1 по mod 4; тогда c будет равно соответственно $-\frac{1}{4}d$ или $-\frac{1}{4}(d-1)$. Таким образом, мы получаем для каждого d фиксированную форму дискриминанта d:

$$\left(1, 0, -\tfrac{1}{4}d\right) \quad \text{или} \quad \left(1, 0, -\tfrac{1}{4}(d-1)\right)$$

в зависимости от того, какое из сравнений $d \equiv 0 \pmod 4$ или $d \equiv 1 \pmod 4$ имеет место. Такая форма называется *главной формой* дискриминанта d. В частности, главной формой дискриминанта -4 является форма $(1, 0, 1)$ или $x^2 + y^2$, а главная форма дискриминанта 5 — это форма $(1, 1, -1)$, или $x^2 + xy - y^2$.

Имеется важное отличие между формами положительного и формами отрицательного дискриминантов. (Мы не будем рассматривать формы нулевого дискриминанта, так как такие формы являются просто квадратами линейных форм.) Рассмотрим сначала формы *отрицательного* дискриминанта. Умножим форму на $4a$ и «выделим полный квадрат»:

$$\begin{aligned} 4a(ax^2 + bxy + cy^2) &= 4a^2x^2 + 4abxy + 4acy^2 = \\ &= (2ax + by)^2 + (4ac - b^2)y^2. \end{aligned}$$

Число $4ac - b^2$ положительно. Поэтому если хоть одно из чисел x и y не равно 0, то полученное выражение положительно; если же и x и y равны 0, то это выражение также равно 0.

Таким образом, все числа, представляемые формой, имеют один и тот же знак: все они положительны, если a положительно, и все отрицательны, если a отрицательно. Такую форму называют *определенной*, точнее, *положительно определенной* или *отрицательно определенной* в зависимости от знака a. Отрицательно определенную форму можно всегда преобразовать в положительно определенную, изменив знаки всех коэффициентов, поэтому, исследуя определенные формы, достаточно рассматривать положительно определенные формы. Примерами положительно определенных форм могут служить формы $(1, 3, 7)$ дискриминанта -19 или $(5, -7, 5)$ дискриминанта -51.

Рассмотрим теперь формы *положительного* дискриминанта. Для этих форм выражение $4ac - b^2$ отрицательно; положим

$4ac-b^2=-d$, тогда $d>0$ и мы можем, выделив полный квадрат, разложить форму на множители. Получим

$$4a(ax^2+bxy+cy^2)=(2ax+by+\sqrt{d}y)(2ax+by-\sqrt{d}y)=\\ =4a^2(x-\theta y)(x-\varphi y),$$

где θ и φ имеют вид

$$\frac{-b\pm\sqrt{d}}{2a}.$$

Мы предполагаем здесь, что a не равно нулю. Числа θ и φ — действительные числа, которые, вообще говоря, не являются рациональными. Знак произведения $(x-\theta y)(x-\varphi y)$ зависит от того, будет ли дробь $\frac{x}{y}$ лежать между θ и φ или вне промежутка между ними. Так как имеются и те, и другие дроби, то форма может принимать как положительные, так и отрицательные значения. Такая форма называется *неопределенной*. Случай, когда a равно нулю, еще проще; в этом случае форма разлагается на множители $y(bx+cy)$ и, очевидно, принимает как положительные, так и отрицательные значения. Примеры неопределенных форм: $(3,1,-1)$ дискриминанта 13 или $(1,4,1)$ дискриминанта 12. Отметим, что, как это было в последнем примере, форма с положительными коэффициентами может быть неопределенной.

Итак, формы отрицательного дискриминанта являются определенными, а формы положительного дискриминанта — неопределенными. Первый шаг теории, на котором проблема представления сводится к проблеме эквивалентности, проводится одинаково для определенных и для неопределенных форм.

В дальнейшем теория неопределенных форм существенно отличается от теории определенных форм; из-за недостатка места мы будем рассматривать только определенные формы.

4. Представление числа формой. Выясняя, какие числа представляются формой (a,b,c), достаточно рассматривать собственные представления. Зная, какие числа собственно представимы, можно получить несобственно представимые числа умножением на квадраты целых чисел.

Предположим, что число n собственно представляется формой (a, b, c). Обозначим через p и r числа, осуществляющие представление

$$n = ap^2 + bpr + cr^2 \tag{6}$$

(считаем, что p и r взаимно просты). Если форма определенная, скажем положительно определенная, то мы будем предполагать n положительным; если же форма неопределенная, то n может быть как положительным, так и отрицательным. Будем предполагать, что n не равно нулю; возможность $n = 0$ лучше рассмотреть отдельно (и она мало интересна).

Так как p и r взаимно просты, то можно найти такие целые q и s, что $ps - qr = 1$. Применим теперь унимодулярную подстановку (1) с найденными здесь p, q, r, s форме (a, b, c). Из сопоставления равенства (6) с первым из равенств (4) следует, что первый коэффициент получающейся формы равен n. Таким образом, мы нашли форму (n, h, l), эквивалентную форме (a, b, c) и имеющую своим первым коэффициентом число n. Обратно, всякая форма, первый коэффициент которой равен n, собственно представляет число n (при $x = 1$, $y = 0$), значит, n собственно представляется всякой формой (a, b, c), эквивалентной форме с первым коэффициентом n. Итак, (a, b, c) *собственно представляет те и только те числа, которые встречаются среди первых коэффициентов эквивалентных ей форм.*

С первого взгляда может показаться, что этот подход к задаче вряд ли что-нибудь даст; тем не менее на нем основана вся последующая теория. Проблема представления сводится теперь к проблеме эквивалентности в том смысле, что теперь достаточно выяснить, эквивалентна ли какая-нибудь форма с первым коэффициентом n данной форме (a, b, c).

Из сформулированного выше общего принципа можно сделать один простой, но важный вывод. Форма (n, h, l) не может быть эквивалентна форме (a, b, c), если ее дискриминант не равен дискриминанту формы (a, b, c), т. е. если не выполняется равенство

$$h^2 - 4nl = d, \tag{7}$$

где $d = b^2 - 4ac$ — дискриминант данной формы. Другими словами, должно существовать h, для которого $h^2 - d$ кратно $4n$,

т. е. должно быть разрешимо сравнение

$$h^2 \equiv d \pmod{4n'}, \tag{8}$$

где $n' = |n|$. (Мы должны взять в качестве модуля сравнения $4n'$, а не $4n$, так как n может быть отрицательно.) При некоторых ограничениях верно и обратное. Именно: если сравнение (8) разрешимо, то существует форма вида (n, h, l) дискриминанта d; правда, эта форма не обязана быть эквивалентной заранее выбранной форме (a, b, c). Отсюда можно сделать следующий вывод: *если n собственно представляется какой-нибудь формой дискриминанта d, то разрешимо сравнение* (8). *Обратно, если это сравнение разрешимо, то n собственно представляется некоторой формой дискриминанта d.*

В некоторых простых случаях бывает, что все формы дискриминанта d эквивалентны между собой. Тогда разрешимость сравнения (8) есть необходимое и достаточное условие для того, чтобы n было собственно представимо данной формой (a, b, c) дискриминанта d. В следующем пункте мы применим этот принцип в трех таких случаях.

Однако прежде необходимо сделать одно дополнительное замечание. Сформулированный выше общий принцип предписывает решить сравнение (8), а затем выяснить, эквивалентна ли данной форме (a, b, c) форма (n, h, l), где l находится из равенства $h^2 - 4nl = d$. Это приведет к рассмотрению бесконечного числа случаев, если выбирать в качестве h всевозможные решения сравнения (8). На самом деле, однако, достаточно рассмотреть значения h, удовлетворяющие условию

$$0 \leqslant h < 2n'. \tag{9}$$

Действительно, если h — какое-нибудь решение сравнения (8) и (n, h, l) — соответствующая ему квадратичная форма, то мы можем применить к этой форме подстановку

$$x = X + uY, \quad y = Y,$$

где u — произвольное целое число. Эта подстановка преобразует форму (n, h, l) в форму

$$n(X + uY)^2 + h(X + uY)Y + lY^2.$$

Первый коэффициент получившейся формы по-прежнему равен n, а средний коэффициент равен уже $h + 2un$. Следовательно, две формы с первым коэффициентом n и средними коэффициентами, отличающимися на кратное $2n$, эквивалентны. Поэтому достаточно рассматривать формы, для которых h удовлетворяет и сравнению (8), и неравенству (9).

5. Три примера. Рассмотрим сначала форму $x^2 + y^2$ дискриминанта -4. В п. 7 будет доказано, что все формы дискриминанта -4 эквивалентны между собой. В силу общего принципа отсюда следует, что целое число n собственно представляется формой $x^2 + y^2$ тогда и только тогда, когда разрешимо сравнение

$$h^2 \equiv -4 \pmod{4n}.$$

Так как удовлетворяющее этому сравнению h должно быть четным, то мы можем разделить обе части этого сравнения на 4 и вместо него рассматривать сравнение

$$h^2 \equiv -1 \pmod{n}. \tag{10}$$

Вопрос о разрешимости этого сравнения, очевидно, связан с теорией квадратичных вычетов. Во-первых, согласно общему принципу, управляющему разрешимостью сравнений по составному модулю (II, 6), достаточно выяснить, разрешимо ли сравнение

$$h^2 \equiv -1 \pmod{p^r} \tag{11}$$

для каждой степени простого числа, входящей в разложение n.

Сравнение (11) не может быть разрешимо, если p имеет вид $4k + 3$, так как -1 является квадратичным невычетом по такому модулю (III, 3). Если p есть простое вида $4k + 1$, то, как известно, при $r = 1$ это сравнение разрешимо, ибо -1 является квадратичным вычетом по такому модулю. Легко доказать по индукции, что рассматриваемое сравнение разрешимо тогда и для любого показателя r. Например, если r равно 2, то возьмем h_1 для которого $h_1^2 \equiv -1 \pmod{p}$, и попытаемся удовлетворить сравнению $h^2 \equiv -1 \pmod{p^2}$, полагая $h = h_1 + tp$, где t — новое неизвестное. Имеем

$$h^2 + 1 = (h_1 + tp)^2 + 1 = h_1^2 + 1 + 2th_1p + t^2p^2.$$

Это выражение разделится на p^2, если

$$\frac{1}{p}(h_1^2+1)+2th_1 \equiv 0 \pmod{p};$$

здесь первое слагаемое является по предположению целым числом. Это — линейное сравнение относительно t; оно разрешимо, так как $2h_1$ не сравнимо с 0 по mod p. То же рассуждение применимо и для более высоких показателей; чтобы решить сравнение при $r=3$, возьмем число h_2 такое, что $h_2^2 \equiv -1 \pmod{p^2}$, и положим $h=h_2+tp^2$, в результате мы опять получим для t линейное сравнение по модулю p.

Этим решается вопрос о разрешимости сравнения (11) для простых вида $4k+1$ и $4k+3$. Остается простое число 2. Для этого числа при $r=1$ сравнение очевидным образом разрешимо (решением служит $h=1$). Но оно не разрешимо, если $r \geqslant 2$, так как любой квадрат сравним с 0 или с 1 по mod 4 и, следовательно, не может быть сравним с -1 по mod 2^r при $r \geqslant 2$.

Поэтому сравнение (10) разрешимо в том и только том случае, если n *не имеет простых множителей вида* $4k+3$ *и не делится на* 4. Это — необходимое и достаточное условие собственной представимости n в виде x^2+y^2. Если допустить умножение на любой квадрат, то получится условие представимости числа суммой двух квадратов (собственно или несобственно), уже установленное в главе V.

В качестве второго примера рассмотрим положительно определенную форму x^2+xy+y^2 дискриминанта -7. В п. 7 будет доказано, что все формы дискриминанта -7 эквивалентны между собой. Предположив это, мы должны решить, для каких чисел n разрешимо сравнение

$$h^2 \equiv -7 \pmod{4n}. \tag{12}$$

Предположим для простоты, что n нечетно, так что 4 и n взаимно просты. Сравнение $h^2 \equiv -7 \pmod 4$, конечно, разрешимо, например, при $h=1$. Сравнение $h^2 \equiv -7 \pmod{p}$ разрешимо для тех простых p, для которых -7 является квадратичным вычетом по mod p. Квадратичный закон взаимности указывает нам, какие p обладают этим свойством. Если p не равно 7, то

$$\left(\frac{-7}{p}\right)=\left(\frac{-1}{p}\right)\left(\frac{7}{p}\right)=\left(\frac{-1}{p}\right)(-1)^{\frac{p-1}{2}}\left(\frac{p}{7}\right)=\left(\frac{p}{7}\right),$$

что равно $+1$ для простых p вида $7k+1$, $7k+2$ или $7k+4$ и равно -1 для простых p вида $7k+3$, $7k+5$ или $7k+6$. Так же как и раньше, можно доказать, что если рассматриваемое сравнение разрешимо для некоторого простого модуля, то оно разрешимо и для любой степени этого простого. Остается рассмотреть случай $p=7$. В этом случае сравнение $h^2 \equiv -7 \pmod 7$, очевидно, разрешимо ($h=0$), но сравнение $h^2 \equiv -7 \pmod{7^2}$ уже не разрешимо. Таким образом, сравнение (12) разрешимо в том и только том случае, когда n *не делится на простые вида* $7k+3$, $7k+5$, $7k+6$ *и не делится на* 49. Это и есть необходимое и достаточное условие собственной представимости нечетного числа n в виде x^2+xy+y^2.

В качестве последнего примера рассмотрим неопределенную форму x^2-2y^2 дискриминанта 8. Все формы дискриминанта 8 эквивалентны между собой, хотя мы этого доказывать не будем. В этом случае следует рассматривать сравнение

$$h^2 \equiv 8 \pmod{4n'}, \quad \text{где } n' = |n|,$$

которое можно заменить сравнением $h^2 \equiv 2 \pmod{n'}$. Сравнение $h^2 \equiv 2 \pmod{p^2}$ разрешимо для простых p вида $8k+1$ или $8k-1$, но не разрешимо, если p — простое вида $8k+3$ или $8k-3$. Если $p=2$, то сравнение разрешимо при $r=1$ и не разрешимо при $r \geqslant 2$. Таким образом, *число* n *(положительное или отрицательное) собственно представляется формой* x^2-2y^2 *тогда и только тогда, когда* $|n|$ *не имеет простых делителей вида* $8k+3$ *и* $8k-3$ *и не делится на* 4.

Конечно, не всегда бывает так, что условие представимости неопределенной формой зависит только от $|n|$ и одинаково для n и $-n$. Причина, по которой это обстоятельство имеет место здесь, в том, что формы x^2-2y^2 и $-x^2+2y^2$ эквивалентны (так как все формы дискриминанта 8 эквивалентны между собой).

6. Редукция положительно определенных форм. Существует бесконечно много форм данного дискриминанта d; их можно разбить на классы, относя две формы в один класс тогда и только тогда, когда они эквивалентны. Как мы увидим позднее, число классов эквивалентных форм данного дискриминанта d конечно.

Пусть дана какая-нибудь форма, среди эквивалентных ей форм желательно найти форму простейшего вида. Это достигается с помощью теории приведения. Теория приведения строится по-разному для определенных и неопределенных форм; мы ограничимся здесь рассмотрением определенных форм. Теория приведения неопределенных форм труднее, и недостаток места не позволяет нам изложить основы этой теории.

Теория приведения положительно определенных форм была построена Лагранжем. Заметим прежде всего, что у положительно определенной формы коэффициенты a и c положительны; b может быть и положительным, и отрицательным. Мы сосредоточим внимание на a и $|b|$ и рассмотрим две операции эквивалентности, с помощью которых одно из этих чисел можно уменьшить, не изменяя другого. Эти операции таковы:

(I) Если $c < a$, форма (a, b, c) заменяется на эквивалентную ей форму $(c, -b, a)$.

(II) Если $|b| > a$, форма (a, b, c) заменяется эквивалентной ей формой (a, b_1, c_1), где $b_1 = b + 2ua$, и целое число u выбрано так, что $|b_1| \leqslant a$; при этом c_1 можно найти из равенства

$$b_1^2 - 4ac_1 = d.$$

Эквивалентность в (I) осуществляется подстановкой $x = Y$, $y = -X$, а эквивалентность в (II) — подстановкой $x = X + uY$, $y = Y$, использованной в конце п. 4.

С помощью операции (I) мы уменьшаем a, не изменяя значения $|b|$, а с помощью операции (II) уменьшаем b, не изменяя значения a. Если дана какая-нибудь форма, мы можем применять эти операции до тех пор, пока не найдем форму, которая не удовлетворяет ни одному из предположений, необходимых для осуществления этих операций; очевидно, что такая форма может быть получена в конечное число шагов. Коэффициенты этой формы удовлетворяют неравенствам

$$c \geqslant a \quad \text{и} \quad |b| \leqslant a. \tag{13}$$

Таким образом, мы доказали, что любая положительно определенная форма эквивалентна форме, коэффициенты которой удовлетворяют условиям (13).

В качестве иллюстрации применим описанный процесс приведения к форме $(10, 34, 29)$ дискриминанта -4. Здесь $b > a$, применяя (II), добьемся, чтобы b принадлежало интервалу от -10 до 10; для этого вычтем из b подходящее кратное 20, в данном случае 40. После вычитания получим форму $(10, -6, ?)$, в которой неизвестный третий коэффициент находится по дискриминанту. Обозначая этот коэффициент через c_1, находим $(-6)^2 - 40c_1 = -4$, откуда $c_1 = 1$. Новая форма имеет вид $(10, -6, 1)$; применяя теперь операцию (I), получаем форму $(1, 6, 10)$. Далее применим (II), что приведет нас в этом случае к форме с нулевым средним коэффициентом. Мы получим форму $(1, 0, ?)$, в которой неизвестный третий коэффициент равен 1 (он находится по значению дискриминанта). Таким образом, мы доказали, что данная форма эквивалентна форме $(1, 0, 1)$.

Бывает, что исходная форма удовлетворяет условиям применения обеих операций (I) и (II). Например, если дана форма $(15, 17, 10)$, то мы можем начать либо применением (I), тогда получится форма $(10, -17, 15)$; либо применением (II), тогда получится форма $(15, -13, 8)$.

Возвращаясь к неравенствам (13), отметим, что в двух случаях мы можем с успехом применить одну из этих операций, даже если условия (13) выполняются. Во-первых, если $b = -a$, то, применяя (II), можно заменить b на a. Во-вторых, если $c - a$, то, применяя операцию (I), можно изменить знак b, добившись таким образом, чтобы b было положительным или равнялось нулю. Принимая во внимание эти две возможности, находим, что *любая положительно определенная форма эквивалентна форме, коэффициенты которой удовлетворяют условию*:

$$\begin{array}{lll} \textit{либо} & c > a & \textit{и} \quad -a < b \leqslant a, \\ \textit{либо} & c = a & \textit{и} \quad 0 \leqslant b \leqslant a. \end{array} \tag{14}$$

Если коэффициенты формы удовлетворяют условиям (14), то она называется *приведенной*.

Имеет место замечательная и важная теорема о том, что существует одна и *только одна* приведенная форма, эквивалентная данной форме. Соответствующее доказательство не очень трудно, но требует все же рассуждений более искусных, чем использованные выше. Основная идея доказательства состоит в

нахождении инвариантной интерпретации коэффициентов приведенной формы, показывающей, что приведенная форма, эквивалентная данной форме, единственна. Например, можно доказать, что первый коэффициент a приведенной формы есть наименьшее число, собственно представляемое этой формой. Из-за недостатка места мы опускаем доказательство упомянутого факта.

Вопрос об эквивалентности двух данных форм может быть (благодаря указанной теореме) решен посредством приведения этих форм. Если обе приведенные формы окажутся одинаковыми, то исходные формы эквивалентны, в противном случае эти формы не эквивалентны.

7. Приведенные формы. Из неравенств (14) легко следует, что существует лишь конечное число приведенных форм данного отрицательного дискриминанта d. Действительно, положим $d = -D$, тогда D положительно и

$$4ac - b^2 = D. \tag{15}$$

Так как в силу (14) $b^2 \leqslant a^2 \leqslant ac$, то $3ac \leqslant D$. Существует только конечное число положительных целых чисел a и c, удовлетворяющих этому условию; при каждом выборе a и c (в силу (15)) для b имеется не более двух возможностей, откуда и следует требуемый результат. Число приведенных форм совпадает, конечно, с числом классов эквивалентных форм, ибо в каждом классе имеется ровно одна приведенная форма. Это число называется *числом классов* дискриминанта d.

Вероятно, быстрейший способ пересчета приведенных форм данного дискриминанта состоит в следующем. Заметим прежде всего, что $b^2 \leqslant ac \leqslant \frac{1}{3}D$ и $4ac = D + b^2$. Кроме того, b четно, если $D \equiv 0 \pmod 4$, и нечетно, если $D \equiv 3 \pmod 4$ (это соответствует случаю $d \equiv 1 \pmod 4$). Заметив это, следует перебрать все значения b подходящей четности (положительные и отрицательные) вплоть до $\sqrt{\frac{1}{3}D}$ и представить $\frac{1}{4}(D + b^2)$ в виде ac всеми возможными способами, а затем выбросить тройки a, b, c, не удовлетворяющие условию (14).

Например, если $d = -4$, так что $D = 4$, то $|b| \leqslant \sqrt{\frac{4}{3}}$ и b

четно, откуда $b = 0$. Далее, $4ac = 4$, поэтому $a = c = 1$. Здесь имеется только одна приведенная форма $(1, 0, 1)$. Это первый пример п. 5.

Рассмотрим теперь второй пример п. 5; предположим, что $d = -7$ и, значит, $D = 7$. Тогда $|b| \leqslant \sqrt{\frac{7}{3}}$, кроме того, b нечетно, значит, $b = 1$ или -1. Далее, $4ac = 1 + 7 = 8$, откуда $a = 1$, $c = 2$. Возможность $b = -1$ должна быть отброшена, так как она не согласуется с (14). Таким образом, в этом случае мы имеем единственную форму $(1, 1, 2)$.

Действуя подобным образом, легко составить таблицу приведенных форм. В таблицу II входят формы с дискриминантами от -3 до -83. Знаком * отмечены так называемые imprимитивные формы — это формы, у которых коэффициенты a, b, c имеют отличный от 1 общий делитель. Imprимитивная форма равна точному кратному некоторой примитивной формы меньшего дискриминанта.

Приведенные формы данного дискриминанта образуют *систему представителей* форм этого дискриминанта; эта система представителей содержит по одной форме из каждого класса эквивалентных форм. Теория п. 4 указывает необходимое и достаточное условие собственной представимости числа хотя бы одной из этих приведенных форм; это и есть тот результат, о котором говорилось в п. 1. В тех случаях, когда существует только одна приведенная форма, проблема представления решается полностью. Единственной приведенной формой в этом случае является главная форма, так как она удовлетворяет условиям (14).

Иногда удается решить проблему представления и в тех случаях, когда имеется больше одной приведенной формы. Рассмотрим первый (исключая imprимитивные формы) из таких случаев, именно случай $d = -15$. Здесь имеются две приведенные формы: $(1, 1, 4)$ и $(2, 1, 2)$. Пусть n представляется первой из них; тогда

$$4n = (2x + y)^2 + 15y^2 \equiv (2x + y)^2 \pmod{15}.$$

Если n не делится на 15, то легко установить, что n сравнимо с одним из чисел 1, 4, 6, 9, 10 по mod 15. Аналогично, если n представляется второй формой, находим, что n сравнимо с одним из

Таблица II

Приведенные положительно определенные формы дискриминанта $-D$

D	a, b, c	D	a, b, c	D	a, b, c
3	1, 1, 1	43	1, 1, 11	64	1, 0, 16
4	1, 0, 1	44	1, 0, 11		2, 0, 8*
7	1, 1, 2		2, 2, 6*		4, 0, 4*
8	1, 0, 2		3, 2, 4		4, 4, 5
11	1, 1, 3		3, −2, 4	67	1, 1, 17
12	1, 0, 3	47	1, 1, 12	68	1, 0, 17
	2, 2, 2*		2, 1, 6		2, 2, 9
15	1, 1, 4		2, −1, 6		3, 2, 6
	2, 1, 2		3, 1, 4		3, −2, 6
16	1, 0, 4		3, −1, 4	71	1, 1, 18
	2, 0, 2*	48	1, 0, 12		2, 1, 9
19	1, 1, 5		2, 0, 6*		2, −1, 9
20	1, 0, 5		3, 0, 4		3, 1, 6
	2, 2, 3		4, 4, 4*		3, −1, 6
23	1, 1, 6	51	1, 0, 13		4, 3, 5
	2, 1, 3		3, 3, 5		4, −3, 5
	2, −1, 3	52	1, 0, 13	72	1, 0, 18
24	1, 0, 6		2, 2, 7		2, 0, 9
	2, 0, 3	55	1, 1, 14		3, 0, 6*
27	1, 1, 7		2, 1, 7	75	1, 1, 19
	3, 3, 3*		2, −1, 7		3, 3, 7
28	1, 0, 7		4, 3, 4		5, 5, 5*
	2, 2, 4*	56	1, 0, 14	76	1, 0, 19
31	1, 1, 8		2, 0, 7		2, 2, 10*
	2, 1, 4		2, 2, 5		4, 2, 5
	2, −1, 4		3, −2, 5		4, −2, 5
32	1, 0, 8	59	1, 1, 15	79	1, 1, 20
	2, 0, 4*		3, 1, 5		2, 1, 10
	3, 2, 3		3, −1, 5		2, −1, 10
35	1, 1, 9	60	1, 0, 15		4, 1, 5
	3, 1, 3		3, 0, 5		4, −1, 5
36	1, 0, 9		2, 2, 8*	80	1, 0, 20
	2, 2, 5		4, 2, 4*		2, 0, 10*
	3, 0, 3*	63	1, 1, 16		3, 2, 7
39	1, 1, 10		2, 1, 8		3, −2, 7
	2, 1, 5		2, −1, 8		4, 0, 5
	2, −1, 5		4, 1, 4		4, 4, 6*
	3, 3, 4		3, 3, 6*	83	1, 1, 21
40	1, 0, 10				3, 1, 7
	2, 0, 5				3, −1, 7

чисел 2, 3, 5, 8, 12 по mod 15. Поэтому можно легко отличить числа, представимые первой формой, от чисел, представимых второй формой, исключая, возможно, числа, кратные 15. Чтобы выразить это различие, Гаусс ввел понятие рода; о двух только что рассмотренных формах говорят, что они принадлежат разным родам. Теория родов, однако, слишком сложна и обширна, чтобы излагать ее здесь.

Указанная возможность различать числа, представляемые двумя различными приведенными формами, обеспечивалась существованием такого модуля (в рассмотренном случае 15), для которого числа, представимые двумя разными формами, удовлетворяют разным сравнениям по этому модулю. Когда такого модуля нет (что иногда бывает), задача представления конкретной формой, по существу, еще не решена. Мы можем, например, найти условие представимости числа одной из форм x^2+55y^2 и $5x^2+11y^2$, но нам не известно никакого простого общего правила, решающего вопрос, для какой из этих форм осуществляется представление.

8. Число представлений. Теория п. 4 дает необходимое и достаточное условие собственной представимости числа одной из форм дискриминанта d; это условие состоит в разрешимости сравнения (8). Можно сделать следующий шаг и перейти к определению полного числа собственных представлений n всеми приведенными формами дискриминанта d. Обозначим это число через $R(n)$. Если существует лишь одна приведенная форма дискриминанта d (например, x^2+y^2 при $d=-4$), то мы получим в результате число представлений этой формой.

Мы наметим здесь теорию, с помощью которой определяется $R(n)$; детали доказательства нам придется опустить. Предположим для простоты, что n взаимно просто с d. Из этого, в частности, следует, что любая форма дискриминанта d, представляющая n, примитивна, так как общий множитель a, b, c делит и n, и d.

Начнем с тех же рассмотрений, что и в п. 4. Мы видели, что каждому собственному представлению n формой (a,b,c), скажем

$$n=ap^2+bpr+cr^2, \tag{16}$$

отвечает подстановка, преобразующая форму (a, b, c) в эквивалентную ей форму (n, h, l) с первым коэффициентом n, второй коэффициент формы (n, h, l) удовлетворяет сравнению

$$h^2 \equiv d \pmod{4n} \tag{17}$$

и неравенству

$$0 \leqslant h < 2n. \tag{18}$$

Чтобы подсчитать полное число представлений $R(n)$, мы должны подсчитать, сколько чисел h удовлетворяет (17) и (18) и сколько представлений (16) отвечает одному и тому же h.

Начнем со второго подсчета. Одно и то же h не может происходить от двух различных приведенных форм: эти формы были бы эквивалентны одной и той же форме (n, h, l), что невозможно. Если два представления n формой (a, b, c) приведут к одному и тому же числу h, то соответствующие им подстановки можно скомбинировать (применяя сначала первую подстановку, а затем подстановку, обратную ко второй) так, что получится подстановка, переводящая (a, b, c) в себя. Легко видеть, что число представлений n, приводящих к одному и тому же значению h, равно числу унимодулярных подстановок, преобразующих (a, b, c) в себя.

Здесь возникает вопрос, который мы еще не рассматривали. Унимодулярная подстановка, переводящая форму в себя, называется автоморфной подстановкой, или *автоморфизмом* формы. У всякой формы имеются два очевидных автоморфизма: тождественная подстановка $x = X$, $y = Y$ и отрицательная тождественная подстановка $x = -X$, $y = -Y$. В общем случае других автоморфизмов нет; но есть два исключения. Форма $x^2 + y^2$ имеет еще два автоморфизма: $x = Y$, $y = -X$ и $x = -Y$, $y = X$, т. е. всего четыре автоморфизма. Форма $x^2 + xy + y^2$ имеет четыре дополнительных автоморфизма:

$$\begin{array}{lll} \text{(I)} & x = X + Y, & y = -X, \\ \text{(II)} & x = -X - Y, & y = X, \\ \text{(III)} & x = Y, & y = -X - Y, \\ \text{(IV)} & x = -Y, & y = X + Y, \end{array}$$

т. е. всего шесть автоморфизмов. Можно доказать, что этот список автоморфизмов полный; число автоморфизмов, скажем w, равно 6, если $d = -3$; w равно 4, если $d = -4$; и w равно 2 в

остальных случаях. Это относится только к примитивным формам; импримитивная форма $2x^2 + 2y^2$ имеет, конечно, те же автоморфизмы, что и форма $x^2 + y^2$.

Таким образом, *полное число $R(n)$ собственных представлений n всеми приведенными формами дискриминанта d есть взятое w раз число значений h, удовлетворяющих сравнению* (17) *и неравенству* (18).

Остается теперь найти число решений сравнения (17); мы ограничимся рассмотрением частного случая $d = -4$. Наше предыдущее предположение о том, что n взаимно просто с d, в этом случае означает просто нечетность n. Сократив сравнение (17) на 4 и неравенство (18) на 2, найдем число решений

$$h^2 \equiv -1 \pmod{n} \tag{19}$$

при условии

$$0 \leqslant h < n. \tag{20}$$

Согласно общему принципу (II, 6), это число равно произведению количеств решений сравнений

$$h^2 \equiv -1 \pmod{p^r} \tag{21}$$

по всем простым степеням, составляющим n.

Сравнение (21) не разрешимо для всех p вида $4k+3$ и имеет два решения, если $p = 4k + 1$ и $r = 1$. С помощью метода, использованного в п. 5, можно легко доказать, что если $p = 4k+$ $+ 1$, то и при $r > 1$ будет ровно два решения. Таким образом, число решений (19) равно 0, если n имеет хоть один множитель вида $4k + 3$, и равно 2^s, если n имеет s различных простых множителей вида $4k + 1$ и ни одного множителя вида $4k + 3$. Так как для формы $x^2 + y^2$ будет $w = 4$, то *число собственных представлений нечетного числа n формой $x^2 + y^2$ равно $4 \cdot 2^s$, если n имеет s различных простых множителей вида $4k+1$ и не имеет ни одного множителя вида $4k + 3$*. Если же n имеет хоть один простой множитель вида $4k + 3$, то форма $x^2 + y^2$ собственно не представляет числа n.

Представления можно разбить на группы по 8 представлений; в каждой группе представления получаются из одного изменением знаков x и y и заменой y на x. Поэтому число существенно различных представлений равно не $4 \cdot 2^s$, а 2^{s-1}.

9. Число классов. Обозначим через $C(d)$ число классов форм дискриминанта d, равное числу приведенных форм дискриминанта d. Ограничимся для простоты рассмотрением дискриминантов, для которых каждая форма примитивна; такие дискриминанты называются *фундаментальными*. Вот несколько примеров из таблицы II:

$$C(-3)=1, \quad C(-4)=1, \quad C(-51)=2, \quad C(-71)=7.$$

Мы можем, конечно, интерпретировать $C(d)$ как число троек (a,b,c), удовлетворяющих равенству $b^2-4ac=1$ и неравенствам (14) из п. 6.

Имеется замечательная формула для $C(d)$, дающая возможность определить это число, не обращаясь к квадратичным формам. Формула принимает простейший вид, если $d=-p$, где p простое (p должно иметь вид $4k+3$, так как $d\equiv 0 \pmod 4$ или $d\equiv 1 \pmod 4$). Образуем сумму, скажем, A всех квадратичных вычетов по mod p и сумму B всех квадратичных невычетов. Тогда

$$C(-p)=\frac{B-A}{p}. \tag{22}$$

Например, если $p=23$, квадратичными вычетами будут числа 1, 2, 3, 4, 6, 8, 9, 12, 13, 16, 18 с суммой 92, а квадратичными невычетами — числа 5, 7, 10, 11, 14, 15, 17, 19, 20, 21, 22 с суммой 161.

Из формулы (22) следует, что

$$C(-23)=\frac{161-92}{23}=3.$$

Это совпадает с табличным значением $C(-23)$. Открытие этой замечательной формулы, вероятно, принадлежит Якоби, хотя, возможно, она была независимо переоткрыта Гауссом. Якоби доказал, что число $\frac{B-A}{p}$ обладает некоторым общим с числом классов $C(-p)$ свойством, а затем, исследовав много численных примеров, пришел к выводу, что эти числа тождественны. Он сообщил об этом в 1832 году, отметив, что доказательство этой формулы ему найти не удалось. Первое опубликованное (1838 г.) доказательство принадлежит Дирихле; общая формула называется формулой Дирихле для числа классов. В доказательстве Дирихле использовались бесконечные ряды, и оно было тесно

связано с его доказательством существования простых в арифметических прогрессиях. Несмотря на многочисленные попытки, *все еще нет элементарного доказательства* *) этой формулы, т. е. доказательства, использующего только целые числа и не содержащего никаких предельных переходов. Это весьма удивительно, ибо формула устанавливает равенство двух натуральных чисел. Элементарно даже не доказано, что $B > A$, хотя из формулы для числа классов следует, что это так.

Тот факт, что $B - A$ кратно p и даже что и A, и B кратны p, вполне элементарен. Квадратичные вычеты сравнимы с 1^2, 2^2, ..., $\left(\frac{1}{2}(p-1)\right)^2$, а эту сумму легко вычислить и доказать, что она кратна p **). Значит, A кратно p, а так как $A+B=1+2+\cdots+(p-1)=\frac{1}{2}p(p-1)$, то B также делится на p.

Имеются и другие формулы для $C(-p)$, эквивалентные (22); некоторые из них более удобны для вычислений, чем указанная формула (22). Мы отметили эту формулу, так как она просто формулируется и не требует рассмотрения нескольких частных случаев. Ряд формул можно обобщить и на случай, когда d не равно $-p$.

Что касается величины числа классов, то Гаусс, проделав большие вычисления, высказал предположение, что $C(d)$ стремится к бесконечности, если d стремится к бесконечности. Эту гипотезу впервые доказал в 1934 году Гельбронн (Heilbronn), его доказательство — важный вклад в аналитическую теорию чисел. Последний *известный* дискриминант, для которого $C(d)=1$, равен -163. Гельбронн и Линфут (Linfoot) доказали, что существует еще не более одного дискриминанта, обладающего этим свойством. Вероятно, такого дискриминанта вообще нет; Лемер (Lehmer) доказал, что такого дискриминанта нет вплоть до $-500\,000\,000$.

Замечания к главе VI. п. 1. Для квадратичных форм используют два разных обозначения. Одно из них употреблялось нами: $ax^2+bxy+cy^2$. Другое обозначение $ax^2+2bxy+cy^2$ предполагает четность второго коэффициента. Второе обозначение исключает такую форму, как x^2+xy+y^2, но, конечно, свойства этой формы

*) В 1927 году ленинградским математиком Б. А. Венковым было получено элементарное доказательство этой формулы для случая $p \neq 8k+7$.

**) Здесь $p > 3$. (*Прим. перев.*)

можно вывести из свойств допускаемой этим обозначением формы $2x^2 + 2xy + 2y^2$. Обозначение без множителя 2 употреблялось Лагранжем, Кронекером и Дедекиндом, второе обозначение употребляли Лежандр, Гаусс и Дирихле. Взглянув на эти великие имена, мы вправе предположить, что ни одно из обозначений не обладает решающим преимуществом перед другим; и действительно, некоторые результаты имеют более простой вид при первом обозначении, другие выглядят проще при втором.

Наиболее доступное изложение теории квадратичных форм можно найти в книгах Матье ([11]) и Диксона ([6,7,8])*). У Диксона, как и у нас, использовано обозначение Лагранжа, Матье пользуется обозначениями Гаусса. Доказательство различных результатов, которые приводятся в этой главе без доказательства, можно найти в упомянутых книгах. Изложение общей теории квадратичных форм имеется в книгах: B. W. Jones, The Arithmetic Theory oi Quadratic Forms (Carus Monograph, № 10, 1950) и G. L. Watson, Integral Quadratic Forms (Cambridge Math. Tracts, № 51, 1960).

п. 2. Формы, которые переводятся одна в другую подстановкой с определителем -1, называют несобственно эквивалентными. Использование подстановок с определителем -1 усложняет теорию автоморфизмов и для определенных, и для неопределенных форм.

п. 8. Из количества *собственных* представлений числа суммой двух квадратов, найденного в этом пункте, можно вывести, что полное число представлений (собственных и несобственных) равно $4(D_1 - D_3)$, как это указывалось в замечаниях к главе V; при этом n может быть как четным, так и нечетным.

п. 9. Об исследовании Якоби см. книгу Бахмана (Bachmann, Die Lehre von der Kreisteilung, Teubner, 1927). Доказательство формулы Дирихле для числа классов имеется в книгах ([18], vol. 1, p. 127—180), ([11], ch. 8) **).

О работах Гельбронна (Heilbronn) и Гельбронна и Линфута (Linfoot) см. Quart. J. of Math., 5 (1934), 150—160 и 293—301.

Примечание переводчика. Элементарное доказательство формулы для числа классов, принадлежащее Б. А. Венкову, изложено в его книге ([1], гл. 6) и в работе в *Mathematische Zeitschrift*, 33:3, 1931, 350—374.

---

*) Эта теория излагается также в книгах Б. А. Венкова ([1]) и Дирихле ([3]). (*Прим. перев.*)

**) Эта формула доказана также в книге Дирихле ([3], гл. 5). (*Прим. перев.*)

ГЛАВА VII

НЕКОТОРЫЕ ДИОФАНТОВЫ УРАВНЕНИЯ

1. Введение. Диофантовым или неопределенным уравнением называют уравнение, которое должно быть решено в целых числах. Мы уже встречали некоторые диофантовы уравнения, например, уравнение $x^2 + y^2 = n$ в главах V и VI, уравнение $x^2 - Ny^2 = 1$ в главе IV.

Вероятно, ни одна из областей теории чисел не сталкивается с такими трудностями, как теория диофантовых уравнений, если исследование диофантовых уравнений можно назвать теорией. При беглом взгляде на обширную литературу создается впечатление, что установлено с помощью различных искусственных приемов много результатов, связанных с отдельными уравнениями; кажется весьма затруднительным объединить эти результаты в общую теорию. Иногда, решив уравнение искусственным приемом, удается создать общую теорию, связанную с найденным решением, разумно объясняющую возникновение этого решения и показывающую, насколько найденное решение можно обобщить. Но внутренние трудности предмета настолько велики, что область применения такой теории обычно очень ограничена. Если удается развить достаточно глубокую теорию диофантовых уравнений специального вида (например, теорию квадратичных форм), то такая теория получает право на самостоятельное существование.

В этой главе мы рассмотрим некоторые диофантовы уравнения, допускающие элементарное исследование, и укажем, где это возможно, общие теории, связанные с этими уравнениями.

2. Уравнение $x^2+y^2 = z^2$. Целочисленные решения этого уравнения (например, $3^2 + 4^2 = 5^2$) были известны издавна. Вавилонские таблицы, датируемые примерно 1700 годом до нашей эры, содержат обширный список решений, причем некоторые

из этих решений довольно велики. Это уравнение интересовало греческих математиков в связи с теоремой Пифагора; его общее решение дал Евклид (книга X, лемма 1 к предложению 29).

Разделив уравнение на z^2 и вводя новые переменные $x/z = X$, $y/z = Y$, получаем

$$X^2 + Y^2 = 1; \tag{1}$$

задача сводится к нахождению решения этого уравнения в рациональных X, Y. Решение этого уравнения основано на представлении его в виде

$$Y^2 = 1 - X^2 = (1 - X)(1 + X).$$

Мы не можем выразить X как рациональную функцию от $(1 - X)(1 + X)$, но X рационально выражается через $\frac{1-X}{1+X}$. Разделив обе части уравнения на $(1 + X)^2$, получим

$$\left(\frac{Y}{1 + X}\right)^2 = \frac{1 - X}{1 + X}.$$

Если положить $t = \frac{Y}{1+X}$, то X и Y будут рациональными функциями от t, именно:

$$\frac{1 - X}{1 + X} = t^2,$$

откуда

$$X = \frac{1 - t^2}{1 + t^2}, \qquad Y = \frac{2t}{1 + t^2}. \tag{2}$$

Для каждого рационального значения t эти формулы дают рациональные значения X, Y, удовлетворяющие (1). Обратно, каждое рациональное решение (1) получается таким образом (за исключением решения $X = -1$, $Y = 0$, которое получается в пределе при стремлении t к бесконечности, но не представимо в виде (2)).

На предыдущее рассуждение можно посмотреть и с геометрической точки зрения. Уравнение $X^2 + Y^2 = 1$ есть уравнение окружности с центром в начале координат и радиусом 1. Возьмем какую-нибудь точку на этой окружности, скажем точку $X = -1$, $Y = 0$. Переменная прямая, проходящая через эту точку, пересечет окружность в некоторой другой точке (если эта прямая не является касательной), координаты этой точки можно найти из уравнений окружности и прямой с помощью рацио-

нальных операций. Переменная прямая, проходящая через точку $(-1, 0)$, имеет уравнение вида $Y = t(X + 1)$; по формуле (2) можно выразить координаты точки пересечения через t. Аналогичный метод можно применить для нахождения рациональных точек на любой кривой второго порядка в предположении, что уравнение этой кривой имеет рациональные коэффициенты и что на ней можно найти хотя бы одну рациональную точку. Однако рациональных точек на кривой с рациональными коэффициентами может и не быть; например, на кривой $X^2+Y^2 = 3$ нет рациональных точек. Но даже если на кривой второго порядка и есть рациональные точки, часто нелегко найти хотя бы одну из них.

Формулы (2), в которых t — произвольное рациональное число, дают общее решение уравнения $X^2+Y^2 = 1$ в рациональных числах, а потому они в принципе дают и общее решение уравнения

$$x^2 + y^2 = z^2 \tag{3}$$

в целых числах. Но переход от рациональных решений уравнения (1) к целым решениям (3) все же заслуживает рассмотрения, так как иногда (в других задачах) такой переход представляет серьезные трудности. Положим $t = \frac{q}{p}$, где p и q — взаимно простые целые числа. Тогда в силу (2)

$$\frac{x}{z} = \frac{p^2 - q^2}{p^2 + q^2}, \quad \frac{y}{z} = \frac{2pq}{p^2 + q^2}. \tag{4}$$

В качестве x, y, z *можно*, конечно, взять числа $p^2 - q^2$, $2pq$, $p^2 + q^2$ или числа $a(p^2 - q^2)$, $2apq$, $a(p^2 + q^2)$, но x, y, z *не обязаны* иметь такой вид. Если три числа p^2-q^2, $2pq$, p^2+q^2 имеют общий множитель, больший 1, то можно разделить их на этот общий множитель и получить новое решение (3) в целых числах.

Рассмотрим две возможности для взаимно простых чисел p и q. Предположим сначала, что одно из них четно, а другое нечетно. Тогда три числа p^2-q^2, $2pq$, p^2+q^2 не имеют общих множителей, больших 1; действительно, такой множитель должен быть нечетным (ибо $p^2 - q^2$ нечетно) и должен делить $(p^2 - q^2) + (p^2 + q^2) = 2p^2$, аналогично этот множитель должен делить $2q^2$, а это невозможно, так как p и q взаимно просты. Значит, в этом случае из (4) следует, что

$$x = m(p^2 - q^2), \quad y = 2mpq, \quad z = m(p^2 + q^2), \tag{5}$$

где m — целое число.

Рассмотрим теперь случай, когда числа p и q нечетны. Так, полагая $p+q = 2P$ и $p-q = 2Q$, получим два взаимно простых числа P и Q. Одно из них четно, а другое нечетно (так. как $p = P+Q$ нечетно). Подставив в (4) вместо p и q их выражения через P и Q, получим после сокращения на 2

$$\frac{x}{z} = \frac{2PQ}{P^2 + Q^2}, \qquad \frac{y}{z} = \frac{P^2 - Q^2}{P^2 + Q^2}.$$

В результате получились уравнения, аналогичные прежним, только x и y поменялись местами и вместо p и q стоят P и Q.

Следовательно, *все решения уравнения* $x^2 + y^2 = z^2$ *в целых числах задаются формулой* (5), *где* m, p, q — *целые числа,* p *и* q *взаимно просты, причем одно из чисел* p *и* q *четно, а другое нечетно*; кроме того, в формулах (5) можно, конечно, заменить x на y и y на x.

Это — формулы Евклида. Простейшим решением (кроме тривиального решения, в котором одно из неизвестных равно нулю) является решение $x = 3$, $y = 4$, $z = 5$, получающееся при $m = 1$, $p = 2$, $q = 1$. Приведем несколько первых примитивных решений (т. е. решений, в которых x, y, z взаимно просты, так что $m = 1$): $(3, 4, 5)$, $(5, 12, 13)$, $(8, 15, 17)$, $(7, 24, 25)$, $(21, 20, 29)$, $(9, 40, 41)$.

Так как формула для z (если взять $m = 1$) имеет вид $z = p^2 + q^2$, то мы можем сделать z точным квадратом, выбрав соответствующим образом p и q, и, значит, получить параметрическое решение уравнения $x^2 + y^2 = z^4$. Повторение этого процесса дает возможность найти решение уравнения $x^2 + y^2 = z^k$, где число k равно степени числа 2. Кроме того, формулы для этого уравнения можно вывести и из формул для уравнения $x^2 + y^2 = z^2$, применив тождества (1) главы V.

3. Уравнение $ax^2 + by^2 = z^2$. Метод, использованный ранее для уравнения $x^2 + y^2 = z^2$, приводит к формуле для общего решения уравнения $ax^2 + y^2 = z^2$. В этом случае также имеется бесконечно много примитивных решений. Но вышеописанный метод неприменим к более общему уравнению

$$ax^2 + by^2 = z^2, \tag{6}$$

где a и b — натуральные числа, ни одно из которых не является точным квадратом.

Отметим прежде всего, что такое уравнение может вообще не иметь решения (кроме тривиального решения $x = y = z = 0$, которое мы не рассматриваем). Например, уравнение

$$2x^2 + 3y^2 = z^2$$

не разрешимо. В самом деле, можно считать, что x, y, z не имеют общего множителя, большего 1, откуда, в частности, следует, что ни x, ни z не делятся на 3. Но тогда сравнение $2x^2 \equiv z^2 \pmod 3$ невозможно, ибо 2 — квадратичный невычет по модулю 3.

Аналогичные рассмотрения применимы и к общему уравнению (6) и дают необходимые условия разрешимости этого уравнения; эти условия требуют разрешимости некоторых сравнений. Мы можем предполагать, что a и b *свободны от квадратов*, т. е. не делятся ни на какой квадрат, больший 1, ибо введение в коэффициенты a и b квадратных множителей не влияет на разрешимость уравнения.

Если уравнение (6) разрешимо, то, разделив в его решении числа x, y, z на их общий множитель, мы получим решение этого уравнения, в котором x, y, z не имеют отличного от 1 общего множителя. Из уравнения (6) вытекает сравнение $ax^2 \equiv z^2 \pmod b$. При этом x и b должны быть взаимно просты; действительно, если бы они имели общий простой множитель, то этот множитель должен был бы делить x и z, поэтому by^2 делилось бы на квадрат этого множителя; с другой стороны, так как b свободно от квадратов, то этот простой множитель делил бы y, что невозможно. Умножая сравнение на x'^2, где $xx' \equiv 1 \pmod b$, получаем сравнение вида

$$a \equiv \alpha^2 \pmod b, \tag{7}$$

в котором $\alpha = x'z$. Аналогично

$$b \equiv \beta^2 \pmod a \tag{8}$$

для некоторого целого β. Другими словами, a должно быть квадратичным вычетом по mod b, а b — квадратичным вычетом

по mod a. Здесь мы используем термин *квадратичный вычет* в более общем смысле, чем в главе III: модули a и b не обязаны быть взаимно простыми.

Если Н. О. Д. чисел a и b равен $h > 1$, то, помимо сравнений (7) и (8), для разрешимости уравнения (6) должно выполняться еще одно сравнение. Положим $a = ha_1$ и $b = hb_1$, так что a_1, b_1, h попарно взаимно просты. В каждом решении (6) z должно делиться на h, значит, $a_1x^2 + b_1y^2$ также должно делиться на h. Умножая $a_1x^2 + b_1y^2$ на $b_1x'^2$, получаем сравнение вида

$$a_1b_1 \equiv -\gamma^2 \pmod{h}. \tag{9}$$

Разрешимость сравнений (7), (8), (9) накладывает ограничения на a и b, необходимые для разрешимости уравнения (6). Заранее, однако, не очевидно, что если сравнения (7), (8), (9) разрешимы, то разрешимо и уравнение (6). Докажем теперь, следуя Лежандру, что в действительности это все же так. Мы хотим доказать таким образом, что *уравнение* (6), *в котором* a *и* b *— свободные от квадратов натуральные числа, разрешимо тогда и только тогда, когда разрешимы сравнения* (7), (8), (9).

Если a или b равно 1, то уравнение (6), очевидно, разрешимо. Если $a = b$, то условия (7) и (8) удовлетворяются тривиальным образом, а (9) приводит к $1 \equiv -\gamma^2 \pmod{a}$. В силу (VI, 5) отсюда следует, что a представимо в виде $p^2 + q^2$, и наше уравнение удовлетворяется при $x = p$, $y = q$, $z = p^2 + q^2$.

Предположим теперь, что $a > b > 1$. План доказательства состоит в том, чтобы вывести из (6) аналогичное уравнение с тем же самым b и с A вместо a, где $0 < A < a$ и A, b удовлетворяют таким же трем сравнениям, как a и b. Повторение этого процесса приводит к уравнению, в котором либо один из коэффициентов равен 1, либо оба коэффициента равны между собой. Как мы видели, такое уравнение разрешимо.

По предположению сравнение (8) разрешимо. Выберем решение β, для которого $|\beta| \leqslant \frac{1}{2}a$. Число $\beta^2 - b$ кратно a, поэтому можно выбрать A и k так, чтобы выполнялось сравнение

$$\beta^2 - b = aAk^2, \tag{10}$$

где k и A — целые числа и A свободно от квадратов (все квадратные множители $\beta^2 - b$ входят в k^2). Заметим, что k взаимно

просто с b, так как b свободно от квадратов. Отметим также, что A положительно, так как

$$aAk^2 = \beta^2 - b > -b > -a,$$

значит, $Ak^2 \geqslant 0$ и поэтому >0, ибо b не является точным квадратом.

Подставив выражения y и z через новые переменные Y и Z по формулам*)

$$z = bY + \beta Z, \quad y = \beta Y + Z, \tag{11}$$

получим

$$z^2 - by^2 = (\beta^2 - b)(Z^2 - bY^2).$$

Ввиду (10) уравнение (6) в новых переменных x, Y, Z принимает вид

$$aAk^2(Z^2 - bY^2) = ax^2.$$

Полагая $x = kAX$, получим новое уравнение

$$AX^2 + bY^2 = Z^2.$$

Если это уравнение разрешимо, то разрешимо и (6), так как подстановка (11) и уравнение $x = kAX$ дают целые, отличные от нуля значения x, y, z по X, Y, Z.

Новый коэффициент a положителен, свободен от квадратов и удовлетворяет условию

$$A = \frac{1}{ak^2}(\beta^2 - b) < \frac{\beta^2}{ak^2} \leqslant \frac{\beta^2}{a} \leqslant \frac{a}{4},$$

так что A меньше a. Остается доказать, что A и b удовлетворяют сравнениям, аналогичным (7), (8), (9). Аналог (8) очевиден, ибо $b \equiv \beta^2 \pmod{A}$ ввиду (10).

Для доказательства аналога (7) заметим, что (10) можно разделить на h, получив равенство

$$h\beta_1^2 - b_1 = a_1 Ak^2.$$

Кроме того, (7) эквивалентно сравнению $a_1 \equiv h\alpha_1^2 \pmod{b_1}$. Следовательно,

$$h\beta_1^2 \equiv hA(\alpha_1 k)^2 \pmod{b_1},$$

*) Вид подстановки (11) определяется равенством

$$z - y\sqrt{b} = (\beta - \sqrt{b})(Z - Y\sqrt{b}).$$

а так как h, k, a_1 взаимно просты с b_1, то отсюда следует, что A сравнимо с квадратом по mod b_1. Далее, $-a_1Ak^2 \equiv b_1 \pmod{h}$ ввиду (9) и того факта, что числа k, a_1, b_1 взаимно просты с h; отсюда следует, что A сравнимо с квадратом по mod h, а значит, также и по mod b; это — аналог сравнения (7).

Для доказательства аналога (9) с A вместо a обозначим через H наибольший общий делитель A и b, и пусть $A = HA_2$, $b = Hb_2$. Разделив равенство (10) на H, получим

$$H\beta_2^2 - b_2 = aA_2k^2.$$

Отсюда

$$-A_2b_2 \equiv a(A_2k)^2 \pmod{H}.$$

Так как $a \equiv \alpha^2 \pmod{H}$ в силу (7), отсюда следует, что $-A_2b_2$ сравнимо с квадратом по mod H, а это и есть аналог (9).

Мы показали, что коэффициенты A и b удовлетворяют условиям, которым должны были удовлетворять числа a и b. Поэтому здесь применим вышеописанный метод доказательства, который и устанавливает разрешимость уравнения (6).

Для иллюстрации проведенного доказательства применим описанный процесс к уравнению

$$41x^2 + 31y^2 = z^2. \tag{12}$$

Так как здесь коэффициенты взаимно просты, то остаются только два условия:

$$41 \equiv \alpha^2 \pmod{31} \quad \text{и} \quad 31 \equiv \beta^2 \pmod{41}.$$

Оба сравнения разрешимы:

$$\alpha \equiv \pm 14 \pmod{31}, \qquad \beta \equiv \pm 20 \pmod{41}.$$

Здесь из разрешимости одного сравнения следует разрешимость другого, благодаря квадратичному закону взаимности: числа 31 и 41 — простые и не каждое из них имеет вид $4k + 3$.

Согласно вышеописанному методу, нужно выбрать некоторое значение β и затем определить A и k по формуле (10). В рассматриваемой теории предполагалось, что $|\beta| \leqslant \frac{1}{2}a$; возьмем поэтому $\beta = 20$, тогда $\beta^2 - b = 400 - 31 = 9 \cdot 41$, откуда $k = 3$, $A = 1$. (Так как $A = 1$, то продолжать процесс нет необходимости.) Получающееся из (12) новое уравнение имеет вид

$$X^2 + 31Y^2 = Z^2,$$

и мы можем взять очевидное решение $X = 1$, $Y = 0$, $Z = 1$. Соотношения между x, y, z и X, Y, Z в данном случае таковы:

$$z = 31Y + 20Z, \quad y = 20Y + Z, \quad x = 3X.$$

Это дает решение $x = 3$, $y = 1$, $z = 20$ для исходного уравнения (12).

Вернемся теперь к общей теории. Мы доказали, что разрешимость сравнений (7), (8), (9) есть необходимое и достаточное условие разрешимости уравнения (6) в предположении, что a и b свободны от квадратов. Лежандр легко вывел отсюда необходимое и достаточное условие разрешимости уравнения $ax^2 + by^2 = cz^2$, где a, b, c — натуральные числа. В предположении, что a, b, c свободны от квадратов и попарно взаимно просты (здесь эти ограничения не являются существенными ограничениями), это условие состоит в разрешимости следующих трех сравнений:

$$bc \equiv \alpha^2 \pmod{a}, \quad ca \equiv \beta^2 \pmod{b}, \quad ab \equiv -\gamma^2 \pmod{c}.$$

В заключение этого пункта сделаем несколько замечаний об условиях разрешимости уравнения, которые выражаются в терминах сравнений. Каждое диофантово уравнение приводит к сравнению по любому выбранному нами модулю, и каждое такое сравнение должно быть разрешимо, если разрешимо уравнение. Но обычно имеется только конечное число модулей, разрешимость сравнений по которым накладывает какие-либо условия на коэффициенты уравнения. Получающиеся условия являются *необходимыми* условиями разрешимости уравнения. Эти условия не всегда достаточны; выяснение связей между разрешимостью сравнений и уравнений приводит к рассмотрению глубоких и тонких вопросов. Как мы уже говорили, разрешимость нескольких сравнений есть необходимое и достаточное условие разрешимости уравнения Лежандра $ax^2 + by^2 = cz^2$. В 1923 году Хассе доказал, что аналогичный результат имеет место для однородных квадратных уравнений от любого числа переменных.

Мы уже встречались с различными примерами, в которых путем рассмотрения сравнений доказывалась неразрешимость уравнения. Иногда можно доказать неразрешимость уравнения,

используя сравнение по модулю, *зависящему от неизвестных уравнения*. В этом основная идея доказательства неразрешимости уравнения

$$y^2 = x^3 + 7.$$

Это доказательство дал Лебег (V. A. Lebesgue) в 1869 году. Во-первых, x должно быть нечетно, так как число вида $8k + 7$ не может быть точным квадратом. Перепишем теперь уравнение в виде

$$y^2 + 1 = x^3 + 8 = (x + 2)(x^2 - 2x + 4).$$

Число $x^2 - 2x + 4 = (x-1)^2 + 3$ есть число вида $4k+3$. Значит, оно имеет некоторый простой делитель q такого вида, но сравнение $y^2 + 1 \equiv 0 \pmod{q}$ не разрешимо, следовательно, не разрешимо и рассматриваемое уравнение.

4. Проблема Ферма. О многих открытиях Ферма мы узнали из пометок, которые он делал на полях имевшегося у него экземпляра «*Арифметики*» Диофанта. Рядом с исследованием Диофанта уравнения $x^2 + y^2 = z^2$ Ферма написал: «... тем не менее, нельзя представить куб в виде суммы двух кубов, четвертую степень в виде суммы двух четвертых степеней и вообще какую-либо степень, большую второй, как сумму двух таких же степеней. Я открыл поистине удивительное доказательство, но поля слишком малы, чтобы его поместить». Эту известную гипотезу Ферма обычно называют последней теоремой Ферма: уравнение

$$x^n + y^n = z^n \tag{13}$$

не разрешимо в натуральных числах x, y, z, если $n > 2$ целое число. Несмотря на трехсотлетние попытки многих величайших математиков, это предложение остается недоказанным в общем виде, хотя истинность его доказана для всех n, не превосходящих 600. Весьма вероятно, что Ферма ошибся, думая, что обладает доказательством этой теоремы.

Привлекательность задачи лежит частично в интригующей простоте формулировки. По этой причине на нее набросились многие любители, смелость которых значительно превосходит их математические способности, так что эта теорема выделяется среди арифметических задач наибольшим числом ее невер-

ных «доказательств».

По-видимому, любой новый метод, развитый для доказательства гипотезы Ферма, явился бы важным вкладом в теорию чисел. Примером этого могут служить работы Куммера (1810—1893). Куммер был уверен вначале, что доказал гипотезу Ферма. Ошибку в его рассуждениях указал ему Дирихле, попытки Куммера исправить ошибку привели его к созданию новой далеко идущей теории — теории *идеалов* в полях алгебраических чисел.

В этом элементарном обзоре мы вынуждены ограничиться доказательством гипотезы Ферма для какого-нибудь частного значения n. Простейшим является случай $n = 4$, в котором неразрешимость уравнения была доказана самим Ферма.

Ферма получил даже более общий результат: он доказал, что уравнение

$$x^4 + y^4 = z^2 \tag{14}$$

не имеет решений в натуральных числах, его доказательство дает простой пример «бесконечного спуска», являющегося особой формой метода доказательства по индукции. Исходя из любого предполагаемого решения уравнения (14) в натуральных числах, Ферма строит другое решение с меньшим значением z. Повторение этого процесса необходимо ведет к противоречию, так как не существует бесконечно убывающих последовательностей натуральных чисел. Принцип доказательства такой же, как и в доказательстве Лежандра, описанном в предыдущем пункте, только здесь он применяется для доказательства неразрешимости уравнения, в то время как там он применялся для доказательства разрешимости уравнения.

Пусть x, y, z — натуральные числа, удовлетворяющие (14). Можно предполагать, что x и y не имеют общего делителя, большего 1, так как на четвертую степень такого общего делителя уравнение можно сократить. В этом предположении числа x^2, y^2, z^2 образуют примитивное решение уравнения $X^2 + Y^2 = Z^2$, поэтому в силу установленного в п. 2 результата они представимы (возможно, после изменения порядка x и y) в виде

$$x^2 = p^2 - q^2, \quad y^2 = 2pq, \quad z^2 = p^2 + q^2,$$

где p и q — взаимно простые натуральные числа, одно из кото-

рых четно, а другое нечетно. Посмотрев на первое равенство и вспоминая, что любой квадрат должен быть сравним с 0 или 1 по mod 4, мы видим, что p нечетно, а q четно. Полагая $q = 2r$, имеем

$$x^2 = p^2 - (2r)^2, \quad \left(\tfrac{1}{2}y\right)^2 = pr$$

Так как p и r взаимно просты и их произведение является точным квадратом, то каждое из них должно быть точным квадратом. Полагая $p = v^2$ и $r = w^2$, приводим первое уравнение к виду

$$x^2 + (2w^2)^2 = v^4.$$

Это уравнение похоже на исходное уравнение (14). Применив аналогичное рассуждение к новому уравнению, мы получим уравнение, в точности подобное (14). Из последнего уравнения следует, что

$$x = P^2 - Q^2, \quad 2w^2 = 2PQ, \quad v^2 = P^2 + Q^2,$$

где P и Q — взаимно простые натуральные числа, одно из которых четно, а другое нечетно. Так как $PQ = w^2$, то P и Q должны быть точными квадратами. Если $P = X^2$, $Q = Y^2$, то рассматриваемое уравнение принимает вид

$$X^4 + Y^4 = v^2,$$

что по форме совпадает с (14). В этом уравнении X, Y, v — натуральные числа и

$$v^2 = p < \sqrt{z},$$

откуда $v < z$. Ввиду сказанного ранее, этим неразрешимость уравнения (14) доказана.

Почти все современные исследования проблемы Ферма основаны на работе Куммера. В них доказывается, что если n удовлетворяет одному из некоторых условий, то уравнение (13) не разрешимо. Достаточно рассматривать простые значения n, большие двух, так как любое число, большее двух, либо делится на некоторое простое число, большее двух, либо делится на 4; и если уравнение не разрешимо для какого-нибудь значения n, то оно и подавно не разрешимо для любого кратного этого значения. До сих пор, когда находилось число n, не удовлетворяющее ни одному из существующих критериев, всегда удавалось спра-

виться с ним с помощью какого-нибудь нового критерия.

5. Уравнение $x^3 + y^3 = z^3 + w^3$. Хотя уравнение $x^3 + y^3 = z^3$ (частный случай уравнения Ферма) не разрешимо, уравнение $x^3 + y^3 = z^3 + w^3$ имеет бесконечно много решений, отличных от очевидных решений, в которых $x = z$ или $x = w$ или $x = -y$. Формулы, дающие решение, нашел Виет (Viete) в 1591 году, но формулы, открытые в 1756—1760 годах Эйлером, являются более общими. В 1841 году формулы Эйлера упростил Бине (Binet).

Чтобы исследовать уравнение

$$x^3 + y^3 = z^3 + w^3, \tag{15}$$

положим $x + y = X$, $x - y = Y$, $z + w = Z$, $z - w = W$. В этих обозначениях наше уравнение принимает вид

$$X(X^2 + 3Y^2) = Z(Z^2 + 3W^2), \tag{16}$$

Имеет место тождество, аналогичное тождеству (1) главы V, оно выражает произведение двух чисел вида $X^2 + 3Y^2$ в такой же форме, именно:

$$(X^2 + 3Y^2)(Z^2 + 3W^2) = (XZ + 3YW)^2 + 3(YZ - XW)^2.$$

Умножая (16) на X^2+3Y^2 и деля на Z, мы получаем с помощью этого тождества такое равенство:

$$\frac{X}{Z}(X^2 + 3Y^2)^2 = (XZ + 3YW)^2 + 3(YZ - XW)^2.$$

Это равенство показывает, что рациональное число $\frac{X}{Z}$ имеет вид p^2+3q^2, где p и q — рациональные числа, задаваемые формулами

$$p = \frac{XZ + 3YW}{X^2 + 3Y^2}, \qquad q = \frac{YZ - XW}{X^2 + 3Y^2}. \tag{17}$$

Для упрощения выкладок положим $Z = 1$ и будем считать X, Y, W рациональными числами. В силу (17) с $Z = 1$ имеем

$$pX + 3qY = 1, \quad pY - qX = W.$$

Эти формулы дают возможность выразить Y и W в терминах p, q и X, где $X = p^2 + 3q^2$. Они дают

$$3qY = 1 - pX, \quad 3qW = p - X^2.$$

Возвращаясь к x, y, z, w и отбрасывая их общий знаменатель, получаем

$$\begin{cases} x = 1 - (p - 3q)(p^2 + 3q^2), & y = -1 + (p + 3q)(p^2 + 3q^2), \\ z = p + 3q - (p^2 + 3q^2)^2, & w = -(p - 3q) + (p^2 + 3q^2)^2. \end{cases} \quad (18)$$

Это и есть формулы Эйлера и Бине. Для любых рациональных чисел p и q эти формулы дают рациональные значения x, y, z, w, удовлетворяющие уравнению (15); наше доказательство показывает, что верно и обратное: каждое рациональное решение (15) пропорционально решению, получаемому по этим формулам.

Если, в частности, придавать p и q целые значения, мы получим целые решения (15), но нет оснований считать, что так можно получить любое целое решение. Одно частное решение получается при $p = 1$, $q = 1$; это решение $x = 9$, $y = 15$, $z = -12$, $w = 18$ приводит к любопытному тождеству: $3^3 + 4^3 + 5^3 = 6^3$. Значения $p = 4$, $q = 1$ отвечают равенству

$$3^3 + 60^3 = 22^3 + 59^3.$$

Простейшее решение (15) с положительными x, y, z, w есть

$$1^3 + 12^3 = 9^3 + 10^3.$$

На самом деле число 1729 — наименьшее из целых чисел, представимых суммой двух положительных целых кубов двумя различными способами *).

Интересное тождество, на которое обратил внимание в 1936 году К. Малер, получается при $p = 3q$; тогда

$$x = 1, \quad y = -1 + 72q^3, \quad z = 6q - 144q^4, \quad w = 144q^4.$$

Полагая $2q = t$, получим тождество

$$(1 - 9t^3)^3 + (3t - 9t^4)^3 + (9t^4)^3 = 1.$$

Оно интересно тем, что показывает, как можно представить число 1 бесконечным числом способов в виде суммы трех це-

*) Навестив больного Рамануджана, лежавшего в Патней (Putney), Харди между прочим сказал ему, что приехал на такси № 1729 и что это число, как ему кажется, выделяется каким-то специальным свойством. Рамануджан тут же указал это свойство.

лых кубов. Аналогичное тождество имеет место для числа 2. Мне неизвестно какое бы то ни было тождество, устанавливающее представимость числа 3 в виде суммы трех целых кубов бесконечно многими способами.

Упомянем здесь еще об одной нерешенной задаче. Не каждое число представимо в виде суммы трех целых кубов; в самом деле, ни одно число, сравнимое с 4 или 5 по mod 9, нельзя представить таким образом. Действительно, легко видеть, что любой куб сравним с 0, 1 или -1 по модулю 9 и, следовательно, сумма любых трех кубов должна быть сравнима с 0, ± 1, ± 2 или ± 3 по mod 9 и не может быть сравнима с ± 4. Спрашивается, *каждое ли число представимо в виде суммы четырех целых кубов*? Несмотря на многочисленные попытки, эта задача до сих пор не решена.

Существует очень простой способ представления любого числа в виде суммы пяти целых кубов. Имеем

$$(x+1)^3 + (x-1)^3 + (-x)^3 + (-x)^3 = 6x.$$

Значит, каждое кратное 6 представляется в виде суммы четырех целых кубов. Кроме того, любое число можно привести к кратному 6, вычтя из него подходящий куб. В самом деле, легко видеть, что $n - n^3$ всегда делится на 6. Это дает результат, который впервые, по-видимому, доказал Ольтрамаре (Oltramare) в 1894 году.

6. Теорема Туэ—Зигеля—Рота. Многие современные исследования диофантовых уравнений основаны на методе, впервые примененном норвежским математиком Акселем Туэ в 1908 году. Этот метод связан с рассмотрением рациональных приближений алгебраического числа; мы поясним, о чем идет речь.

Пусть $f(x, y)$ — какая-либо однородная форма от x и y степени n, скажем

$$f(x, y) = a_0 x^n + a_1 x^{n-1} y + \ldots + a_{n-1} x y^{n-1} + a_n y^n,$$

где a_0, a_1, …, a_n — целые и n не меньше 3. Предположим, что эта форма неприводима, т. е. не может быть представлена в виде

произведения двух форм с рациональными коэффициентами *). В силу так называемой основной теоремы алгебры эту форму можно разложить на множители следующим образом:

$$f(x, y) = a_0(x - \theta_1 y)(x - \theta_2 y) \dots (x - \theta_n y),$$

где θ_1, θ_2, ..., θ_n — иррациональные числа, действительные или комплексные. Эти числа являются корнями неприводимого алгебраического уравнения

$$a_0\theta^n + a_1\theta^{n-1} + \dots + a_n = 0$$

и называются *алгебраическими числами* степени n.

При любых целых x и y полином $f(x, y)$ является целым числом. Значит, если x и y не равны нулю одновременно, то

$$|a_0(x - \theta_1 y)(x - \theta_2 y) \dots (x - \theta_n y)| \geqslant 1.$$

Предположим, что $\frac{x}{y}$ есть рациональное приближение к θ_1 и y — большое положительное число. Тогда все множители $x - \theta_2 y$, ... меньше (по модулю) некоторого постоянного кратного y, откуда, разделив на y^n, получим

$$\left|\frac{x}{y} - \theta_1\right| > \frac{K}{y^n}, \tag{19}$$

где K — положительная постоянная, зависящая от коэффициентов формы f. Таким образом, алгебраическое число степени n не может обладать последовательностью рациональных приближений, сходящихся к нему слишком быстро. Этот результат установил в 1844 году Лиувилль, что дало ему возможность построить числа, не являющиеся алгебраическими.

Туэ длинным и трудным способом доказал, что имеет место более сильное неравенство, именно:

$$\left|\frac{x}{y} - \theta_1\right| > \frac{K}{y^\nu} \tag{20}$$

для всех, кроме конечного числа, рациональных приближений к θ_1, где ν — любое число, большее $\frac{1}{2}n + 1$. Число $\frac{1}{2}n + 1$ умень-

*) Здесь не важно, будем ли мы считать коэффициенты полиномов, участвующих в разложении, рациональными или целыми, так как можно доказать, что если имеется разложение на множители с рациональными коэффициентами, то имеется и разложение с целыми коэффициентами.

шено в 1921 году Зигелем до $2\sqrt{n}$, а в 1947 году Дайсон (Dyson) уменьшил это число до $\sqrt{2n}$.

В 1955 году Рот доказал замечательную теорему о том, что если ν — какое-нибудь число, большее 2, то неравенство (20) имеет место для всех, кроме конечного числа, рациональных приближений θ_1. Это наилучший из возможных результатов такого рода, ибо, как мы видели в (IV, 7), неравенство $\left|\frac{x}{y}-\theta_1\right|<\frac{1}{y^2}$ всегда имеет бесконечно много решений, если θ_1 иррациональное (безразлично, алгебраическое или трансцендентное). Доказательство Рота является, конечно, очень трудным.

Неравенство (20) дает нижнюю границу для значения формы $f(x, y)$. Если x, y — любые большие целые числа, для которых число $|f(x, y)|$ мало по сравнению с $|y|^n$, то $\frac{x}{y}$ должно быть рациональным приближением к одному из корней $\theta_1, \ldots, \theta_n$. Предположим, не нарушая общности, что $\frac{x}{y}$ является приближением к θ_1; из (20) тогда следует неравенство

$$|f(x, y)| > K_1 y^{n-\nu},$$

где K_1 — некоторая положительная постоянная. В качестве ν по теореме Рота можно взять любое число, большее 2. Поэтому любое диофантово уравнение, из которого следует, что $f(x, y)$ меньше определенной степени $|y|$, может иметь лишь конечное число решений. Пусть, в частности, $g(x, y)$ — полином, быть может, неоднородный, в котором каждый член имеет степень меньшую, чем $n-2$. Тогда уравнение

$$f(x, y) = g(x, y)$$

имеет конечное число решений. Например, это утверждение верно, если $g(x, y)$ является константой. Существенно, конечно, что n должно быть не меньше 3: как известно, уравнение Пелля $x^2 - Ny^2 = 1$ степени 2 имеет бесконечно много решений.

В качестве примера можно рассмотреть любое уравнение вида

$$ax^4 + bx^3y + cx^2y^2 + dxy^3 + fy^4 = kx + ly + m.$$

Если левая часть неприводима, то это уравнение имеет только конечное число решений. Действительно, правая часть равенства степени 1, и $1 < n-2$ при $n = 4$.

Метод Туэ — Зигеля — Рота обладает одной особенностью. Устанавливая, что различные типы уравнений от двух переменных имеют лишь конечное число решений, он не дает каких-либо промежутков для x и y, в которых решений нет. Причина этого недостатка метода в том, что он основан на рассмотрении *двух или более* предполагаемых приближений к алгебраическому числу. Противоречие получается, если все они «слишком хороши». Поэтому, вообще говоря, можно указать пределы (в каждом конкретном случае), в которых уравнение имеет *не более одного* решения или не более заданного числа решений, но не пределы, в которых уравнение не имеет решений.

Замечания к главе VII. п. 3. Об уравнении $ax^2 + by^2 = cz^2$ см. также L. J. Mordell, Monatshefte für Math., **55** (1951), 323—327.

Имеется теорема Диксона, устанавливающая, что если уравнение $ax^2 + by^2 = cz^2$ разрешимо (a, b, c свободны от квадратов и попарно взаимно просты), то форма $ax^2 + by^2 - cz^2$ представляет все целые числа. Таким образом, из указанного в тексте примера следует, что любое число представимо формой $41x^2 + 31y^2 - z^2$.

Интересный обзор различных методов исследования уравнения $y^2 = x^3 + k$ можно найти в книге L. J. Mordell, A chapter in the theory of numbers, Cambridge, 1921.

п. 5. См. ([7], vol. II, ch. 21) и K. Mahler, J. London Math. Soc., **11** (1936), 136—138.

О случае с Рамануджаном см. записки Харди в Collected Papers of S. Ramanujan (Cambridge, 1927) или Proc. London Math. Soc. (2), **19** (1921). О проблеме четырех кубов см. H. W. Richmond, Messenger of Math., **51** (1922), 177—186 и L. J. Mordell, J. London Math. Soc., **11** (1936), 208—218.

п. 6. Теорема Рота была опубликована в Mathematika, **2** (1955), 1—20 (с исправлением на стр. 168). Немного другие формы доказательства приводятся в книгах: Cassels, Introduction to Diophantine Approximation (Cambridge Math. Tracts № 45, 1957), ([10], vol. II) и Schneider, Einführung in die transzendenten Zahlen (Springer, 1957).

БИБЛИОГРАФИЯ

Этот список содержит перечень книг по общей теории чисел. Ссылки на работы по специальным областям предмета имеются в замечаниях, приведенных в конце каждой главы.

На русском языке *)

([1]) Б. А. Венков, Элементарная теория чисел, М. — Л., 1937.

([2]) И. М. Виноградов, Основы теории чисел, М., «Наука», 1965.

([3]) П. Г. Лежен-Дирихле, Лекции по теории чисел, М. — Л., 1936.

([4]) Г. Хассе, Лекции по теории чисел, М., ИЛ, 1953.

([5]) А. О. Гельфонд, Ю. В. Линник, Элементарные методы в аналитической теории чисел, М., Физматгиз, 1962.

На английском языке

([6]) L. E. Dickson, Introduction to the Theory of Numbers (Chicago University Press, 1929).

([7]) L. E. Dickson, History oi the Theory of Numbers (Carnegie Institute, Washington; vol. I, 1919; vol. II, 1920; vol. III, 1923).

([8]) L. E. Dickson, Modern Elementary Theory of Numbers (Chicago University Press, 1939).

([9]) G. H. Hardy and E. M. Wright, Introduction to the Theory of Numbers (Clarendon Press, Oxford, 4th ed., 1960).

([10]) W. J. LeVeque, Topics in Number Theory (2 vols. Addison—Wesley, Reading, Mass., 1956).

([11]) G. B. Mathews, Theory of Numbers (Deighton Bell, Cambridge, 1892; Part I only published).

([12]) T. Nagell, Introduction to Number Theory (John Wiley, New York, 1951).

([13]) O. Ore, Number Theory and its History (McGraw-Hill, New York, 1948).

([14]) J. V. Uspensky and M. A. Heaslet, Elementary Number Theory (McGraw-Hill, New York, 1939).

*) Список книг на русском языке добавлен мною. (*Прим. перев.*)

На французском языке

([15]) E. Cahen, Theorie des nombres (2 vols., Hermann, Paris, 1924).

На немецком языке

([16]) P. Bachmann, Niedere Zahlentheorie (Teubner, Leipzig; vol. I, 1902; vol. II, 1910).

([17]) E. Bessel-Hagen, Zahlentheorie (Pascals Repertorium, vol. I, part 3; Teubner, Leipzig, 1929).

([18]) E. Landau, Vorlesungen über Zahlentheorie (3 vols., Hirzel, Leipzig, 1927; reprinted by Chelsea, New York).

([19]) Scholz, Einführung in die Zahlentheorie (Sammlung Göschen, №1131, de Gruyter, Berlin, 1939).

([20]) P. G. Lejeune Dirichlet, Vorlesungen über Zahlentheorie, herausgegeben von R. Dedekind (Vieweg, Braunschweig, 4th ed., 1894) *).

([21]) H. Hasse, Vorlesungen über Zahlentheorie (Springer, Berlin, 1950) **).

*) Имеется русский перевод, см. ([3]). (*Прим. перев.*)

**) Имеется русский перевод, см. ([4]). (*Прим. перев.*)

УКАЗАТЕЛЬ

Г. Дэвенпорт

Высшая арифметика.
Введение в теорию чисел.

М., 1965 г., 176 стр. с илл.

Редактор *И. Е. Морозова*

Техн. редактор *Л. А. Пыжова*

Корректор *Н. В. Гераськина*

Сдано в набор 12/V 1965 г. Подписано к печати 12/VIII 1965 г. Бумага $84 \times 108^{1}/_{32}$. Физ. печ. л. 5,5. Условн. печ. л. 9,02. Уч.-изд. л. 8,8. Тираж 30 000 экз. Цена книги 44 коп. Заказ № 1719.

Издательство «Наука».

Главная редакция
физико математической литературы.

Москва, В-71, Ленинский проспект, 15.

Ленинградская типография № 1 «Печатный Двор» имени А. М. Горького Главполиграфпрома Государственного комитета Совета Министров СССР по печати, Гатчинская, 26.

Отпечатано с матриц в Гостипографии «Вайздас», г. Вильнюс, Страздялио, 1. Заказ № 3924.