Теория чисел во Второй школе

П. В. Бибиков, К. В. Козеренко, А. И. Малахов

Оглавление

Предисловие

Во второй половине XIX века произошло событие, сопоставимое по своему историческому, культурному и цивилизационному значению с появлением в Древней Греции (Фалес, VI в. до н.э.) дедуктивной математической системы. Математики начали понимать, что «должно быть позволено рассуждать об объектах, не имеющих никакой наглядной или чувственной интерпретации». В 1870 году получила признание геометрия Лобачевского. Это произошло после того, как Феликс Клейн обнаружил в одной работе Артура Кэли «модель», позволяющую отождествить объекты и соотношения геометрии Лобачевского с некоторыми объектами и соотношениями евклидовой геометрии. Этим он доказал, что геометрия Лобачевского непротиворечива в той же мере, что и евклидова, — противоречие в одной из них необходимо влечет противоречие в другой. В 1872 году почти одновременно Кантор, Дедекинд и Вейерштрасс дали определние (правда, довольно различными методами) вещественного числа, затем Вейерштрасс определил отрицательные числа в виде классов пар натуральных чисел, и, наконец, в 1888 году Дедекинд сформулировал полную систему аксиом для арифметики (аксиомы Пеано). В геометрии похожие процессы завершились выходом в 1899 году книги Гильберта «Основания геометрии», где он объяснил, что прямая, точка и плоскость появляются только в связи с теми аксиомами, которые для них выбираются. Другими словами, назвать ли их точками, прямыми, плоскостями или же столами, стульями, пивными кружками, — это будут те объекты, для которых справедливы соотношения, выражаемые аксиомами. *Родился новый формальный язык! И новая интуиция!* Теперь «интуиция отнюдь не обязательно имеет пространственную или чувственную природу, как часто думают, а скорее представляет собой некоторое *знание поведения матема-*

тических объектов, часто прибегающее к помощи образов самой различной природы, но основанное прежде всего на повседневном знакомстве с этими объектами» (Бурбаки).

В математическом образовании вследствие такого развития науки появились разрывы. Точка разрыва — это неверно сформированная интуиция или, иными словами, это разрыв в математической культуре.

Точки разрыва в математическом образовании сильно снижают его уровень и качество. С некоторыми разрывами справиться легко, просто указав на них. Есть разрывы более серьезные. Открытие геометрии Лобачевского оказало огромное влияние на развитие математики. Такую же роль эта геометрия должна играть и в образовании. Но представление о том, что через точку, не лежащую на прямой, можно провести более одной прямой, параллельной данной, не стало общедоступным. Это даже, скорее, точка отрыва школьной математики от общекультурных достижений.

Наконец, имеются и узаконенные, но устранимые, как мы полагаем, разрывы. Школьная математика уже в средних классах становится, в основном, конкурсно-олимпиадной. Мы учим детей решать вычурные, никому не нужные и при этом очень сложные задачи, которые зачастую не хочется не то что решать, даже вникать в условия! Причем в подборе этих задач, как правило, нет системы. Ни о каком формировании интуиции в этой ситуации говорить просто не приходится. Какое представление о математике после этого складывается у наших учеников, одному Богу известно.

Что же делать? Перейдем теперь к конкретным предложениям.

Теория чисел в школе

Теория чисел, на наш взгляд, доставляет замечательный пример теории, которая может способствовать устранению разрывов в образовании по следующим причинам.

Во-первых, следуя Пуанкаре и Арнольду, мы полагаем, что математика является частью теоретической физики, то есть экспериментальной наукой. Слово «математика» означает «точное знание», и соответствующие открытия были получены из наблюдений явлений

природы. Решая огромное количество задач, школьники учатся не наблюдать явления, а отвечать на *уже поставленные* вопросы, и на то, что математика — искусство их задавать, не обращают внимание. В теории чисел эксперимент играет огромную роль и позволяет самостоятельно пройти путь от наблюдения и формулирования гипотезы до доказательства теоремы. Неоценимый опыт!

Во-вторых, школьники не подозревают, что они не знают, что такое натуральное число, $0, -1$, многочлен. Может быть, строгих определений здесь давать и не надо (хотя наш опыт показывает, что такие разговоры идут «на ура», и мы приводим в тексте соответствующие определения), но отметить их отсутствие необходимо. Иначе сложится ситуация, когда привыкание заменяет понимание, что, конечно, есть точка разрыва. Но в школьной математике есть точки разрыва и посерьезнее.

Формальный язык обладает одной важной особенностью — фиксацией аналогий. Похожие структуры выделяются термином и определенным набором аксиом, которые отражают те или иные свойства этих структур. Так, например, множества целых чисел $\mathbb{Z}$, многочленов $\mathbb{R}[x]$ и гауссовых чисел $\mathbb{Z}[\sqrt{-1}]$ похожи. Эта похожесть фиксируется определением: все эти множества называются *кольцами*. Такая особенность формального языка позволяет изучать одновременно целые классы структур и постоянно используется в современной математике. Когда вчерашний школьник приходит на первую лекцию и слышит определение кольца (группы, поля, ...), то воспринимает эти объекты по выражению Арнольда, как «множества с операциями, удовлетворяющими длинному ряду труднозапоминаемых аксиом ...». Нет смысла комментировать, сколь пагубно это сказывается на его образовании. Новоиспеченный студент не подготовлен к восприятию такого языка потому, что «школьная» математика конкретна, а к формальному языку нужно привыкнуть, уметь его узнавать. На это требуется время. Мы надеемся, что наш курс теории чисел позволяет достичь этой цели, постепенно, начиная с конкретных примеров, давая почувствовать ученикам аналогию между кольцами $\mathbb{Z}$, $\mathbb{R}[x]$ и $\mathbb{Z}[\sqrt{-1}]$ и подготавливая их (как будущих студентов) к восприятию формального языка.

В-третьих, теория чисел, безусловно, предоставляет возможность научить решать задачи. По всей видимости, именно последнее соображение является для многих смыслом и целью изучения теории чисел. Научить решать задачи! И все! Как правило, для сборников задач характерно полное отсутствие сюжетов, задачи в них появляются так, как будто они вылетели из «датчика случайных чисел». А то, что среди них есть такие, которые высвечивают целый спектр ключевых идей, остается за кадром! Одним из таких примеров является задача описания пифагоровых троек — она не только мотивирует идею рассмотрения алгебраических кривых, но и возможность или невозможность их рациональной параметризации, приводит к понятию рода римановой поверхности, который в свою очередь отвечает за топологию вещественных и комплексных кривых, элементарность абелевых интегралов, ...

О таких вещах необходимо говорить со школьниками! Необходимо подчеркивать, что математика — это прожектор, который высвечивает значительную часть картины мира, а не представляет собой набор несвязанных между собой методов, или что это — игра вроде шахмат.

Ну, и, наконец, что касается логических пробелов при обучении, то просто надо стремиться к тому, чтобы их было поменьше. Сама по себе эта деятельность нам кажется очень полезной. Впрочем, как и во всем, и здесь уместно «чувство меры и сообразности». Степень погружения зависит от уровня класса. Однако обязательно надо придерживаться принципа Н.Н.Константинова о «честном умолчании»: если учитель в каком-то месте, либо пропустил доказательство (умолчал), понимая, что ученики воспримут этот факт как нечто естественное и не вызывающее возражений, либо просто сослался на очевидность, то добавить строгое доказательство можно, *не разрушая структуры курса.*

Мотивировки

Математическое образование должно строго придерживаться принципа естественности (так сказать, «снежного кома»), т.е. необходимо давать только тот материал, который ляжет на уже усвоенный и который будет мотивирован. Перед тем как вводить новое понятие или

начинать новый курс, нужно обязательно объяснять (хотя бы «на пальцах»), ради чего это делается. Отсутствие мотивировок вводимых понятий и направления исследования недопустимо и является не просто разрывом, а делает такое «образование» бессмысленным. Спросите любого школьника, зачем нужны логарифмы или тригонометрические уравнения, и вы поймете то, что мы имеем в виду. И, конечно, объяснения должны соответствовать уровню математической культуры слушателя. Л. Выготский, имея в виду только что сказанное, говорил про зоны ближайшего развития. Мы старались придерживаться этого правила и вводить такие классические понятия и теоремы, как китайскую теорему об остатках, функцию Эйлера, теорему Вильсона и прочие, мотивируя это необходимостью исследовать определенные явления, как, например, устройство колец остатков, периодичность геометрических прогрессий по модулю, разрешимость квадратных сравнений и другие.

Ничья земля

Образование похоже на освоение земель. Каждый чертит свою карту новой для него земли. Откуда берутся задачи? Сколько и каких задач надо решить, чтобы считать, что курс освоен? По всей видимости, наши ученики даже не думают об этом, хотя, казалось бы, здесь, как в туристическом походе, они должны постоянно спрашивать, а долго ли осталось идти. И только тогда, когда основные реки, горы и равнины нанесены на карту (то есть, когда не осталось больших пробелов — проблем), можно начинать осваивать другую территорию.

Ничья земля — это разделы математики вроде аффинной геометрии, проективной геометрии, геометрии Лобачевского, теории нефакториальных колец и т.д. На изучение этих разделов не хватает времени ни в школе, ни в вузе, но без них нельзя представить себе полноценного математического образования. Разрыв здесь состоит в том, что в школе формируют такие навыки, которые очень затрудняют их освоение. Такие разрывы мы называем блокировками.

Хороший способ устранения этих точек разрыва — исследовательские работы школьников. В Добавлениях мы приводим работы, выполненные нашими учениками Данилой Байгушевым и Григорием

Юргиным. Некоторые из них были опубликованы.

- Байгушев Д.А. *Об асимптотике эргодических перестановок Арнольда.* Мат. Просвещение. Сер. 3. Вып. 16 (2012), 89-93
- Baygushev D. *On Geometry of Young Diagrams for Arnold Permutations.* Lobachevskii Journal of Mathematics, 2012, Vol. 33, No. 2, pp. 109-114.
- Байгушев Д.А. *О матричной функции Эйлера.* Труды математического центра им. Н.И. Лобачевского т. 45, с. 12-14, 2012.

В заключение, может быть, самое главное.

«Математика - это язык»

Язык науки, особенно язык математики, принципиальнейшим образом отличается от естественного, так сказать, бытового языка. Для естественного языка характерно наличие неоднозначности, т.е. наличие взаимоисключающих смыслов. В процессе восприятия естественной речи человек пользуется различными инструментами разрешения неоднозначности, такими как аналогии, апелляции к наглядным образам, но, прежде всего, контекстом. Поэтому естественно-языковые тексты информационно избыточны. Язык математики использует особый инструмент разрешения неоднозначности. В математическом тексте каждый (!) термин или понятие должны быть определены, что абсолютно исключает возможность неоднозначного их понимания.

На то, что математика — это совершенно особый язык, очень редко кто в школе обращает внимание, и, как правило, этому вообще не учат, хотя, может быть, математика включена в школьную программу именно для того, чтобы хотя бы с этим языком познакомиться, или, что значительно лучше, научиться хоть немного на нем разговаривать. Преподаватели же вузов считают, что студенты, конечно, им владеют свободно, и сразу начинают говорить с ними на незнакомом формальном языке, а бедные студенты не понимают, что происходит. В результате приходится просто зазубривать непонятные определения и доказательства теорем, что приводит к потере смысла образования.

Почти сто лет тому назад Эмиль Борель говорил, «что по существу образование ума при помощи точных знаний гораздо важнее, чем приобретение этих знаний и, что преподавание математики может получить полную воспитательную ценность лишь при условии, если оно будет избегать слишком распространенного *софизма*, будто реальные трудности можно разрешить с помощью простых словесных определений». Здесь мы хотим быть правильно понятыми и не перепугать тех, кто призывает учителей «быть реалистами и не пытаться научить строгим определениям и понятиям с самого начала». Конечно, о таких вещах, на наш взгляд, надо сначала просто рассказывать на уроках и не только не требовать, но и не ожидать немедленного понимания. Семя брошено в почву, надо подождать пока оно прорастет.

Как отметил В.А. Успенский, способность отличать осмысленное от бессмысленного и истинное от ложного следует неуклонно и ненавойливо прививать уже с начальных классов школы. И не является ли это главным в школьном преподавании?

П. В. Бибиков, К. В. Козеренко, А. И. Малахов

Глава I

Простые числа

Натуральные числа — видимо, первый математический объект, который был принят человеком на самой заре нашей цивилизации многие тысячи лет тому назад с вполне понятной целью пересчета предметов. Натуральные числа можно складывать, перемножать и, казалось бы, что нет ничего проще. Однако, не будем торопиться с выводами. В математике часто бывает так, что объекты, которые появляются из вполне «приземленных» соображений, таят в себе неожиданную глубину.

1. Простые числа и неприводимые многочлены

Числа называют *составными*, если их можно разложить на два меньших сомножителя. Например, число $6 = 2 \cdot 3$ является составным. А вот число 7 нельзя разложить подобным образом. Поэтому число 7 называют *простым* числом. Итак, дадим

Определение. *Составным числом* называется натуральное число, которое может быть разложено в произведение двух натуральных чисел, отличных от него самого.

Простым числом называется натуральное число, большее 1 и не являющееся составным.

Замечание. Часто в качестве определения простых чисел дается следующее их свойство: у простого числа есть *ровно два* делителя — 1 и оно само. В дальнейшем (рассматривая, например, примеры Яглома и Гильберта) мы увидим, насколько важно давать верные определения.

Замечание. Как следует из определения, 1 не является ни простым, ни составным числом. Почему — мы подробно обсудим в III.2.

Давайте выпишем все простые числа, меньшие 100:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43,$$
$$47, 53, 59, 61, 67, 71, 73, 79, 83, 89, 97, \ldots$$

Возникает естественный вопрос: как искать простые числа? Опишем старинный способ, придуманный еще в III в. до н. э. Эратосфеном Киренским, хранителем Александрийской библиотеки.

Выпишем натуральные числа, начиная с 2. Двойку обведем, а остальные числа, которые делятся на 2, зачеркнем. Ближайшим незачеркнутым числом будет 3. Обведем и его, а все остальные числа, кратные 3, зачеркнем. Следующее наименьшее незачеркнутое число — это 5. Обводим пятерку, а остальные числа, кратные 5, зачеркиваем. Повторяя эту процедуру снова и снова, мы в конце концов добьемся того, что незачеркнутыми останутся лишь простые числа — они словно просеялись сквозь решето.

Замечание. Такой способ не позволяет достоточно быстро находить большие простые числа, и тем более определять, является ли данное число простым или нет, поскольку нужно выписать все числа до него.

Казалось бы, какие сложности можем мы испытать при изучении простых чисел? Еще какие! Чтобы продемонстрировать, о чем идет речь, рассмотрим, например, среди простых чисел пары таких, разность между которыми минимальна. Как легко видеть, эта разность, кроме одного исключительного случая 2 и 3, равна 2:

$$2, \mathbf{3}, \mathbf{5}, \mathbf{7}, \mathbf{11}, \mathbf{13}, \mathbf{17}, \mathbf{19}, 23, \mathbf{29}, \mathbf{31}, 37, \mathbf{41}, \mathbf{43}, 47, 53, \mathbf{59}, \mathbf{61},$$
$$67, \mathbf{71}, \mathbf{73}, 79, 83, 89, 97, \mathbf{101}, \mathbf{103}, \mathbf{107}, \mathbf{109}, 113, 127, 131 \ldots$$

Такие пары простых чисел называются близнецами. Как устроено множество чисел-близнецов? Можно ли хотя бы утверждать, что это множество бесконечно? До сих пор ответы на эти вопросы неизвестны.

Прежде чем двигаться дальше, отметим одно ключевое соображение для всего нашего курса. В математике первостепенную роль играет *установление аналогий* между, казалось бы, совершенно различными структурами.

Рассмотрим, на первый взгляд, неожиданный в данном контексте пример: множество многочленов от одной переменной. Странно было бы ожидать, что множество многочленов имеет нечто общее с множеством натуральных чисел. Однако, заметим, что многочлены можно складывать и перемножать и даже раскладывать на множители. Не правда ли, очень похоже на натуральные числа? Возникает естественный вопрос: *существуют ли в множестве многочленов аналоги простых чисел?* Да, существуют! Они называются *неприводимыми многочленами.*

Определение. *Приводимым многочленом* называется многочлен, который может быть разложен в произведение двух многочленов меньшей положительной степени.

Неприводимым многочленом называется многочлен положительной степени, не являющийся приводимым.

Замечание. Обратите внимание, что в этом определении существенную роль играет степень многочлена.

Замечание. Как следует из определения, многочлены нулевой степени (т.е. ненулевые числа) не является ни приводимыми, ни неприводимыми многочленами. Причины этому те же, по которым 1 не является ни простым, ни составным числом (вновь аналогия!).

Замечание. В большинстве школьных учебников и пособий многочленом называют «выражение вида...». Вообще говоря, эта фраза определением не является. Здесь просто одно слово — «многочлен», заменено другим — «выражение». А что такое выражение?... В главе VII мы дадим строгое определние многочлена.

Рассмотрим несколько примеров. Начнем с приводимых многочле-

нов. Таковыми являются, например, следующие.

$$x^2 = x \cdot x,$$
$$x^2 - 1 = (x-1) \cdot (x+1),$$
$$x^2 - \frac{1}{4} = \left(x - \frac{1}{2}\right) \cdot \left(x + \frac{1}{2}\right),$$
$$x^3 - x = x \cdot (x^2 - 1) = x \cdot (x-1) \cdot (x+1),$$
$$x^3 + x = x \cdot (x^2 + 1).$$

Приведем несколько примеров неприводимых многочленов.

$$x,$$
$$x + c, \quad \text{где } c \text{ — произвольное число},$$
$$x^2 + 1,$$
$$x^2 + x + 1.$$

Необходимо заметить, что вопрос о неприводимости того или иного многочлена зависит от множества, которому принадлежат его коэффициенты. Если рассматривать многочлены, коэффициенты которых являются произвольными действительными числами (это множество обозначается через $\mathbb{R}[x]$), то, как оказывается, существует критерий, который позволяет явно описать все неприводимые многочлены! Ими являются

1. Все многочлены первой степени;
2. Многочлены второй степени с отрицательным дискриминантом. А именно, рассмотрим многочлен

$$ax^2 + bx + c.$$

Используя процедуру выделения полного квадрата (смотрите, как эта процедура заработала!), получаем

$$ax^2 + bx + c = a\left(x^2 + \frac{b}{a}x\right) + c = a\left(x^2 + 2\frac{b}{2a}x + \frac{b^2}{4a^2} - \frac{b^2}{4a^2}\right) + c =$$
$$a\left(x^2 + 2\frac{b}{2a}x + \frac{b^2}{4a^2}\right) - \frac{b^2}{4a} + \frac{ac}{a} = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a}.$$

Выражение $D = b^2 - 4ac$ играет ключевую роль при рассмотрении многочленов второй степени и называется *дискриминантом* многочлена $ax^2 + bx + c$. Рассмотрим три случая.

- $D < 0$. В таком случае, как видно из полученного нами представления, многочлен не имеет корней и всегда принимает значения одного знака. Положительные, если $a > 0$, и отрицательные, если $a < 0$. Именно в этом случае многочлен $ax^2 + bx + c$ является неприводимым. Приведем соответствующие примеры

$$\begin{aligned} x^2 + 1 &> 0, \\ x^2 + x + 1 = \left(x + \frac{1}{2}\right)^2 + \frac{3}{4} &> 0, \\ -x^2 - 1 &< 0, \\ -2x^2 + 3x - 2 = -2\left(x - \frac{3}{4}\right)^2 - \frac{7}{8} &< 0. \end{aligned}$$

- $D = 0$. В таком случае $ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2$, откуда очевидно следует приводимость исходного многочлена.
- $D > 0$. В этом случае оказывается, что у исходного многочлена обязательно будут корни x_1, x_2 и разложение будет иметь вид

$$ax^2 + bx + c = a\left(x + \frac{b}{2a}\right)^2 - \frac{b^2 - 4ac}{4a} = a(x - x_1) \cdot (x - x_2).$$

Например,

$$\begin{aligned} x^2 - 1 &= (x - 1) \cdot (x + 1), \\ x^2 - x - 2 &= (x - 2) \cdot (x + 1), \\ 3x^2 - 5x - 2 &= (3x + 1) \cdot (x - 2) = 3\left(x + \frac{1}{3}\right) \cdot (x - 2), \\ -2x^2 + 7x - 3 &= -(2x - 1) \cdot (x - 3) = -2\left(x + \frac{1}{2}\right) \cdot (x - 3). \end{aligned}$$

Однако, могут возникнуть следующие сложности. Многочлен $x^2 - 2$ является приводимым, поскольку $D = 0^2 - 4 \cdot (-2) = 8 > 0$. Значит, имеется его разложение

$$x^2 - 2 = (x - ?) \cdot (x - ?).$$

Что будет стоять на месте вопросительных знаков? Пока мы лишь можем сказать, что это будут такие числа, квадрат которых равен 2 (иными словами, корни уравнения $x^2 - 2 = 0$). Окончательный ответ на этот вопрос таит в себе неожиданные сложности, ибо *не существует такого рационального числа* (дроби), квадрат которого равен 2. В главе XI мы сможем прояснить ситуацию и ответить на поставленный вопрос.

В множестве $\mathbb{R}[x]$ все многочлены третьей степени и выше приводимы, т.е. их можно разложить на множители! Например,

$$x^4 + 4 = (x^2 - 2x + 2) \cdot (x^2 + 2x + 2).$$

Замечание. Множество $\mathbb{R}[x]$ *похоже* на множество $\mathbb{N}$. Но, как мы ранее сказали, не существует эффективного алгоритма, который позволяет определить, является ли данное натуральное число простым. Даже с использованием самых современных компьютеров это потребует значительного времени. Однако, в случае многочленов с действительными коэффициентами вопрос решается много проще.

Если рассматривать многочлены, коэффициенты которых являются произвольными рациональными числами (это множество обозначается через $\mathbb{Q}[x]$), то ситуация становится намного сложнее. Не существует явного описания неприводимых многочленов в $\mathbb{Q}[x]$. Приведем несколько примеров многочленов, которые приводимы, как

элементы $\mathbb{R}[x]$, но неприводимы, как элементы $\mathbb{Q}[x]$.

$$x^2 - 2,$$
$$x^4 + 1,$$
$$x^4 + x^3 + x^2 + x + 1,$$
$$x^5 - 3,$$
$$x^6 + x^5 + x^4 + x^3 + x^2 + x + 1.$$

Замечание. В множестве $\mathbb{Q}[x]$ существуют неприводимые многочлены любой степени. Например,

$$x^n - p,$$

где p — простое число.

При повседневном знакомстве с теми или иным объектами, например, с натуральными числами, вы вырабатываете соответствующую интуицию, приобретаете знание поведения этих объектов. Устанавливаемые аналогии наподобие той, что мы отметили между числами и многочленами, позволяют *перебрасывать интуицию* на новые структуры.

Вернемся к рассмотрению простых чисел и докажем следующее утверждение, которое понадобится нам в дальнейшем.

Лемма (О простом делителе)**.** *У любого натурального числа, большего 1, существует простой делитель.*

Доказательство. Пусть n — произвольное натуральное число, большее 1. Рассмотрим *наименьший* натуральный делитель p числа n, больший 1.

Почему такой делитель p существует?

Действительно, предположим, что наименьшего делителя не существует. В таком случае для любого делителя q найдется делитель q' такой, что $q > q'$. Если наименьшего делителя нет, то эту цепочку можно продолжать $q > q' > q'' > \ldots$. Но она обязательно оборвется, поскольку множество натуральных чисел $\mathbb{N}$ *ограничено снизу.*

Докажем, что число p простое. Предположим противное: пусть число p составное. Тогда по определение составного числа найдутся два натуральных числа a и b, отличные от p, такие, что $p = ab$. Но тогда $a < p$ и n делится на a. В самом деле, если $n = p \cdot n_1$, то, подставляя в это равенство $p = ab$, получаем $n = a \cdot (bn_1)$. Значит, n делится на a и p — не наименьший делитель n. Противоречие. □

Замечание. Обратите внимание, что утверждения о простых числах отнюдь не просты, если доказывать их в общем виде. Например, утверждение о возможности разложения натурального числа на простые множители единственным образом является на самом деле теоремой! Теоремой, которая требует доказательства (совсем не простого), которое мы дадими в главе IX.

Замечание. После того как мы установили аналогию между числами и многочленами, имеет смысл спросить, а существует ли для многочленов аналог утверждения, которое мы только что доказали для чисел. Как правило, такой аналог существует. В качестве примера приведем следующее утверждение.

Лемма (О неприводимом делителе). *У любого многочлена положительной степени существует неприводимый делитель.*

Введем для удобства общеприятное обозначение для степени многочлена: deg.

Доказательство. Пусть f — произвольный многочлен положительной степени. Рассмотрим многочлен p *наименьшей* положительной степени, который делит f.
Почему такой многочлен p существует?
Действительно, предположим, что такого многочлена не существует. В таком случае для любого делителя q найдется делитель q' такой, что $\deg q > \deg q'$. Если нет делителя нименьшей степени, то эту цепочку можно продолжать $\deg q > \deg q' > \deg q'' > \ldots$. Но она обязательно оборвется, поскольку степень многочлена принимает лишь

неотрицательные значения и поэтому множество значений степени *ограничено снизу*.

Докажем, что многочлен p неприводим. Предположим противное: пусть p — приводимый многочлен. Тогда по определение приводимого многочлена найдутся многочлены g и h положительной степени, такие, что $p = gh$. Но тогда $\deg g < \deg p$ и f делится на g. В самом деле, если $f = p \cdot f_1$, то, подставляя в это равенство $p = gh$, получаем $f = g \cdot (hf_1)$. Значит, f делится на g и p — делитель f не наименьшей степени — противоречие. □

2. Распределение простых чисел. Функция π.

Посмотрим на ряд из нескольких первых простых чисел:

$$2, 3, 5, 7, 11, 13, 17, 19, 23, 29, 31, 37, 41, 43, 47, 53, 59, 61, 67, \ldots$$

Видно, что простых чисел довольно много, и встречаются они достаточно часто, хотя и «хаотично». Чтобы продемонстрировать это явление, рассмотрим функцию распределения простых чисел. Для натурального числа n ее значение $\pi(n)$ определяется как количество таких простых чисел p, что $p \leqslant n$. Эта функция несет в себе достаточно полную информацию о ряде простых чисел. Например, количество простых чисел, принадлежащих отрезку $[n, m]$, равно разности $\pi(m) - \pi(n-1)$.

На рисунке ниже приведен график этой функции для натуральных чисел от 1 до 60.

Обращают на себя внимание отрезки постоянства функции π. Как легко сообразить, это такие отрезки в натуральном ряду, которые не содержат *ни одного простого числа*. К их изучению мы вернемся чуть позже.

Кроме того, функция π несет в себе информацию о плотности, с которой простые числа встречаются в натуральном ряду.

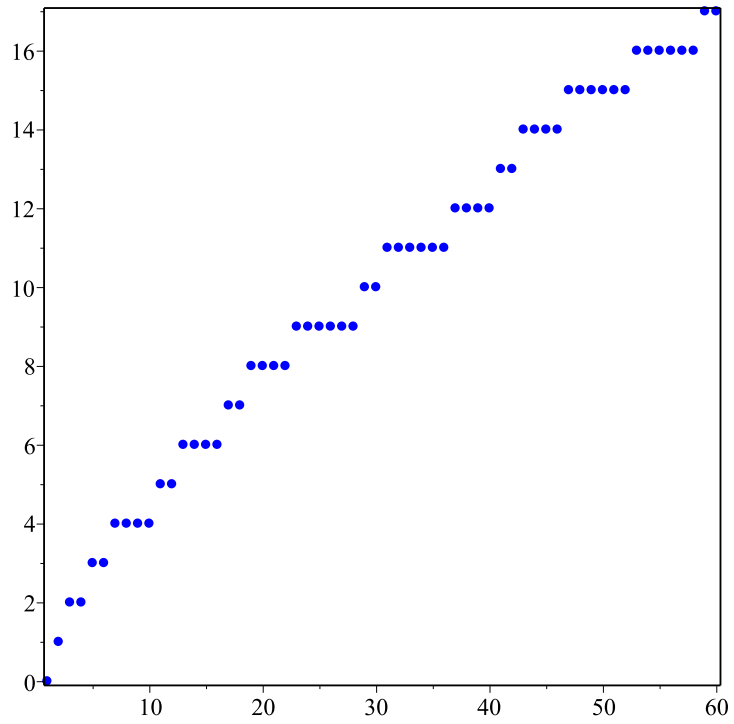


График $\pi(n)$ на отрезке от 1 до 60.

n	10	10^2	10^3	10^4	10^5	10^6	10^7
$\pi(n)$	4	25	168	1229	9592	78498	664579
$\pi(n)/n$	0,4	0,25	0,168	0,123	0,096	0,078	0,066

Видно, что эта плотность уменьшается с ростом n, т.е. простых чисел все же не так много. Теперь рассмотрим график функции распределения на больших масштабах. Для первой тысячи натуральных чисел он приведен ниже.

Что обращает на себя внимание на этом графике? Несмотря на отсутствие какой-либо явной закономерности в ряду простых чисел, функция π на больших масштабах растет довольно регулярно. Еще в конце XVIII века Лежандр и Гаусс независимо сформулировали гипотезу о возможности приблизить $\pi(n)$ известной функцией (эта функция есть $\frac{n}{\ln n}$). Усилиями многих математиков спустя сто лет в 1896 году эта гипотеза была доказана.

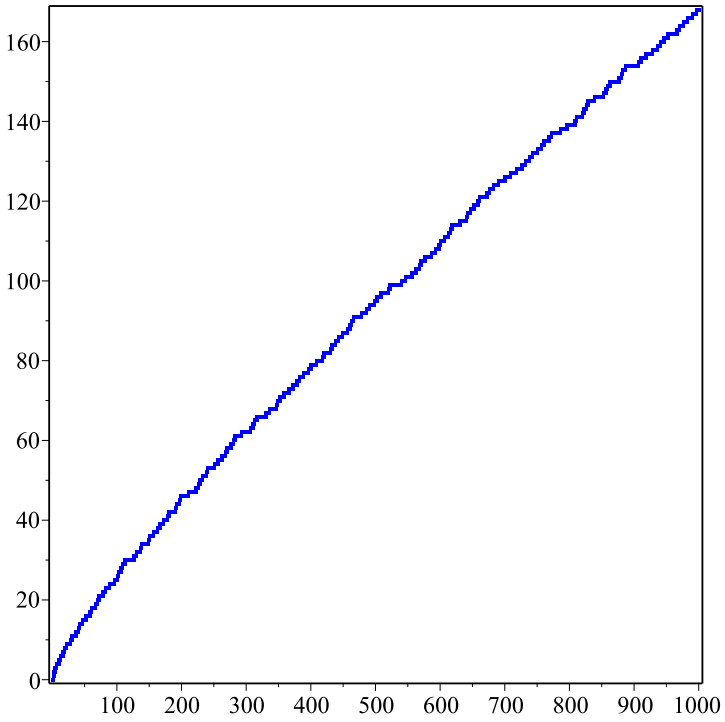
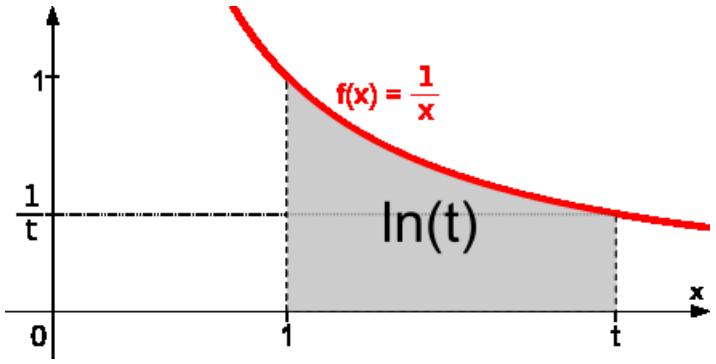


График $\pi(n)$ на отрезке от 1 до 1000.

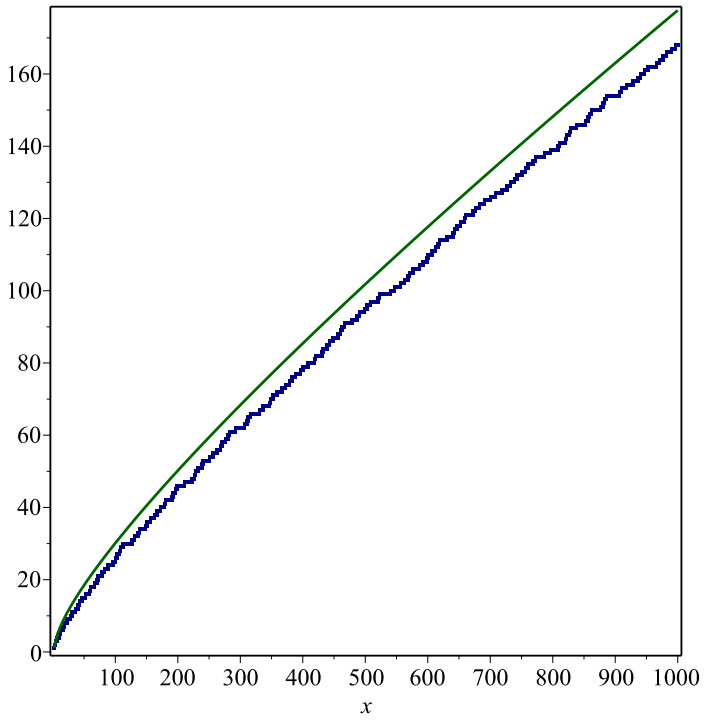
Замечание. Что такое $\ln n$, наверняка спрашиваете вы. Ответ совершенно потрясающий! Рассмотрим график функции $f(x) = \frac{1}{x}$, который называется гиперболой. Тогда $\ln t$ — это величина заштрихованной площади(!) под этим графиком на отрезке $[1, t]$.



Определение функции $\ln(t)$.

Каким образом в задаче из теории чисел появляется такая функция? Математика воистину удивительная наука!

Ниже приведен график функции π и ее приближения.



Функция π и ее приближение.

n	10	10^2	10^3	10^4	10^5	10^6	10^7
$\pi(n)$	4	25	168	1229	9592	78498	664579
$\frac{\pi(n)}{n/\ln n}$	0,921	1,151	1,161	1,132	1,104	1,085	1,071

Из приведенной таблицы можно видеть, как отношение $\frac{\pi(n)}{n/\ln n}$ стремится к 1 с ростом n. Формально это записывается так:

$$\pi(n) \sim \frac{n}{\ln n} \text{ при } n \to \infty$$

До сих пор в распределении простых чисел в натуральном ряду есть множество открытых вопросов, не решенных современными математиками, несмотря на значительные усилия. Сейчас мы попробуем поисследовать ряд простых чисел и доказать некоторые закономерности в нем.

3. Теорема Евклида

Первый вопрос, который естественно возникает при рассмотрении ряда простых чисел, следующий. Мы выписали несколько простых

чисел. Современные компьютеры могут выписать простые числа, состоящие из многих тысяч знаков. Однако можно ли утверждать, что всегда существуют простые числа, помимо тех, что уже найдены? Иными словами, верно ли, что *простых чисел бесконечно много*?

Ответ на этот вопрос дает следующая

Теорема (Евклид)**.** *Простых чисел бесконечно много.*

Доказательство. Предположим *противное*, а именно что простых чисел лишь *конечное* число и P — наибольшее из них. Тогда число $N = 2 \cdot 3 \cdot 5 \cdot \ldots \cdot P + 1$ не простое. Рассмотрим p — его простой делитель, который существует по лемме о простом делителе. Среди набора $2, 3, \ldots P$ его нет, поскольку число N не делится ни на одно из чисел данного набора. Мы пришли к противоречию. Теорема доказана. □

Рассмотрим несколько чисел вида

$$e_k = p_1 p_2 \ldots p_k + 1.$$

Здесь $p_1 = 2 < p_2 < p_3 < \ldots < p_k$ - простые числа, взятые подряд в порядке возрастания. Числа e_k называют *числами Евклида.*

$$\begin{aligned} e_1 &= 2 + 1 = 3 \\ e_2 &= 2 \cdot 3 + 1 = 7 \\ e_3 &= 2 \cdot 3 \cdot 5 + 1 = 31 \\ e_4 &= 2 \cdot 3 \cdot 5 \cdot 7 + 1 = 211 \\ e_5 &= 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 + 1 = 2311 \end{aligned}$$

Заметим, что пока числа получаются простыми. Однако,

$$e_6 = 2 \cdot 3 \cdot 5 \cdot 7 \cdot 11 \cdot 13 + 1 = 30031 = 59 \cdot 509.$$

Замечание. Арабские цифры, к которым вы так привыкли и которыми пользуетесь, не задумываясь, возникли в Индии через более чем 500 лет после смерти Евклида. Представьте, каким трудным делом было раскладывать числа на простые множители без столь удобного инструмента!

Данная теорема доставляет первый для нас пример *теоремы существования*. Приведенное доказательство не дает способа находить новые простые числа. Однако небольшая модификация доказательства Евклида дает *алгоритм*, который позволяет строить новые простые числа из уже имеющихся.

Рассмотрим простое число $q_1 = 2$. Давайте попробуем найти какое-нибудь число, которое не делится на 2. Для этого проще всего прибавить к числу $q_1 = 2$ единицу. Мы получим простое число $q_1 + 1 = 3$. Обозначим его через q_2.

Теперь давайте попробуем найти число, не делящееся ни на $q_1 = 2$, ни на $q_2 = 3$. Для этого рассмотрим число

$$q_1 \cdot q_2 + 1 = 2 \cdot 3 + 1 = 7.$$

Оно также оказалось простым. Обозначим его через q_3.

Будем продолжать этот процесс. Возьмем уже найденные нами простые числа $q_1 = 2$, $q_2 = 3$ и $q_3 = 7$ (обратите внимание, что пока что мы не получили простого числа 5) и возьмем число

$$q_1 q_2 q_3 + 1 = 2 \cdot 3 \cdot 7 + 1 = 43.$$

Как мы помним, это число также просто. Обозначим его через q_4.

Возникает гипотеза: *чтобы найти новое простое число, нужно перемножить все найденные до этого простые числа и прибавить 1.*

Давайте это проверим.

Еще раз запишем наш процесс с самого начала:

2	$2 + 1 = 3$ — простое
2, 3	$2 \cdot 3 + 1 = 7$ — простое
2, 3, 7	$2 \cdot 3 \cdot 7 + 1 = 43$ — простое
2, 3, 7, 43	$2 \cdot 3 \cdot 7 \cdot 43 + 1 = 1807$ —???

Является ли число 1807 простым? Оказывается, что $1807 = 13 \cdot 139$! Значит, число 1807 составное, и наша гипотеза оказалась неверной.

Что же делать?

А вот что. Рассмотрим минимальный натуральный делитель числа 1807, отличный от 1. Это число равно 13, и его нет среди уже найденных нами простых чисел q_1, q_2, q_3 и q_4. Обозначим его через q_5.

Теперь мы готовы описать наш алгоритм нахождения новых простых чисел в общем виде.

Рассмотрим уже найденные простые числа $q_1, q_2, \ldots, q_n$. Построим число

$$Q_n = q_1 q_2 \ldots q_n + 1$$

и рассмотрим наименьший натуральный делитель q_{n+1} числа Q_n, отличный от 1. Тогда q_{n+1} — новое простое число.

Заметим, что таким образом мы не сможем получить все простые числа. В частности, мы не получим простого числа 5.

4. Простые числа в арифметических прогрессиях

Рассмотрим последовательности вида $3k$, $3k+1$ и $3k+2$, где k пробегает все целые неотрицательные числа. (Такие последовательности, имеющие вид $a, a+d, a+2d, a+3d, \ldots$, называются арифметическими прогрессиями.)

$$0, \mathbf{3}, 6, 9, 12, 15, 18, 21, 24, 27, 30, 33, \ldots$$
$$1, 4, \mathbf{7}, 10, \mathbf{13}, 16, \mathbf{19}, 22, 25, 28, \mathbf{31}, 34, \ldots$$
$$\mathbf{2}, \mathbf{5}, 8, \mathbf{11}, 14, \mathbf{17}, 20, \mathbf{23}, 26, \mathbf{29}, 32, 35, \ldots$$

Легко видеть, что в первой из них содержится только одно простое число 3. Интересно, каково количество простых чисел в двух других.

Теорема. *Последовательность $3k+2$ содержит бесконечно много простых чисел.*

Доказательство. Попробуем действовать по аналогии с доказательство теоремы Евклида. Допустим противное, т.е. что множество простых чисел вида $3k+2$ конечно: 2, 5, 11, 17, …, p_n.

Рассмотрим число

$$N = 3 \cdot (2 \cdot 5 \cdot 11 \cdot \ldots \cdot p_n) + 2.$$

Замечание. Почему N имеет такой вид? Наша задача — «не выходить» за пределы арифметической прогрессии $3k + 2$ и найти в ней простое число, которого нет среди 2, 5, 11, 17, ..., p_n.

При доказательстве теоремы Евклида мы находили у построенного таким образом числа N простые делители, которых нет в нашем списке. Однако в нашем случае у числа N есть простой делитель из списка: $N \vdots 2$. Чтобы провести доказательство, аналогичное доказательсвту теоремы Евклида, рассмотрим число

$$N' = 3 \cdot (5 \cdot 11 \cdot \ldots \cdot p_n) + 2.$$

Рассмотрим множество *простых делителей* числа N'. Пусть q_1 — простой делитель числа N', который существует в силу леммы о простом делителе. Имеем

$$N' = q_1 \cdot N_1.$$

Если $N_1 > 1$, то процесс можно продолжить и рассмотреть q_2 — простой делитель N_1, который будет также делителем числа N':

$$N' = q_1 \cdot q_2 \cdot N_2.$$

Продолжим далее этот процесс, пока не получим

$$N' = q_1 \cdot q_2 \cdot \ldots \cdot q_l,$$

где $q_1, q_2, \ldots, q_l$ — все простые (не обязательно все различные) делители числа N. (Заметим, что процесс обязательно закончится, поскольку $N' > N_1 > N_2 > \ldots > 1$).

Очевидно, что среди простых делитей нет ни одного из чисел 2, 5, 11, 17, ..., p_n (именно для этого мы перешли от N к N'). Докажем, что найдется q_i такой, что $q_i = 3m + 2$ для некоторого m. Тем самым мы найдем простое число вида $3k + 2$, которого нет среди 2, 5, 11, 17, ..., p_n, что будет противоречить нашему предположению.

Предположим, что все $q_1, q_2, \ldots, q_l$ имеют вид $3k+1$. Докажем, что в таком случае само число N' должно иметь вид $3k+1$, что, очевидно, не так.

Имеем $q_1 = 3k_1 + 1$ и $q_2 = 3k_2 + 1$. Тогда

$$q_1 \cdot q_2 = (3k_1+1)\cdot(3k_2+1) = 9k_1k_2+3k_1+3k_2+1 = 3(3k_1k_2+k_1+k_2)+1.$$

Откуда, перемножая далее q_i, получаем

$$N' = q_1 \cdot q_2 \cdot \ldots \cdot q_l = 3K + 1.$$

Полученное противоречие доказывает теорему. □

Замечание. Вместо числа N' мы могли бы рассмотреть число

$$N'' = 3 \cdot (2 \cdot 5 \cdot 11 \cdot \ldots \cdot p_n) - 1.$$

Такая замена на первый взгляд совершенно неочевидна. Однако, в ней есть некоторый красивый смысл, с которым вы познакомитесь в главе IV.

Замечание. Обратите внимание, что мы нигде не ссылались на основную теорему арифметики, поскольку она сама по себе является фактом, который требует непростого доказательства!

По аналогии с только что доказанной теоремой можно показать, что простых чисел вида $4k+3$ или $6k+5$ бесконечно много. Однако, доказать аналогичным образом то же самое для чисел вида $3k+1$ не удастся.

Замечание. Полезно подумать, почему доказательство не работает для этих случаев.

На самом деле верен следующий общий результат:

Любая последовательность вида $an + d$, где наибольший общий делитель чисел a и d равен 1, содержит бесконечно много простых чисел. Эта теорема была доказана великим немецким математиком Иоганном Леженем Дирихле в 1837 году, и доказательство носит отнюдь не элементарный характер.

Теперь возникает вопрос, а что будет, если рассматривать не линейные последовательности вида $an + d$, а, скажем, квадратичные? Рассмотрим, например, последовательность $n^2 + 1$:

$$1, \mathbf{2}, \mathbf{5}, 10, \mathbf{17}, 26, \mathbf{37}, 50, 65, 82, \mathbf{101}, 122, 145, 170, \mathbf{197}, 226, \ldots$$

Верно ли, что она содержит бесконечно много простых чисел? Ответ на этот вопрос неизвестен.

А что если искать арифметические прогрессии в самом ряду простых чисел? Давайте выпишем несколько:

$3, 5, 7$
$5, 11, 17, 23, 29$
$7, 37, 67, 97, 127, 157$
$7, 157, 307, 457, 607, 757, 907$
$199, 409, 619, 829, 1039, 1249, 1459, 1669, 1879, 2089$

Можно ли утверждать, что в ряду простых чисел существуют арифметические прогрессии сколь угодно большой длины? Положительный ответ на этот вопрос дает теорема, доказанная в 2004 году выдающимися математиками Беном Грином и Теренсом Тао.

5. Отрезки с простыми числами и без

Давайте еще раз посмотрим на выписанный нами ряд простых чисел. Зафиксируем некоторое натуральное число n. Можно ли указать такое натуральное число N (зависящее от n), что на отрезке $[n; N]$ обязательно найдется простое число?

Теорема (Отрезки с простыми)**.** *Для любого натурального n на отрезке $[n; n! + 1]$ всегда есть простое число.*

Это означает, что можно взять $N = n! + 1$, и среди чисел

$$n, \quad n+1, \quad n+2 \quad \ldots \quad n!, \quad n!+1$$

обязательно найдется простое число.

Давайте проверим эту теорему для небольших значений n. Составим таблицу для n и N и выпишем простые числа на отрезке $[n; N]$:

n	N	простые на отрезке $[n; N]$
2	3	2, 3
3	7	3, 5, 7
4	25	5, 7, 11, 13, 17, 19, 23
5	121	5, 7, 11, 13, 17, 19, 23, 29, . . . , 119

Как видно, на нашем отрезке действительно есть простые числа, причем их очень много.

Доказательство. Рассмотрим простой делитель p числа $n!+1$, который существует по лемме о простом делителе. Очевидно, что $p \leqslant n! + 1$. Поскольку $n! + 1$ не делится ни на одно число, меньшее n, мы можем заключить, что $n < p$. □

На самом деле верно более сильное утверждение, которое называется *постулатом Бертрана* (хотя доказано оно было российским математиком Пафнутием Львовичем Чебышевым в 1852 году): *на отрезке* $[n; 2n]$ *всегда существует простое число.*

Как мы уже отмечали, существует много открытых проблем в теории чисел. Например, до сих пор недоказанной остается следующая *гипотеза Лежандра*: для любого натурального числа n на отрезке $[n^2; (n + 1)^2]$ найдется простое число.

Мы с вами доказали, что на отрезке $[n; n! + 1]$ всегда есть простое число. С другой стороны, существуют такие отрезки в натуральном ряду, на которых нет ни одного простого числа. Этим отрезкам на графике функции π отвечают отрезки ее постоянства.

Можно ли утверждать, что существуют сколь угодно длинные отрезки, на которых тоже нет ни одного простого? Оказывается, можно.

Теорема (Отрезки без простых)**.** *Для любого натуральноˊго n на отрезке*

$$[(n+1)!+2;(n+1)!+(n+1)]$$

длины n нет ни одного простого числа.

Таким образом, среди чисел

$$(n+1)!+2, \quad (n+1)!+3 \quad \dots \quad (n+1)!+(n+1)$$

нет ни одного простого.

Замечание. В формулировке теоремы участвует число $n+1$, поскольку в таком случае рассматриваемый отрезок имеет длину, равную n.

Доказательство. Как легко видеть,

$$\begin{aligned}
&(n+1)!+2 \text{ делится на } 2,\\
&(n+1)!+3 \text{ делится на } 3,\\
&\dots\dots\dots\dots\\
&(n+1)!+(n+1) \text{ делится на } (n+1).
\end{aligned}$$

Значит, все рассматриваемые числа являются составными. □

Давайте составив таблицу соответствующих отрезков для маленьких n.

n	левый конец	правый конец
2	8	9
3	26	28
4	122	125
5	722	726
6	5042	5047
7	40322	40328

Мы видим, что отрезки, построенные нами, расположены в натуральном ряду очень далеко. Но зато наша конструкция имеет общий характер и годится для любого натурального n.

Теперь возникает следующий вопрос. На отрезке $[1; 1000]$ 168 простых чисел. Мы можем указать отрезок, который не содержит ни одного простого числа, например, $[1001!+2; 1001!+1001]$. А существует ли отрезок из 1000 натуральных чисел, на котором, например, ровно 5 простых чисел? А если рассматривать отрезки произвольной длины? Зафиксируем произвольное натуральное число N. Тогда верно следующее

Утверждение. *Для любого натурального n такого, что*

$$1 \leqslant n \leqslant \pi(N),$$

существует отрезок из N натуральных чисел, на котором ровно n простых чисел.

Доказательство. Справедливость данного утверждения следует из так называемого принципа *дискретной непрерывности*, который оказывается очень полезным и при решении многих других задач.

Рассмотрим последовательность отрезков длины N:

$$\begin{aligned}
&\Delta_1 = [1, 2, \ldots, N];\\
&\Delta_2 = [2, 3, \ldots, N+1];\\
&\ldots\ldots\ldots\\
&\Delta_k = [k, k+1, \ldots, N+(k-1)];\\
&\ldots\ldots\ldots
\end{aligned}$$

При переходе от отрезка Δ_k к отрезку Δ_{k+1} количество простых чисел на них меняется не более, чем на 1. Отрезок Δ_1 содержит $\pi(N)$ простых чисел, на отрезке $\Delta_{(N+1)!+2}$ нет ни одного простого числа. Двигаясь от $\pi(N)$ до 0 с шагом длины 1, мы пройдем через каждое число в промежутке $[1, \pi(N)]$. $\square$

6. Существует ли формула?

В разное время предпринимались попытки найти выражение, значениями которого при разных значениях входящих в него переменных были бы простые числа. Леонард Эйлер, который внес значительный вклад в развитие всей математики и по праву признается одним из величайших мыслителей в истории нашей цивилизации, в том числе славился свои умением проводить огромные вычисления и замечать совершенно удивительные закономерности. Например, ему принадлежит следующий пример: многочлен $x^2 - x + 41$ принимает простые значения при любом целом неотрицательном x, не превосходящем 40. Менее известен также принадлежащий Эйлеру пример многочлена $x^2 - 79x + 1601$, значение которого при любом целом неотрицательном x, не превосходящем 79, является простым числом! Заметим, что до создания первого компьютера оставалось на тот момент чуть менее двухсот лет. Пример Эйлера побуждает задать вопрос: существует ли многочлен $p(x)$, значение которого при *всех* натуральных x является простым числом? Ответ дает следующая

Теорема. *Не существует многочлена положительной степени с целыми коэффициентами, значения которого при всех натуральных x — простые числа.*

Доказательство. Предположим, что такой многочлен существует

$$p(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x + a_0.$$

Будем без ограничения общности считать, что $a_0 \geqslant 0$.

Если $a_0 \neq 1$, то число $p(a_0)$ делится на a_0. Откуда следует, что, во-первых, число a_0 простое, во-вторых, $p(a_0) = a_0$. Более того, для любого натурального числа k число $p(ka_0)$ делится на a_0, а значит, $p(ka_0) = a_0$. Таким образом, многочлен $p(x) - a_0$ имеет *бесконечное множество корней*: $a_0, 2a_0, 3a_0, \ldots$, что невозможно. Дело в том, что количество корней многочлена не превосходит его степени. Соответствующую теорему мы докажем в VII.2.

Если $a_0 = 1$, то найдется такое натуральное число t, что свободный коэффициент многочлена

$$\widehat{p}(x) = p(x + t)$$

отличен от 1. Действительно, если предположить, что такого числа t не существует, то для любого натурального n

$$\widehat{a}_0 = \widehat{p}(0) = p(0 + n) = 1,$$

т.е. многочлен $p(x) - 1$ имеет бесконечное множество корней, что невозможно. Построив многочлен $\widehat{p}(x)$, свободный член которого $\widehat{a}_0$ отличен от 1, мы свели задачу к предыдущей. □

Замечание. Совершенно поразительным фактом является существование многочленов от *многих переменных*, множество положительных значений которых при неотрицательных значениях переменных совпадает с множеством простых чисел! Одним из примеров является следующий многочлен от 26 переменных $a, b, c, \ldots, x, y, z$.

$$\begin{aligned}
&(k+2)(1-[wz+h+j-q]^2-[(gk+2g+k+1)(h+j)+h-z]^2-\\
&[2n+p+q+z-e]^2-[16(k+1)^3(k+2)(n+1)^2+1-f^2]^2-\\
&[e^3(e+2)(a+1)^2+1-o^2]^2-[(a^2-1)y^2+1-x^2]^2-\\
&[16r^2y^4(a^2-1)+1-u^2]^2-[((a+u^2(u^2-a))^2-1)(n+4dy)^2+\\
&1-(x+cu)^2]^2-[n+l+v-y]^2-[(a^2-1)l^2+1-m^2]^2-\\
&[ai+k+1-l-i]^2-[p+l(a-n-1)+b(2an+2a-n^2-2n-2)-\\
&m]^2-[q+y(a-p-1)+s(2ap+2a-p^2-2p-2)-x]^2-\\
&[z+pl(a-p)+t(2ap-p^2-1)-pm]^2)
\end{aligned}$$

Этот многочлен был построен российским математиком Юрием Владимировичем Матиясевичем в работе, которая давала отрицательный ответ на 10-ую проблему Гильберта: не существует универсального алгоритма для решения диофантовых уравнений. Подробно об этих уравнениях мы будем говорить в главе V.

7. Числа Ферма, Мерсенна и правильные многоугольники

Теперь рассмотрим выражения, отличные от многочленов.

Числа Ферма

Начнем с чисел вида $2^n + 1$. Выпишем несколько чисел такого вида:

n	**1**	**2**	3	**4**	5	6	7	**8**	9	10	11	12
$2^n + 1$	**3**	**5**	9	**17**	33	65	129	**257**	513	1025	2049	4097

Внимательно изучив эту таблицу, можно придти к следующему выводу.

Утверждение. *Если у числа n есть нечетный делитель d, то число $2^n + 1$ является составным.*

Доказательство. Если $n = k \cdot d$, где d — нечетное, то

$$2^n + 1 = (2^k)^d + 1 = (2^k + 1) \cdot ((2^k)^{(d-1)} - (2^k)^{(d-2)} + \ldots - 2^k + 1).$$

□

Таким образом, простыми среди чисел $2^n + 1$ могут быть только числа вида

$$2^{2^m} + 1.$$

Эти числа обозначаются через F_m и называются *числами Ферма.*

Сам Ферма рассмотрел числа F_0, F_1, F_2, F_3, F_4.

$$F_0 = 2^{2^0} + 1 = 2^1 + 1 = 3$$
$$F_1 = 2^{2^1} + 1 = 2^2 + 1 = 5$$
$$F_2 = 2^{2^2} + 1 = 2^4 + 1 = 17$$
$$F_3 = 2^{2^3} + 1 = 2^8 + 1 = 257$$
$$F_4 = 2^{2^4} + 1 = 2^{16} + 1 = 65537$$

Как можно видеть, они являются простыми. Ферма выдвинул гипотезу о том, что число F_m будет простым для любого m. Эту гипотезу

опроверг Эйлер в 1732 году, найдя разложения числа F_5 на простые множители:

$$F_5 = 2^{2^5} + 1 = 2^{32} + 1 = 4294967297 = 641 \cdot 6700417.$$

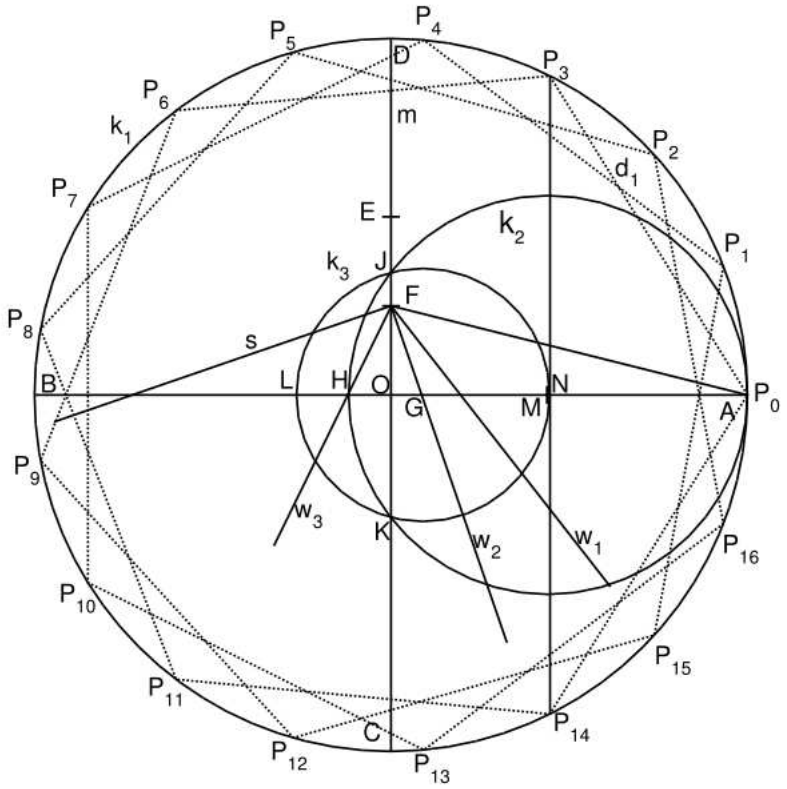
Более того, до сих пор не найдено ни одного простого числа Ферма, кроме F_0, F_1, F_2, F_3, F_4.

Совершенно невероятным образом простые числа Ферма появляются в геометрии. В работах древнегреческих математиков берет свое начало задача исследования построений правильных многоугольников с помощью циркуля и линейки. Античным геометрам были известны способы построения правильных n-угольников для $n = 2^k, 3 \cdot 2^k, 5 \cdot 2^k$ и $3 \cdot 5 \cdot 2^k$. Разумеется, возникает вопрос: *для каких n возможно построение правильных n-угольников с помощью циркуля и линейки*? Как и многие другие задачи, эта вошла в богатое наследство античной математики и вызвала интерес у европейских ученых. В частности, проблемой занимался молодой немецкий математик Гаусс (1777-1855). Ему удалось построить правильный 17-угольник. Приведем цитату из замечательной книги Гиндикина «Рассказы о физиках и математиках»:

«*Открытие, которого ждали две тысячи лет.* 1 июня 1796 г. в газете «Jenenser Intelligenzblatt» появилась заметка следующего содержания: «Всякому начинающему геометру известно, что можно геометрически (т. е. циркулем и линейкой) строить разные правильные многоугольники, а именно: треугольник, пятиугольник, пятнадцатиугольник и те, которые получаются из каждого из них путем последовательного удвоения числа его сторон. Это было известно во времена Евклида, и, как кажется, с тех пор было распространено убеждение, что дальше область элементарной геометрии не распространяется: по крайней мере, я не знаю удачной попытки распространить ее в эту сторону. Тем более кажется мне заслуживающим внимания открытие, что, кроме этих правильных многоугольников, может быть геометрически построено множество других, например семнадцатиугольник». Под заметкой стоит подпись: К.Ф. Гаусс из

Брауншвейга, студент-математик в Геттингене.»

Это событие явилось поворотным пунктом жизни Гаусса. Он принимает решение посвятить себя не филологии, а исключительно математике.



Построение правильного 17-угольника.

Приведем пример алгоритма построения:

1. Проводим большую окружность k_1 (будущую описанную окружность семнадцатиугольника) с центром O.
2. Проводим ее диаметр AB.
3. Строим к нему серединный перпендикуляр m, пересекающий k_1 в точках C и D.
4. Отмечаем точку E — середину DO.
5. Посередине EO отмечаем точку F и проводим отрезок FA.
6. Строим биссектрису w_1 угла $\angle OFA$.
7. Строим w_2 — биссектрису угла между m и w_1, которая пересекает AB в точке G.
8. Проводим s — перпендикуляр к w_2 из точки F.

9. Строим w_3 — биссектрису угла между s и w_2. Она пересекает AB в точке H.
10. Строим окружность k_2 на диаметре HA. Она пересекается с CD в точках J и K.
11. Проводим окружность k_3 с центром G через точки J и K. Она пересекается с AB в точках L и N.
12. Строим касательную к k_3 через N.

Точки пересечения этой касательной с исходной окружностью k_1 — это точки P_3 и P_{14} искомого семнадцатиугольника. Если принять середину получившейся дуги за P_0 и отложить дугу P_0P_{14} по окружности три раза, все вершины семнадцатиугольника будут построены.

Гаусс не был бы Гауссом, если бы после этого не доказал следующую теорему, носящую его имя.

Теорема (Гаусса)**.** *Правильный n-угольник возможно построить при $n = 2^k \cdot F_1 \cdot \ldots \cdot F_m$, где F_i — различные простые числа Ферма.*

Из доказательства этой теоремы следовало не только существование такого построения, но и его алгоритм. Спустя сорок лет в 1836 году французский математик Пьер Ванцель доказал, что других правильных многоугольников, которые можно построить циркулем и линейкой, не существует.

Как можно доказать *невозможность* построения чего-либо с помощью циркуля и линейки?! Еще древнегреческих математиков интересовали следующие знаменитые задачи:

- трисекция угла — разбить произвольный угол на три равные части;
- удвоение куба — построить ребро куба вдвое большего по объему, чем данный куб;
- квадратура круга — построить квадрат, равный по площади данному кругу.

Грекам их решить не удалось, как, впрочем, и их многочисленным последователям. Только в XIX веке было строго доказано, что все эти три задачи неразрешимы при использовании циркуля и линейки. Доказательство этих фактов было достигнуто с помощью алгебры и арифметики, а не геометрии! В частности, невозможность построения квадратуры круга следует из того, что не существует такого многочлена с целыми коэффициентами, корнем которого являлось бы число π. Числа, не являющиеся корнем никакого многочлена с целыми коэффициентами, называются *трансцендентными*. Именно трансцендентность π явилась преградой к построению квадратуры круга!

Числа Мерсенна

Рассмотрим числа вида $2^n - 1$.

n	1	**2**	**3**	4	**5**	6	**7**	8	9	10	11	12
$2^n - 1$	1	**3**	**7**	15	**31**	63	**127**	255	511	1023	2047	4095

По аналогии с числами Ферма укажем такие n, что число $2^n - 1$ заведомо будет составным.

Утверждение. *Если число n составное, то число $2^n - 1$ также является составным.*

Доказательство. Если $n = k \cdot l$, то

$$2^n - 1 = (2^k)^l - 1 = (2^k - 1) \cdot ((2^k)^{l-1} + (2^k)^{l-2} + \ldots + 1).$$

□

Числа

$$M_p = 2^p - 1,$$

где p — простое число, называются *числами Мерсенна* в честь современника Пьера Ферма французского математика Марена Мерсенна.

Рассмотрим несколько чисел M_p.

$$M_2 = 2^2 - 1 = 3$$
$$M_3 = 2^3 - 1 = 7$$
$$M_5 = 2^5 - 1 = 31$$
$$M_7 = 2^7 - 1 = 127$$
$$M_{11} = 2^{11} - 1 = 2047 = 23 \cdot 89$$
$$M_{13} = 2^{13} - 1 = 8191$$

Не все числа M_p являются простыми. Однако, в отличие от чисел Ферма, среди чисел Мерсенна простые встречаются чаще. В 1772 году Эйлер доказал, что число

$$M_{31} = 2^{31} - 1 = 2147483647$$

является простым. В течение более ста лет это число было самым большим известным простым числом. После создания компьютеров задача поиска простых чисел значительно упростилась, хотя до сих пор она требует значительных вычислительных мощностей. Необходимо отметить, что в случае рассмотрения не произвольного числа, а имеющего определенный вид, скажем, числа Мерсенна, существуют алгоритмы, позволяющие достаточно быстро определить, просто число или нет. Именно поэтому самыми большими известными простыми числами всегда были числа Мерсенна. На момент написания этих строк самым большим известным простым числом является следующее число Мерсенна:

$$M_{74207281} = 2^{74207281} - 1.$$

Оно было найдено 7 января 2016 года и его десятичная запись состоит из 22 338 618 знаков!

8. А все-таки формула существует!

Интенсивное развитие математики на рубеже XIX и XX веков привело к появлению совершенно удивительных результатов. Например,

было доказано, что существует такое (не целое!) число a, что *при всех натуральных* n число $[a^{3^n}]$ является простым, где $[x]$ — целая часть числа x. Однако указать, чему равно данное a не удается (нахождение значения требует знания всех простых чисел)! Существует множество подобных результатов, которые были получены только в середине XX века и требуют серьезной математической техники. В частности, в 1952 году польский математик Вацлав Серпинский доказал существование такого числа A, что n-ое простое число p_n можно записать формулой

$$p_n = [10^{2^n} \cdot A] - 10^{2^{n-1}} \cdot [10^{2^{n-1}} \cdot A].$$

Вот она — так долго искомая, но совершенно бесполезная для практических вычислений формула! Ведь указать, чему равно число A, невозможно.

В математике регулярно встречаются подобные результаты, называемые *теоремами существования*. Они гарантируют существование некоторых объектов, но совершенно ничего не говорят о том, как их явно построить и можно ли это сделать вообще.

Глава II

Делимость

В предыдущей главе мы изучали простые и составные числа. Однако множество натуральных чисел обладает одним очевидным недостатком: в нем отсутствует вычитание. Действительно, не существует, например, такого натурального числа x, что $x + 2 = 1$, т.е. среди натуральных числа $1 - 2$ нет. Нам с вами хорошо известно, что такое число есть среди *целых*: $1 - 2 = -1$.

Замечание. Минус единица стала научным теормином только во второй половине XIX века! До этого «это не более чем жаргон, в котором нет ни капли здравого смысла». Ни Декарт, Ферма, Лейбниц, Эйлер, Лагранж, ни Гаусс или Коши не могли дать определения отрицательных чисел! А мы сможем! В главе XI мы подробнее поговорим о причинах такого положения вещей и дадим строгие определения. А пока мы будем свободно пользоваться целыми числами, не давая строгих определений.

Отметим также, что рассмотрение множества целых чисел (которое обозначается буквой $\mathbb{Z}$) углубляет проводимую нами аналогию между числами и многочленами, поскольку последние можно свободно вычитать. Перейдем теперь к исследованию делимости целых чисел.

По определению целое число a *делится* на целое число $b \neq 0$, если существует такое целое число q, что $a = bq$. В таком случае число a называется *делимым*, b — *делителем*, а q — *частным*. Кратко это записывается следующим образом: $a \vdots b$.

Отметим следующие очевидные утверждения.

- Если $a \vdots c$ и $b \vdots c$, то $(a + b) \vdots c$ и $(a - b) \vdots c$.

  Обратите внимание, что в обратную сторону утверждение *невер-*

но: из условия $(a+b) \vdots c$ не следует, что $a \vdots c$ и $b \vdots c$.

- Если $a \vdots b$ и $b \vdots c$, то $a \vdots c$.
- Если $a \vdots c$ и $b \vdots d$, то $(ab) \vdots (cd)$.
- Если $a \vdots b$, то $(ac) \vdots (bc)$, где c — натуральное число.

Если вместо целых чисел рассматривать многочлены, то все высказанное остается истинным с точностью до замены слова «число» на слово «многочлен».

1. Деление с остатком

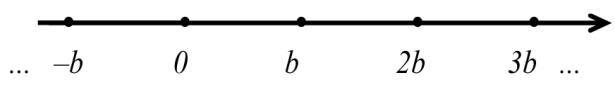
Как легко видеть, целые числа (как и многочлены) зачастую не делятся друг на друга нацело. В таком случае их можно *делить с остатком*.

Теорема (о делении с остатком)**.** *Для любых целых чисел a, b, таких, что $b \neq 0$, существуют единственные целые числа q и r, такие, что*

$$a = bq + r \quad \textit{и} \quad 0 \leqslant r < |b|.$$

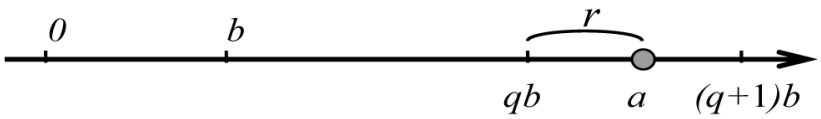
Определение. В этом случае число q называется *частным*, а число r — *остатком* от деления a на b.

Доказательство. Пусть b — некоторое целое число. Отметим на числовой оси целые числа, делящиеся на b. Они будут расположены на ней на равном расстоянии друг от друга. Эти числа называют еще *кратными* числу b.



Теперь отметим на той же числовой оси число a. Оно попадает в единственный полуинтервал между двумя выделенными числами, кратными b. Пусть это числа qb и $(q+1)b$ (см. рисунок).

Тогда число $r = a - qb$ есть целое число, удовлетворяющее неравенствам $0 \leqslant r < |b|$. □



При работе с целыми числами полезным оказывается следующее

Утверждение. *Среди n подряд идущих целых чисел найдется единственное число, кратное n.*

Доказательство. Как и в доказательстве теоремы о делении с остатком, отметим на числовой оси целые числа, делящиеся на n. Между двумя соседними будет ровно $n-1$ чисел. Теперь рассмотрим n подряд идущих целых чисел. Очевидно, что среди подряд идущих ровно одно отмеченное. □

Многочлены также можно делить с остатком. Обратите внимание, что во множестве чисел остаток меньше модуля делителя, а во множестве многочленов степень остатка меньше степени делителя, но деление осуществляется так же процедурой «деления уголком».

Например, разделим $3x^5+2x^4+x^2-x+1$ на x^3+2x^2+x с остатком.

$$
\begin{array}{l|l}
3x^5+2x^4+0\cdot x^3+x^2-x+1 & \underline{x^3+2x^2+x} \\
\underline{3x^5+6x^4+3x^3} & 3x^2-4x+5 \\
-4x^4-3x^3+x^2-x+1 & \\
\underline{-4x^4-8x^3-4x^2} & \\
-5x^3+5x^2-x+1 & \\
\underline{5x^3+10x^2+5x} & \\
-5x^2-6x+1 &
\end{array}
$$

Таким образом, получаем

$$3x^5+2x^4+x^2-x+1=(x^3+2x^2+x)\cdot(3x^2-4x+5)+(-5x^2-6x+1).$$

Теорема. *Пусть a и b — многочлены от одной переменной. Тогда существуют однозначно определенные многочлены q и r такие, что $a=bq+r$ и $\deg r<\deg b$.*

Доказательство. Существование таких многочленов q и r следует из процедуры «деления уголком».
Докажем *единственность* q и r. Пусть

$$a = bq_1 + r_1 = bq_2 + r_2,$$

где $\deg r_1 < \deg b$ и $\deg r_2 < \deg b$. Тогда

$$r_1 - r_2 = (q_2 - q_1)b$$

и, если $q_1 \neq q_2$, то

$$\deg(r_1 - r_2) = \deg(q_2 - q_1) + \deg b \geqslant \deg b,$$

что, очевидно, неверно. □

Замечание. Именно степень многочлена сыграла ключевую роль в доказательстве теоремы. Из доказательства также видно, что степень многочлена похожа на модуль целого числа. Такая похожесть еще больше сближает множества целых чисел и многочленов.

2. Алгоритм Евклида

Напомним, что *наибольшим общим делителем* целых чисел a и b называется наибольшее натуральное число d, на которое делятся оба числа a и b. Наибольший общий делитель обычно обозначается через НОД(a, b) или для краткости через (a, b).

Если НОД$(a, b) = 1$, то говорят, что числа a и b *взаимно просты*.

Наименьшим общим кратным целых чисел a и b называется наименьшее натуральное число k, которое делится на оба числа a и b. Наименьшее общее кратное обычно обозначается через НОК(a, b) или для краткости через $[a, b]$.

Ключевую роль при вычислении НОДа играет следующая

Теорема (Основное свойство НОД)**.**

$$НОД(a, b) = НОД(a - b, b).$$

Доказательство. Если k является общим делителем чисел a и b, то k — также общей делитель чисел $a - b$ и b, и наоборот. Таким образом, множества общих делителей пар чисел a, b и $a - b, b$ совпадают. Значит, совпадают и их НОДы. □

Эта теорема позволяет установить связь между поиском НОДа и делением с остатком.

Следствие. *Разделим число a на b с остатком: $a = bq + r$. Тогда*

$$НОД(a, b) = НОД(b, r).$$

Доказательство.

$$\text{НОД}(a, b) = \text{НОД}(a - b, b) = \ldots = \text{НОД}(a - qb, b) = \text{НОД}(b, r).$$

□

Для нахождения наибольшего общего делителя двух целых чисел существует *эффективный* алгоритм, основанный на доказанном нами следствии и который был описан еще в книге «Начал» и назван в честь их автора *алгоритмом Евклида.*

Пусть a и b — целые числа, причем $a \geqslant b$. Если $a \vdots b$, то b — искомое число. В противном случае надо разделить a на b с остатком r_1, после этого разделить с остатком b на r_1 и продолжить этот процесс, пока деление не будет выполнено без остатка.

Почему описанный процесс обязательно закончится?

Ключевым для ответа на этот вопрос является следующее соображение: при делении с остатком *остатки убывают.*

$$\begin{aligned} a &= bq_1 + r_1, \\ b &= r_1q_2 + r_2, \\ r_1 &= r_2q_3 + r_3, \\ &\ldots\ldots\ldots\ldots \end{aligned}$$

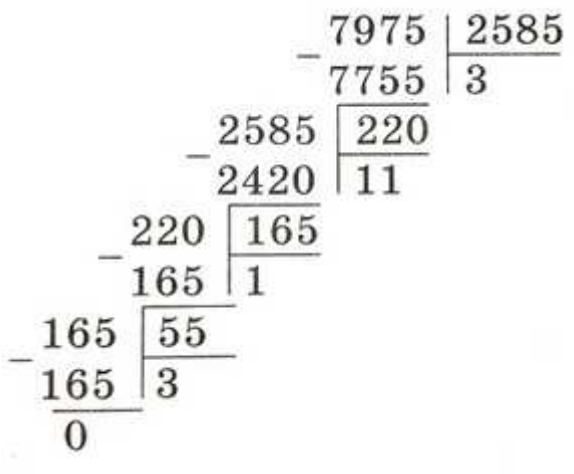
Откуда по теореме о делении с остатком получаем

$$|b| > r_1 > r_2 > r_3 > \ldots$$

Поскольку $r_i \geqslant 0$, в конце концов деление произойдет без остатка.

Если r_n — последний отличный от нуля остаток, то НОД$(a, b) = r_n$.

Ниже приведен пример вычисления НОД(7975, 2585).



Откуда следует, что НОД(7925, 2585) = 55.

Замечание. Существование быстрого алгоритма для определения простоты данного натурального числа породило бы большие трудности при кодировании информациии. Ключевым этапом взламывания алгоритма шифрования RSA является разложения натурального числа n на простые множители (как правило, это число имеет вид $n = pq$, где p и q — очень большие простые числа). Обычному компьютеру для выполнения этой задачи потребуется несколько тысяч лет. Однако, определение взаимной простоты двух натуральных чисел не представляет никакой сложности.

Вновь вернемся к аналогии с многочленами. Чтобы дать определение НОДа и НОКа для многочленов, замены словосочетания «целое число» на «многочлен» будет недостаточно, поскольку неясно, что означают слова *наибольший* или *наименьший* многочлен. Однако, как мы уже сказали, степень многочлена является аналогом модуля числа! Таким образом, получаем следующее

Определение. Наибольшим общим делителем многочленов a и b называется многочлен *наибольшей степени*, на который делятся оба многочлена a и b.

Наименьшим общим кратным многочленов a и b называется многочлен *наименьшей степени*, который делится на оба многочлена a и b.

Для нахождения наибольшего общего делителя двух многочленов успешно используется только что разобранный нами алгоритм Евклида. Необходимо только заменить модуль целого числа на степень многочлена.

Отметим, в частности, почему алгоритм Евклида, примененный к многочленам, *обязательно закончится.*
При делении с остатком *степени остатков убывают.*

$$a = bq_1 + r_1,$$
$$b = r_1q_2 + r_2,$$
$$r_1 = r_2q_3 + r_3,$$
$$\ldots\ldots\ldots\ldots$$

Откуда по теореме о делении с остатком получаем

$$\deg b > \deg r_1 > \deg r_2 > \deg r_3 > \ldots$$

Поскольку $\deg r_i \geqslant 0$, в конце концов деление произойдет без остатка.

Вычислим, например, НОД$(x^3 + x - 2, x^3 + x^2 - x - 1)$.

$$\begin{array}{r|l} x^3+x-2 & x^3+x^2-x-1 \\ \hline x^3+x^2-x-1 & 1 \end{array}$$

$$\begin{array}{r|l} x^3+x^2-x-1 & -x^2+2x-1 \\ \hline x^3-2x^2+x & -x-3 \end{array}$$

$$\begin{array}{r} 3x^2-2x-1 \\ \hline 3x^2-6x+3 \end{array}$$

$$\begin{array}{r|l} -x^2+2x-1 & 4x-4 \\ \hline -x^2+x & -\frac{1}{4}x+\frac{1}{4} \end{array}$$

$$\begin{array}{r} x-1 \\ \hline x-1 \\ \hline 0 \end{array}$$

Получаем, что НОД$(x^3 + x - 2, x^3 + x^2 - x - 1) = 4x - 4$.

Теперь возникает следующий вопрос. Мы определили НОД как многочлен наибольшей степени, на который делятся данные многочлены. $x^3 + x - 2$ и $x^3 + x^2 - x - 1$ делятся на $4x - 4$, но очевидно, что

они делятся и на $x-1$, и на $2x-2$, и вообще на любой многочлен вида $cx - c$, где c — не равное нулю число. При этом степени всех этих многочленов одинаковы и равны 1. Так какой же из них выбрать в качестве НОДа?! Ответ на этот вопрос таков: выбрать можно любой. В случае многочленов НОД (как и НОК) определен с точностью до умножения на ненулевое число (многочлен нулевой степени).

Причины этой неоднозначности кроются в свойстве многочленов нулевой степени, о которых мы подробнее поговорим в III.2.

Из алгоритма Евклида и в случае целых чисел, и в случае многочленов следует важное утверждение.

Теорема (О представлении НОД)**.** *Для любых элементов a и b $НОД(a,b)$ может быть представлен в виде $НОД(a,b) = au+bv$, где u, v — некоторые целые числа или, соответственно, многочлены.*

Доказательство. Для доказательства этого утверждения достаточно двигаться по цепочке равенств в алгоритме Евклида сверху вниз. Имеем:

$$\begin{aligned} a &= bq_1 + r_1, \\ b &= r_1q_2 + r_2, \\ r_1 &= r_2q_3 + r_3, \\ &\cdots\cdots\cdots\cdots \\ r_{n-1} &= r_nq_{n+1}, \end{aligned}$$

где r_n — искомый НОД. Тогда

$$\begin{aligned} r_1 &= a - bq_1, \\ r_2 &= b - r_1q_2 = \\ &= b - (a - bq_1)q_2 = \\ &= a(-q_2) + b(q_1q_2), \\ &\cdots\cdots\cdots\cdots \\ r_n &= au_n + bv_n. \end{aligned}$$

□

В главе IX мы увидим, что это утверждение окажется необходимым для доказательства основной теоремы арифметики.

Глава III

Основная теорема арифметики

Возможность разложения натурального числа на простые множители единственным образом является ключевым фактом в теории чисел. Фактом, который вам прекрасно знаком и которым вы (явно или неявно) почти всегда пользуетесь при работе с натуральными числами. Более того, этот факт кажется совершенно очевидным и воспринимается так же естественно, как восход солнца утром. Тем не менее, это не просто само собой разумеющийся факт, а

Теорема (Основная теорема арифметики)**.** *Каждое натуральное число n, большее 1, может быть разложено в произведение простых чисел*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

где $\alpha_1, \ldots, \alpha_k \geqslant 1$. Это разложение единственно с точностью до порядка простых сомножителей.

Замечание. Если бы 1 считалась простым числом, то нельзя было бы утверждать, что разложение на простые единственно. Как мы уже отмечали, причина этого заключается в том, что на 1 всегда можно делить (иными словами, число 1 является единственным обратимым элементом множества $\mathbb{N}$).

Замечание. Эта теорема отсутствует в «Началах» Евклида. Впервые ее точная формулировка и доказательство появились в книге «Арифметические исследования» (1801 г.) Гаусса, имя которого мы уже упоминали.

Причина, по которой строгое доказательство этой теоремы появилось только у Гаусса, по всей видимости, заключалась в том, что он был первым, кто систематически исследовал множества, кото-

рые похожи на цклые числа, но которые не всегда обладают всеми свойствами целых чисел. Например, мы познакомимся и с такими множествами, для которых не справедлива основная теорема арифметики!

1. Примеры Яглома и Гильберта

Сама по себе основная теорема арифметики не кажется чем-то удивительным. Ведь совершенно очевидно, что любое число раскладывается в произведение простых, разложение это единственно (например, $36 = 2^2 \cdot 3^2$), и что вообще можно извлечь интересного из такого тривиального факта?

Оказывается, что все далеко не так просто. Основная теорема арифметики отражает одно из важнейших свойств натуральных чисел: так называемую *факториальность*.

Рассмотрим несколько примеров, показывающих, как возникают сложности в таких вопросах.

Пример Яглома

Этот пример был предложен советским математиком Исааком Ягломом.

Давайте рассмотрим множество четных чисел:

$$2, \quad 4, \quad 6, \quad 8, \quad 10 \quad \ldots$$

Выясним, как будут выглядеть, так сказать, простые и составные числа в этом множестве.

По нашему определению *составным* называется число, которое может быть разложено в произведение двух чисел, отличных от него самого, а все остальные числа (кроме 1, которой у нас и так нет) называются *простыми*.

Число 2, очевидно, является простым. Число $4 = 2 \cdot 2$ представимо в виде произведения двух двоек и потому является составным.

Каким будет число 6 — простым или составным? Казалось бы $6 = 2 \cdot 3$. Но ведь числа 3 нет в нашем множестве! Значит, число 6 *не может быть разложено в произведение двух чисел из нашего множества* и потому *является простым*!

Утверждение. *Все четные числа, кратные 4, являются составными, а все некратные 4 — простыми.*

Доказательство. Рассмотрим четное натуральное число n.

Если $n \vdots 4$, то

$$n = (2 \cdot n_1) \cdot (2 \cdot n_2).$$

Таким образом, число n разложено в произведение двух четных отличных от него чисел, а значит n — *составное*.

Если $n \not\vdots 4$, оно не может быть разложено в произведение двух четных чисел, а значит является *простым*. □

Теперь давайте проверим, работает ли основная теорема арифметики в таком множестве чисел. Рассмотрим число 36. Имеем:

$$36 = 2 \cdot 18 = 6 \cdot 6.$$

Но и 2, и 6, и 18 являются простыми числами! Значит, число 36 можно разложить на простые множитель *двумя различными способами*!

Пример Гильберта

В примере Яглома мы рассмотрели множество четных чисел и выяснили, что основная теорема арифметики для него неверна. Можно ли подобный пример построить для нечетных чисел? Оказывается, можно. Следующий пример был предложен Давидом Гильбертом.

Рассмотрим следующее множество чисел вида $4k + 1$:

$$1, \quad 5, \quad 9, \quad 13, \quad 17, \quad 21 \quad \ldots$$

Опять попробуем найти простые и составные числа в этом множестве.

Легко видеть, что числа $5, 9, 13, 17, 21, 25$ в начале ряда являются простыми. Приведем несколько примеров составных чисел в этом

множестве (напомним, что число 1 не является ни простым, ни составным).

$$45 = 5 \cdot 9$$
$$65 = 5 \cdot 13$$
$$117 = 9 \cdot 13$$

Попробуем построить контрпример к основной теореме арифметики. Для этого заметим, что произведение двух чисел вида $4k + 3$ является числом вида $4k + 1$. Например, $3 \cdot 7 = 21$. Хотя числа 3 и 7 не принадлежат рассматриваемому нами множеству, ему принадлежит 21 (именно поэтому в примере Гильберта число 21 является простым). Поскольку мы ищем число, имеющее два разложения на простые, попробуем сконструировать его с использованием чисел 3 и 7. Рассмотрим число $441 = 3^2 \cdot 7^2$. Имеем

$$441 = 21 \cdot 21 = 9 \cdot 49.$$

И 9, и 21, и 49 являются простыми числами.

Попробуйте самостоятельно отыскать число, имеющее три различных разложения на простые множители в примерах Яглома и Гильберта.

2. $\mathbb{Z}[\sqrt{-k}]$ и великая теорема Ферма

Обратимся теперь к более содержательным и сложным примерам. Что такое $\mathbb{Z}[\sqrt{-k}]$, наверняка спрашиваете вы. Рассмотрим «числа» вида $a + b\sqrt{-k}$, где k — фиксированное *натуральное* число, а a и b — произвольные *целые* числа. Во-первых, отметим, что обычные целые числа содержатся среди рассматриваемых (достаточно положить $b = 0$). То есть только что определенное нами множество «чисел» представляет собой лишь расширение множества $\mathbb{Z}$ целых чисел. Во-вторых, разберемся с тем, как понимать, что такое $\sqrt{-k}$? Очень просто: $\sqrt{-k}$ — это «число», квадрат которого равен $-k$

$$(\sqrt{-k})^2 = \sqrt{-k} \cdot \sqrt{-k} = -k.$$

Замечание. Пока $\sqrt{-k}$ — это не более, чем символ. О строгом определении $\mathbb{Z}[\sqrt{-k}]$ мы будем говорить в XI.6.

Оперировать с новыми числами так же легко, как и с обычными целыми числами. Сложение определяется очевидным образом

$$(a+b\sqrt{-k})+(c+d\sqrt{-k})=(a+c)+(b+d)\sqrt{-k}.$$

Чтобы определить умножение, используем правила обычной арифметики и не забудем про равенство $(\sqrt{-k})^2=\sqrt{-k}\cdot\sqrt{-k}=-k$

$$\begin{aligned}&(a+b\sqrt{-k})\cdot(c+d\sqrt{-k})=\\ =&(a+b\sqrt{-k})\cdot c+(a+b\sqrt{-k})\cdot d\sqrt{-k}=\\ =&ac+bc\sqrt{-k}+ad\sqrt{-k}+bd\sqrt{-k}\cdot\sqrt{-k}=\\ =&(ac-bdk)+(bc+ad)\cdot\sqrt{-k}\end{aligned}$$

Таким образом, множество чисел вида $a+b\sqrt{-k}$ выдерживает сложение и умножение (что делает его похожим на множество $\mathbb{Z}$ целых чисел). Это множество и обозначается через $\mathbb{Z}[\sqrt{-k}]$. Обратимся теперь к конкретным примерам.

Как ошибся Эйлер

Рассмотрим множество $\mathbb{Z}[\sqrt{-3}]$ чисел вида

$$a+b\sqrt{-3},\quad \text{где}\quad a,b\in\mathbb{Z}.$$

Попробуем отыскать несколько простых и составных чисел в этом множестве. Начнем с целых чисел (которые являются элементами $\mathbb{Z}[\sqrt{-3}]$). Составные целые числа являются, очевидно, и составными числами в $\mathbb{Z}[\sqrt{-3}]$. А что можно сказать про простые целые числа? Будут ли они обязательно простыми в $\mathbb{Z}[\sqrt{-3}]$? Следующий пример показывает, что нет:

$$7=(2+\sqrt{-3})\cdot(2-\sqrt{-3}).$$

Вообще говоря, совершенно непонятно, как выяснить является ли данное число $a+b\sqrt{-3}$ простым. Чтобы ответить на этом вопрос, введем вспомогательное понятие.

Определение. *Нормой* числа $a + b\sqrt{-k}$ называется число

$$N(a + b\sqrt{-k}) = (a + b\sqrt{-k}) \cdot (a - b\sqrt{-k}) =$$
$$a^2 - b^2 \cdot \sqrt{-k} \cdot \sqrt{-k} = a^2 + kb^2.$$

Замечание. Число $a - b\sqrt{-k}$ называется *сопряженным* с числом $a + b\sqrt{-k}$.

Утверждение. *Пусть* $z, w \in \mathbb{Z}[\sqrt{-k}]$. *Тогда*

$$N(z \cdot w) = N(z) \cdot N(w).$$

Доказательство. Пусть $z = a + b\sqrt{-k}$, $w = c + d\sqrt{-k}$. Тогда

$$N((a + b\sqrt{-k}) \cdot (c + d\sqrt{-k})) = N((ac - kbd) + (bc + ad) \cdot \sqrt{-k}) =$$
$$(ac - kbd)^2 + k(bc + ad)^2 = (a^2 + kb^2) \cdot (c^2 + kd^2) =$$
$$N(a + b\sqrt{-k}) \cdot N(c + d\sqrt{-k}).$$

□

Замечание. Тождество Эйлера

$$(ac - kbd)^2 + k(bc + ad)^2 = (a^2 + kb^2) \cdot (c^2 + kd^2)$$

означает, что норма N обладает свойством мультипликативности: $N(z \cdot w) = N(z) \cdot N(w)$. Иными словами, произведение чисел вида $a^2 + kb^2$ является числом такого вида. Если $k = 1$, то мы приходим к рассмотрению чисел, представимых в виде суммы двух квадратов. Подробнее о них мы будем говорить в VIII.1.

Замечание. Норма — целое *неотрицательное* число. Более того, доказанное утверждение делает норму числа $a + b\sqrt{-k}$ еще больше похожей на модуль целого числа и степень многочлена. В дальнейшем мы увидим, что такая аналогия делает множества $\mathbb{Z}$, $\mathbb{Z}[\sqrt{-k}]$ и $\mathbb{R}[x]$ схожими, хотя на первый взгляд между ними нет ничего общего.

Докажем теперь, что число 2 является простым в $\mathbb{Z}[\sqrt{-3}]$. Действительно, предположим, что

$$2 = (a + b\sqrt{-3}) \cdot (c + d\sqrt{-3}).$$

Используя утверждение о норме произведения, получаем

$$N(2) = N((a + b\sqrt{-3}) \cdot (c + d\sqrt{-3})) = N(a + b\sqrt{-3}) \cdot N(c + d\sqrt{-3}),$$

что равносильно равенству

$$4 = (a^2 + 3b^2) \cdot (c^2 + 3d^2).$$

В нем каждый из сомножителей является натуральным числом, поэтому в правой части оба сомножителя либо равны 2, либо один из них равен 1, а другой 4. Первое невозможно, поскольку уравнение $a^2 + 3b^2 = 2$, очевидно, не имеет решений в целых числах. Во втором случае получаем уравнения

$$a^2 + 3b^2 = 1 \quad \text{и} \quad c^2 + 3d^2 = 4,$$

первое из которых имеет следующее решение

$$a = \pm 1,\ b = 0.$$

Тогда в исходном разложении числа 2 на множители один из них равен 1 (или -1), т.е. разложение тривиально и имеет вид $2 = 1 \cdot 2$ (или $2 = (-1) \cdot (-2)$). Значит, число 2 действительно простое.

Замечание. Данным пример в некоторой степени проясняет, почему нужно рассматривать именно множество $\mathbb{Z}[\sqrt{-k}]$, а не $\mathbb{Z}[\sqrt{k}]$. Во втором случае норма числа равнялась бы $a^2 - kb^2$, т.е. переставала бы быть неотрицательной, что не только принципиально само по себе, но и слишком осложнило бы поиск простых и составных чисел. Например, не очень понятно, имеет ли уравнение $a^2 - 3b^2 = 2$ решения и каковы они.

Разберем, на первый взгляд, более замысловатый пример и докажем простоту числа $1 + \sqrt{-3}$. Предположим обратное:

$$1 + \sqrt{-3} = (a + b\sqrt{-3}) \cdot (c + d\sqrt{-3}).$$

Используя утверждение о норме произведения, получаем

$$N(1+\sqrt{-3}) = N((a+b\sqrt{-3})\cdot(c+d\sqrt{-3})) = N(a+b\sqrt{-3})\cdot N(c+d\sqrt{-3}),$$

что, как несложно видеть, равносильно равенству

$$4 = (a^2 + 3b^2) \cdot (c^2 + 3d^2),$$

которое мы только что рассмотрели, доказывая простоту числа 2! Таким образом, число $1 + \sqrt{-3}$ также является простым.

Замечание. Совершенно ясно, что абсолютно также доказывается простота числа $1-\sqrt{-3}$. Более того, если число $a+b\sqrt{-3}$ оказывается простым, то таковым будет и сопряженное ему число $a - b\sqrt{-3}$.

Теперь возникает естественный вопрос, которому и посвящена данная глава. Выполнена ли основная теорема арифметики для $\mathbb{Z}[\sqrt{-3}]$? Оказывается, что нет! Например,

$$4 = 2 \cdot 2 = (1 + \sqrt{3}) \cdot (1 - \sqrt{-3}).$$

Теперь давайте разберемся, причем здесь Эйлер и какую же ошибку он допустил. С этим связана крайне интересная история, которая берет свое начало в III веке до н.э. в исследованиях древнегреческого математика Диофанта. Его имя нам известно благодаря его многотомной «Арифметике» из 13 книг. Основная тема «Арифметики» — решение уравнений в целых числах. К этой теме мы вернемся в дальнейшем, а сейчас нас будет интересовать одна из задач, упомянутых Диофантом, которая связана с описанием так называемых пифагоровых троек, т.е. целочисленных решений уравнения

$$x^2 + y^2 = z^2.$$

В главе V мы подробно поговорим об этой задаче.

Сочинения Диофанта долгое время оставались неизвестны европейским математикам. Только в XVI веке одна из его рукописей была случайно обнаружена в библиотеке Ватикана. Первый перевод «Арифметики» на латинский язык был издан в 1621 году, и владельцем одного из экземпляров стал Пьер Ферма. Значительная часть его наследия в области теории чисел — мысли, идеи, формулировки теорем, гипотезы — дошла до нас в форме записей и замечаний на

полях этого экземпляра «Арифметики». Напротив восьмой задачи второй книги Диофанта «разбить квадратное число на два других квадратных числа» (это и есть упамянутая выше задача) Ферма сделал самую знаменитую свою запись, в которой утверждалось, что при $n > 2$ *не существует отличных от нуля целых чисел, являющихся решением уравнения*

$$x^n + y^n = z^n.$$

Так формулируется «Великая теорема Ферма».

В математике часто бывает так, что задача, имеющая простую формулировку, требует крайне нетривиального решения. Великая теорема Ферма, пожалуй, самый знаменитый тому пример. На протяжении более чем трех сотен лет предпринимались попытки найти ее доказательство. Наконец, оно было получено в 1995 году английским математиком Эндрю Уайлсом после 8 лет работы. Его более чем 100-страничное доказательство отнюдь не элементарно. Не будет преувеличением сказать, что немногие люди во всем мире знакомы со всеми тонкостями этого доказательства. Единственный случай, для которого известно элементарное доказательство, — это случай $n = 4$. Указанное доказательство дано самим Ферма и опирается на упомянутые формулы для пифагоровых троек.

Следующий после Ферма шаг делает Эйлер — он доказывает теорему для случая $n = 3$, то есть показывает, что уравнение

$$x^3 + y^3 = z^3$$

не имеет нетривиальных решений в целых числах. Совершенно поразительным и принципиально новым в доказательстве Эйлера было привлечение чисел вида $a + b\sqrt{-3}$!

Замечание. С великой теоремой Ферма связана одна важная нить исторического развития алгебры. Хотя в формулировке теоремы участвуют только целые числа, в ходе ее доказательства потребовалось обратиться к числам более общего вида (например, $\mathbb{Z}[\sqrt{-k}]$). Глубокое проникновение в арифметику этих чисел потребовало введения но-

вых понятий и придало импульс развитию всей алгебры, теории чисел и других математических наук.

Доказательство, приводимое Эйлером, опиралось на следующее рассуждение: если взаимно простые целые числа a и b таковы, что a^2+3b^2 является кубом целого числа, то найдутся такие целые s и t, что

$$a = s(s^2 - 9t^2), \quad b = 3t(s^2 - t^2).$$

Эйлер рассматривает равенство

$$a^2 + 3b^2 = (a + b\sqrt{-3}) \cdot (a - b\sqrt{-3})$$

и утверждает, что *поскольку левая его часть является кубом, то то же справедливо и для обоих сомножителей правой части.* В частности,

$$a + b\sqrt{-3} = (s + t\sqrt{-3})^3.$$

Тогда, раскрывая скобки, получаем

$$\begin{aligned} &a + b\sqrt{-3} = (s + t\sqrt{-3})^3 = \\ &s^3 + 3s^2t\sqrt{-3} - 9st^2 - 3t^3\sqrt{-3} = \\ &(s^3 - 9st^2) + (3s^2t - 3t^3)\sqrt{-3}, \end{aligned}$$

откуда и следует, что

$$a = s^3 - 9st^2 = s(s^2 - 9t^2), \quad b = 3s^2t - 3t^3 = 3t(s^2 - t^2).$$

По всей видимости, Эйлер полагал очевидным, что взаимная простота a и b влечет взаимную простоту $(a + b\sqrt{-3})$ и $(a - b\sqrt{-3})$ и, тем самым, утверждение о том, что множители в правой части также являются кубами. Вообще говоря, это нуждается в доказательстве. Более того, учитывая, что для $\mathbb{Z}[\sqrt{-3}]$ не выполнена основная теорема арифметики, нельзя быть уверенным, что это вообще так! В данном конкретном случае единственность разложения $a^2 + 3b^2$ на простые множители удается доказать, но в ту же ловушку попадали другие математики, которые делали попытки дать общее доказательство теоремы Ферма.

Гауссовы числа

Множество $\mathbb{Z}[\sqrt{-1}]$ чисел вида $a + b\sqrt{-1},\ a, b \in \mathbb{Z}$ впервые появилось в работах Гаусса в 1832 году. Имеено Гауссом было открыто, что арифметические понятия и теоремы, которые всегда связывались с натуральными и целыми числами, можно переносить на другие объекты. Мы уже отмечали это на примере множества многочленов от одной переменной. Гаусс в своих исследованиях пришел к необходимости перенести известные теоремы на множество $\mathbb{Z}[\sqrt{-1}]$.

Прежде чем по аналогии с предыдущими примерами найти несколько простых и составных чисел в $\mathbb{Z}[\sqrt{-1}]$, мы опишем аналоги 1 в множестве гауссовых чисел. Что понимать под аналогом 1? Когда у нас появилось понятие простого числа, мы отмечали, что 1 не является ни простым, ни составным. Ведь если считать 1 простым, то в таком случае нельзя утверждать, что существует единственное разложение на простые множители. В случае натуральных чисел это замечание не кажется содержательным, однако, пример гауссовых чисел показывает, что не все так просто.

Причина происходящего кроется в том, что 1 является *обратимым* элементом множества натуральных чисел $\mathbb{N}$.

Определение. Элемент q данного множества называется *обратимым* (или *делителем единицы*), если существует такой s, принадлежащий данному множеству, что $q \cdot s = 1$.

Элемент s принято обозначат q^{-1}.

Как легко видеть, $1^{-1} = 1$. Покажем теперь, как наличие именно делителей единицы может «испортить» теорему о разложении на простые множители. Пусть q — делитель единицы и имеется разложение элемента n на простые множители

$$n = p_1 \cdot p_2 \cdot \ldots \cdot p_k.$$

Тогда имеет также место такое разложение

$$n = p_1 \cdot p_2 \cdot \ldots \cdot p_k \cdot q \cdot q^{-1}.$$

Продемонстрируем это явление на примере многочленов:

$$x^2-1=(x-1)\cdot(x+1)=2(x-1)\cdot\frac{1}{2}(x+1)=3(x-1)\cdot\frac{1}{3}(x+1)=\dots$$

Имея это в виду, несколько подкорректируем определение простого и составного элементов.

Определение. *Составным* называется *необратимый* элемент, который может быть разложен в произведение двух *необратимых* элементов.

Простым называется *необратимый* элемент, не являющийся составным.

Делители единицы (обратимые элементы) не являются ни простыми, ни составными.

Разложение на простые множители следует понимать с точностью до умножения на делители единицы (обратимые множители). Точно так же, как мы не различаем разложения натурального числа 36

$$36=2\cdot2\cdot3\cdot3=1\cdot2\cdot2\cdot3\cdot3=1\cdot1\cdot2\cdot2\cdot3\cdot3=\dots,$$

не отличают разложения

$$x^2-1=(x-1)\cdot(x+1)=2(x-1)\cdot\frac{1}{2}(x+1)=3(x-1)\cdot\frac{1}{3}(x+1)=\dots$$

или более общо

$$n=p_1\cdot\ldots\cdot p_k=q_1\cdot q_1^{-1}\cdot p_1\cdot\ldots\cdot p_k=q_1\cdot q_1^{-1}\cdot q_2\cdot q_2^{-1}\cdot p_1\cdot\ldots\cdot p_k=\dots,$$

где $q_1,q_2,\dots$ — делители единицы.

Замечание. Данное определение является универсальным. В частности, оно охватывает и случай натуральных чисел, и случай многочленов.

Найдем все делители единицы в множестве $\mathbb{Z}[\sqrt{-1}]$ гауссовых чисел. В этом нам поможет уже известное понятие нормы.

Теорема. *В* $\mathbb{Z}[\sqrt{-1}]$ *делителями единицы являются в точности следующие числа*

$$1,\ -1,\ \sqrt{-1},\ -\sqrt{-1}.$$

Доказательство. Пусть $q = a + b\sqrt{-1}$ — делитель единицы. Тогда существует $q^{-1} = c + d\sqrt{-1} \in \mathbb{Z}[\sqrt{-1}]$. Имеем

$$\begin{aligned}
&q \cdot q^{-1} = 1 \quad \Rightarrow \\
&N(q) \cdot N(q^{-1}) = N(q \cdot q^{-1}) = 1 \quad \Leftrightarrow \\
&N(a + b\sqrt{-1}) \cdot N(c + d\sqrt{-1}) = 1 \quad \Leftrightarrow \\
&(a^2 + b^2) \cdot (c^2 + d^2) = 1 \quad \Leftrightarrow \\
&a^2 + b^2 = 1, \quad c^2 + d^2 = 1,
\end{aligned}$$

откуда следует, что делители единицы содержатся среди чисел ± 1, $\pm\sqrt{-1}$. Простая проверка показывает, что все эти числа являются обратимыми. Действительно,

$$\begin{aligned}
&1 \cdot 1 = 1 \quad \text{т.е.} \quad 1^{-1} = 1 \\
&(-1) \cdot (-1) = 1 \quad \text{т.е.} \quad (-1)^{-1} = -1 \\
&\sqrt{-1} \cdot (-\sqrt{-1}) = 1 \quad \text{т.е.} \quad (\sqrt{-1})^{-1} = -\sqrt{-1} \\
&(-\sqrt{-1}) \cdot \sqrt{-1} = 1 \quad \text{т.е.} \quad (-\sqrt{-1})^{-1} = \sqrt{-1}.
\end{aligned}$$

□

Замечание. Из приведенного рассуждения сразу видно, что в случае произвольного $\mathbb{Z}[\sqrt{-k}]$ число $a + b\sqrt{-k}$ является делителем единицы только в том случае, если

$$N(a + b\sqrt{-k}) = a^2 + kb^2 = 1.$$

Откуда следует, что в случае $k > 1$ делителями единицы будут только числа ± 1.

Рассмотрим теперь несколько примеров. Число 2 является простым в целых числах и простым в $\mathbb{Z}[\sqrt{-3}]$. Однако, в $\mathbb{Z}[\sqrt{-1}]$ имеем

$$2 = (1 + \sqrt{-1}) \cdot (1 - \sqrt{-1}),$$

а значит, число 2 — составное в $\mathbb{Z}[\sqrt{-1}]$! А что можно сказать о числах $(1 \pm \sqrt{-1})$? Являются ли они простыми? Докажем, что они действительно простые. Предположим, что

$$1 + \sqrt{-1} = (a + b\sqrt{-1}) \cdot (c + d\sqrt{-1}).$$

Тогда получаем

$$N(1+\sqrt{-1}) = N((a+b\sqrt{-1})\cdot(c+d\sqrt{-1})) = N(a+b\sqrt{-1})\cdot N(c+d\sqrt{-1}),$$

что равносильно равенству

$$2 = (a^2 + b^2) \cdot (c^2 + d^2).$$

Из него вытекает, что либо $a^2 + b^2 = 1$, либо $c^2 + d^2 = 1$. В обоих случаях один из сомножителей в разложении

$$1 + \sqrt{-1} = (a + b\sqrt{-1}) \cdot (c + d\sqrt{-1})$$

является делителем единицы. Таким образом, число $1 + \sqrt{-1}$ (и $1 - \sqrt{-1}$) воистину простое.

Разумеется, возникает вопрос: верна ли основная теорема арифметики для гауссовых чисел? Рассмотрим следующий пример

$$5 = (1 + 2\sqrt{-1}) \cdot (1 - 2\sqrt{-1}) = (2 + \sqrt{-1}) \cdot (2 - \sqrt{-1}).$$

Легко убедиться, что числа, участвующие в данных разложениях, простые. Кажется, мы нашли пример, который показывает, что основная теорема арифметики не выполнена. Однако, не будем торопиться, ведь мы уже отмечали, что разложение на простые множители следует понимать с точностью до умножения на делители единицы. Если в случае натуральных чисел или многочленов сразу видно, какие разложения следует счиатать одинаковыми, то в случае гауссовых чисел это не столь очевидно. Обратимся к разложению числа 5:

$$\begin{aligned}
&5 = (1 + 2\sqrt{-1}) \cdot (1 - 2\sqrt{-1}) = \\
&\sqrt{-1} \cdot (2 - \sqrt{-1}) \cdot (-\sqrt{-1}) \cdot (2 + \sqrt{-1}) = \\
&\sqrt{-1} \cdot (-\sqrt{-1}) \cdot (2 - \sqrt{-1}) \cdot (2 + \sqrt{-1}) = \\
&\sqrt{-1} \cdot (\sqrt{-1})^{-1} \cdot (2 - \sqrt{-1}) \cdot (2 + \sqrt{-1})
\end{aligned}$$

Так что полученные нами разложения на самом деле суть одно! Все равно, что $2 \cdot 3$ и $1 \cdot 1 \cdot 2 \cdot 3$.

Совершенно удивительно, но основная теорема арифметики выполнена для множества гауссовых чисел! Правда, если для доказательства того, что теорема *не выполнена*, достаточно было всего лишь привести пример, то совершенно непонятно, как доказать, что теорема верна. Мы знакомы с несколькими примерами множеств, для которых имеет место основная теорема арифметики. Среди них множество $\mathbb{N}$ натуральных чисел, $\mathbb{Z}$ целых чисел, $\mathbb{R}[x]$ и $\mathbb{Q}[x]$ многочленов, теперь появился пример множества $\mathbb{Z}[\sqrt{-1}]$ гауссовых чисел. Позже мы дадим доказательство, которое будет универсальным и будет включать в себя все вышеперечисленные случаи.

3. Десять следствий

Примеры, разобранные нами в предыдущих разделах, показывают, что основная теорема арифметики — не очевидный факт, а глубокое свойство, которым обладают не только множества натуральных и целых чисел, но и множества многочленов и гауссовых чисел, как будет доказано в главе IX.

Естественно ожидать, что из этой теоремы следует много важных фактов. Для нас будут существенны следующие десять утверждений — следствий из основной теоремы арифметики.

Замечание. Эти следствия будут иметь место всегда, когда выполнена основная теорема.

Следствие 1. *Пусть*

$$a = p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_n^{k_n},$$
$$b = p_1^{l_1} \cdot p_2^{l_2} \cdot \ldots \cdot p_n^{l_n}$$

— разложение a и b на простые сомножители, причем $k_i \geqslant 0, l_i \geqslant 0$, но $k_i + l_i > 0$. Тогда

$$НОД(a, b) = p_1^{\min(k_1,l_1)} \cdot p_2^{\min(k_2,l_2)} \cdot \ldots \cdot p_n^{\min(k_n,l_n)}.$$

Пример. а) Пусть

$$a = 2^0 \cdot 3^2 \cdot 5^0 \cdot 11^1,$$
$$b = 2^3 \cdot 3^1 \cdot 5^2 \cdot 11^0.$$

Тогда

$$\text{НОД}(a, b) = 2^0 \cdot 3^1 \cdot 5^0 \cdot 11^0 = 3.$$

б) Пусть

$$a = (x-1)^0 \cdot (x^2+1)^2 \cdot (x^2+x+2)^0 \cdot (3x-1)^1,$$
$$b = (x-1)^3 \cdot (x^2+1)^1 \cdot (x^2+x+2)^2 \cdot (3x-1)^0.$$

Тогда

$$\text{НОД}(a, b) = (x-1)^0 \cdot (x^2+1)^1 \cdot (x^2+x+2)^0 \cdot (3x-1)^0 = x^2+1.$$

Следствие 2. *В обозначениях предыдущего следствия*

$$НОК(a, b) = p_1^{\max(k_1,l_1)} \cdot p_2^{\max(k_2,l_2)} \cdot \ldots \cdot p_n^{\max(k_n,l_n)}.$$

Следствие 3. *Имеет место формула*

$$НОД(a, b) \cdot НОК(a, b) = ab.$$

Замечание. Именно это следствие применяется для вычисления наименьшего общего кратного.

Следствие 4. *Если* $(ab) \vdots c$ *и* $НОД(a, c) = 1$, *то* $b \vdots c$.

Следствие 5 (Лемма Евклида о простом делителе)**.** *Если* p *простое и* $(ab) \vdots p$, *то либо* $a \vdots p$, *либо* $b \vdots p$.

Следствие 6. *Если* $НОД(b, c) = 1$ *и* $a \vdots b$, $a \vdots c$, *то* $a \vdots (bc)$.

Следствие 7. *Если* c *— общее кратное чисел* a *и* b, *то* $c \vdots НОК(a, b)$.

Следствие 8. *Если* d *— общий делитель чисел* a *и* b, *то* $НОД(a, b) \vdots d$.

Следствие 9. *Имеют место следующие формулы:*

$$НОД(ma, mb) = m \cdot НОД(a, b), \quad НОД(a/m, b/m) = НОД(a, b)/m,$$
$$НОК(ma, mb) = m \cdot НОК(a, b), \quad НОК(a/m, b/m) = НОК(a, b)/m.$$

Следствие 10. *Если* $НОД(a, b) = 1$ *и* $c^n = ab$, *то* $a = a_1^n,\ b = b_1^n$.

Замечание. Именно это следствие использовал Эйлер, ошибочно полагая что основная теорема арифметики выполнена для чисел $\mathbb{Z}[\sqrt{-3}]$.

Доказательства всех следствий напрямую следуют из основной теоремы арифметики.

4. Теорема Лежандра

Перед тем, как формулировать и доказывать следующую замечательную теорему, принадлежащую Лежандру (1752-1833), заметим, что задача разложения данного натурального числа на множители представляет собой довольно непростую задачу. Мы уже отмечали, что даже с использованием компьютера для разложения больших чисел потребуется значительное время. Удивительно, но если интересующее вас число имеет вид $n!$, то задача значительно упрощается.

Быстро разложить число $n!$ на множители позволяет следущая

Теорема (Лежандр)**.** *Простое число* p *входит в разложение числа* $n!$ *на простые множители в степени*

$$\left[\frac{n}{p}\right] + \left[\frac{n}{p^2}\right] + \left[\frac{n}{p^3}\right] + \dots$$

Замечание. 1. Здесь квадратные скобки обозначают *целую часть* числа, т.е. наибольшее целое число, не превосходящее данное. Например, $[5] = 5$, $[7{,}5] = 7$, $[1/3] = 0$, $[-1{,}5] = -2$. Ниже приведен график функции $y = [x]$.

2. Обратите внимание, что сумма, указанная в теореме, *конечная*: знаменатели дробей все время возрастают, так что рано или поздно

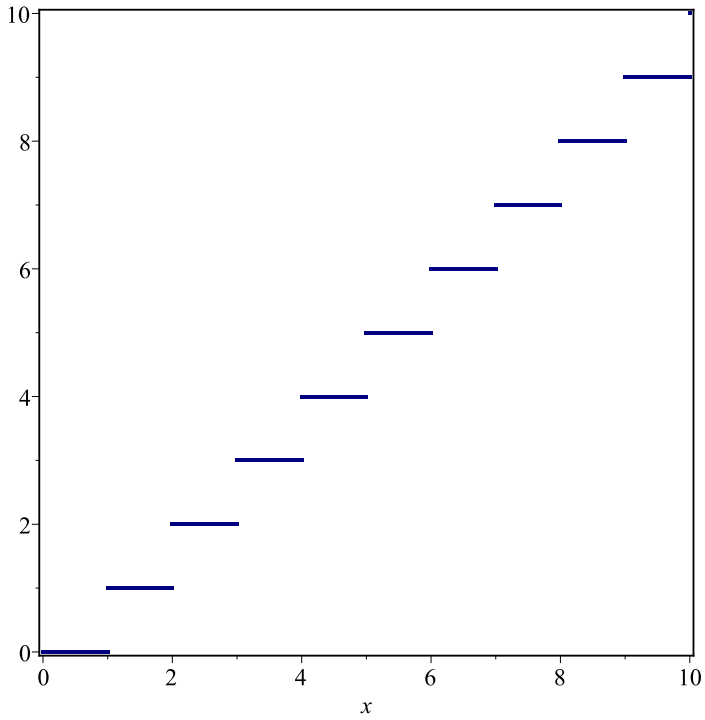


График функции $y = [x]$.

все дроби, начиная с некоторой, станут меньше 1, а значит, их целые части будут равны 0.

Доказательство. Рассмотрим натуральные числа от 1 до n. Ясно, что вклад в степень p дадут только те из этих чисел, которые делятся на p. таких чисел ровно $[n/p]$, и каждое из них даст вклад 1 в степень p.

Но это еще не все. Ведь среди чисел, делящихся на p, есть числа, делящиеся и на p^2. Каждое такое число уже даст вклад 2, а не 1 в степень p. Одну единицу мы уже учли в слагаемом $[n/p]$, а второе учитывается в слагаемом $[n/p^2]$.

Аналогично, рассматривая числа от 1 до n, делящиеся на p^3, p^4, ... мы последовательно будем учитывать вклады этих чисел в степень p в разложении $n!$, прибавляя $[n/p^3]$, $[n/p^4]$, ...

Таким образом, наша формула доказана. □

Теорема Лежандра позволяет также решить следующую интересную задачу. Существует ли многочлен с *нецелыми* коэффициентами,

значение которого в каждой *целой* точке является *целым* числом? Будем называть такие многочлены целозначными.

Теорема. *Многочлены*

$$C_{n+x}^{n} = \frac{(x+1)(x+2)\dots(x+n)}{1\cdot 2\cdot \dots \cdot n}$$

для любого натурального n являются целозначными.

$$C_{1+x}^{1} = x+1;$$
$$C_{2+x}^{2} = \frac{(x+1)(x+2)}{1\cdot 2} = \frac{x^2+3x+2}{2};$$
$$C_{3+x}^{3} = \frac{(x+1)(x+2)(x+3)}{1\cdot 2\cdot 3} = \frac{x^3+6x^2+11x+6}{6}.$$

Прежде чем доказывать теорему, рассмотрим вспомогательное утверждение.

Лемма. *Для любой пары чисел x и y выполнено неравенство*

$$[x]+[y] \leqslant [x+y].$$

Доказательство. Рассмотрим функцию

$$f(x,y) = [x]+[y]-[x+y].$$

Мы хотим доказать, что для любых x и y выполнено неравенство

$$f(x,y) \leqslant 0.$$

Очевидно, что

$$f(x+1,y) = f(x,y+1) = f(x,y).$$

В таком случае можно без ограничения общности считать, что

$$0 \leqslant x,y < 1.$$

Тогда

$$f(x,y) = [x]+[y]-[x+y] = 0-[x+y] \leqslant 0.$$

□

Теперь мы готовы доказать теорему.

Доказательство. Пусть $x = k \in \mathbb{Z}$. Имеем

$$C_{n+k}^{n} = \frac{(k+1)(k+2)\ldots(k+n)}{1 \cdot 2 \cdot \ldots \cdot n} = \frac{(k+n)!}{k! \cdot n!}.$$

Рассмотрим произвольное простое число p и докажем, что степень, в которой оно входит в разложение числителя на простые множители, не меньше, чем степень, в которой оно входит в разложение знаменателя на простые множители (и та, и другая степени могут быть равны 0). Тогда можно будет утверждать, что числитель делится на знаменатель, а значит, дробь является целым числом. В разложение числителя p входит в степени

$$a = \left[\frac{n+k}{p}\right] + \left[\frac{n+k}{p^2}\right] + \left[\frac{n+k}{p^3}\right] + \ldots$$

В разложение знаменателя — в степени

$$b = \left(\left[\frac{n}{p}\right] + \left[\frac{k}{p}\right]\right) + \left(\left[\frac{n}{p^2}\right] + \left[\frac{k}{p^2}\right]\right) + \left(\left[\frac{n}{p^3}\right] + \left[\frac{k}{p^3}\right]\right) + \ldots$$

Из леммы следует, что для любого m

$$\left[\frac{n}{p^m}\right] + \left[\frac{k}{p^m}\right] \leqslant \left[\frac{n+k}{p^m}\right],$$

откуда следует, что $b \leqslant a$. Таким образом, теорема доказана. □

Замечание. Наверняка вы спрашиваете, почему мы использовали именно такое обозначение C_{n+k}^{n}? Дело в том, что указанные числа имеют очень важное *комбинаторное* истолкование. А именно, для неотрицательных целых n и k число C_{n+k}^{k} — это количество способов выбрать любые n элементов из имеющихся $n+k$ различных.

Замечание. На первый взгляд целозначные многочлены могут выглядеть несколько неожиданно. Например,

$$g(x) = \frac{x^5}{5} + \frac{x^3}{3} + \frac{7x}{15}$$

является целозначным.

Оказывается, что в некотором смысле целозначные многочлены исчерпываются многочленами C_{n+x}^n. А именно, справедлива

Теорема. Пусть g — целозначный многочлен и $\deg g = m$. Тогда

$$g(x) = c_0 \cdot C_{m+x}^m + c_1 \cdot C_{(m-1)+x}^{m-1} + \ldots + c_m,$$

где $c_0, c_1, \ldots, c_m$ — целые числа.

Например, для приведенного выше многочлена имеем

$$\frac{x^5}{5} + \frac{x^3}{3} + \frac{7x}{15} = 24C_{5+x}^5 - 72C_{4+x}^4 + 80C_{3+x}^3 - 40C_{2+x}^2 + 9C_{1+x}^1 - 1.$$

5. Делители числа. Функции τ и σ.

Основная теорема арифметики позволяет провести более детальное исследование делителей натурального числа. Один из первых естественных вопросов о делителях — сколько их?

Определение. Функция, значение которой для натурального числа n равно числу его делителей, включая 1 и само число, называется *тау-функцией* и обозначается $\tau(n)$.

Прежде чем явно вычислять значение тау-функции, приведем ее график. Ниже представлен случай первых 60 натуральных чисел. Для наглядности мы соединили точки на втором графике отрезками.

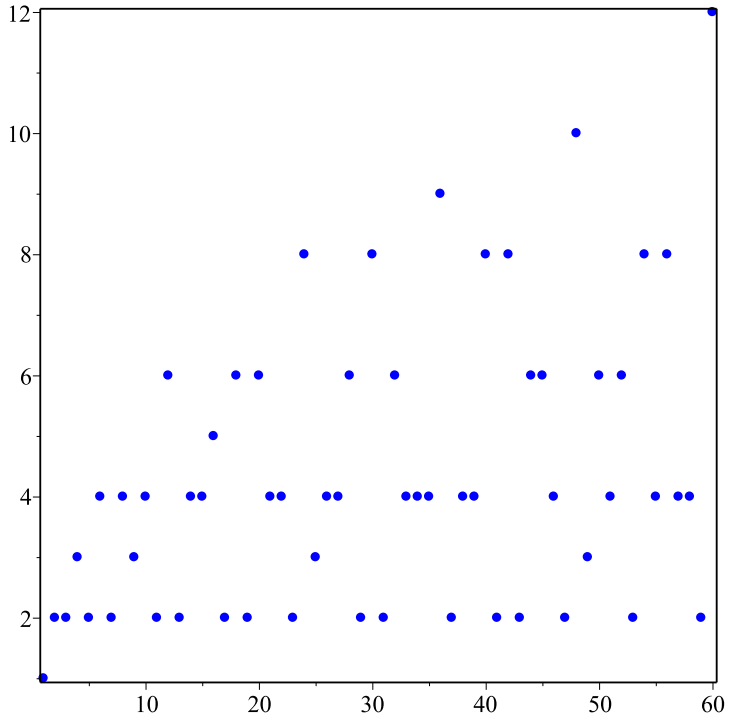


График $\tau(n)$ на отрезке от 1 до 60.

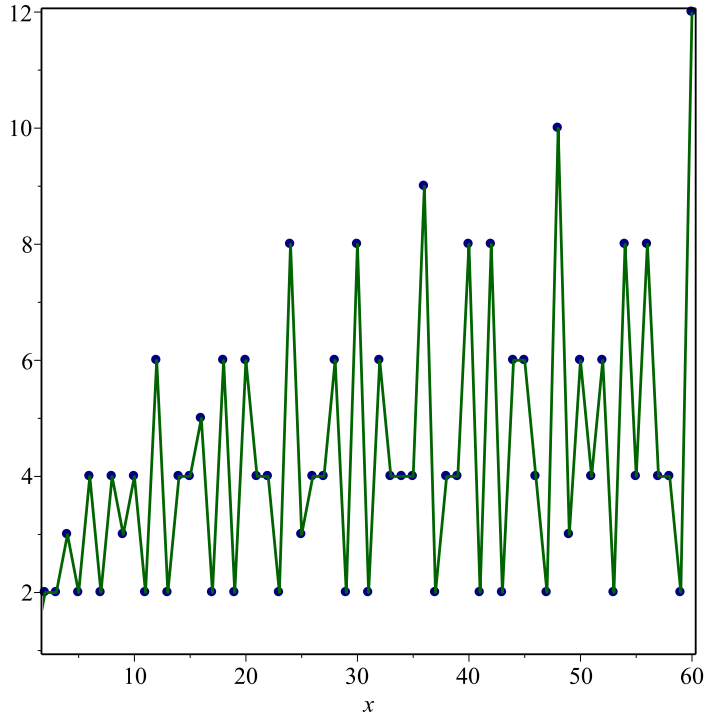


График $\tau(n)$ на отрезке от 1 до 60.

Для первой тысячи натуральных чисел график выглядит следующим образом.

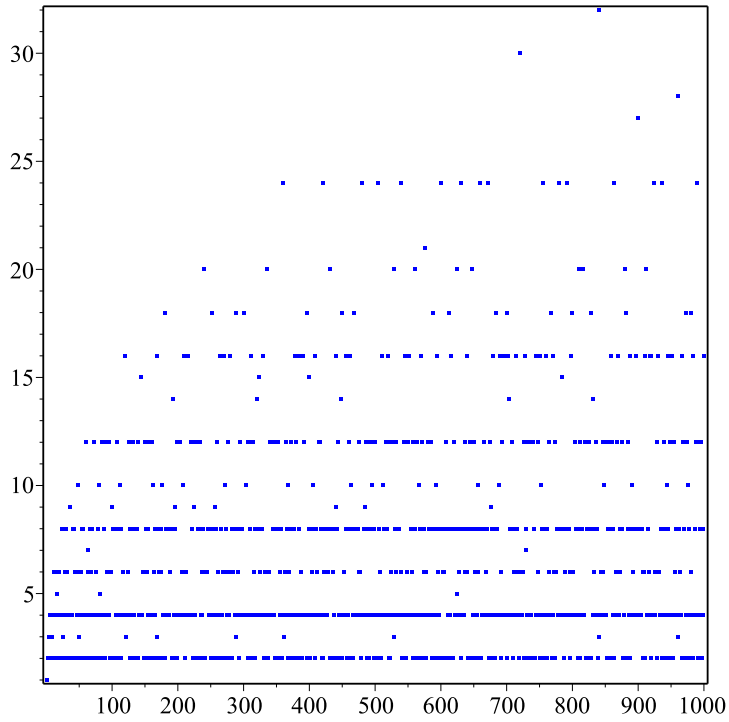


График $\tau(n)$ на отрезке от 1 до 1000.

Зная разложение натурального числа n на простые множители, можно явно вычислить значение $\tau(n)$. Введем необходимое в дальнейшем понятие.

Определение. Функция f, определенная на множестве натуральных чисел, называется *мультипликативной*, если

1. $f(1) = 1$;
2. $f(m \cdot n) = f(m) \cdot f(n)$ при условии НОД$(m, n) = 1$.

Вычисление значений таких функций основано на следующем утверждении.

Лемма (О мультипликативной функции)**.** *Если функция f — мультипликативная и*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

то

$$f(n) = f(p_1^{\alpha_1}) \cdot f(p_2^{\alpha_2}) \cdot \ldots \cdot f(p_k^{\alpha_k}).$$

Замечание. Как правило, значение мультипликативных функций достаточно легко вычисляется для степени простого числа. Благодаря доказанной лемме результат для произвольного натурального числа находится мгновенно.

Теперь мы готовы привести явную формулу для тау-функции.

Теорема (О количестве делителей)**.** *Если*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

то

$$\tau(n) = (\alpha_1 + 1) \cdot (\alpha_2 + 1) \cdot \ldots \cdot (\alpha_k + 1).$$

Доказательство. 1. Докажем, что функция τ является мультипликативной.

Очевидно, что $\tau(1) = 1$. Если НОД$(m, n) = 1$, то по основной теореме арифметики любой делитель числа $m \cdot n$ может быть единственным образом представлен в виде произведения делителей чисел m и n, и обратно, каждое такое произведение является делителем числа $m \cdot n$, откуда следует, что $\tau(m \cdot n) = \tau(m) \cdot \tau(n)$

2. Вычислим $\tau(p^\alpha)$ для простого p. Как легко видеть, число p^α имеет в точности $\alpha + 1$ делителей:

$$1,\ p,\ p^2,\ \ldots,\ p^{\alpha-1},\ p^\alpha.$$

Таким образом,

$$\tau(p^\alpha) = \alpha + 1.$$

По лемме о мультипликативной функции

$$\tau(n) = (\alpha_1 + 1) \cdot (\alpha_2 + 1) \cdot \ldots \cdot (\alpha_k + 1).$$

□

Следующим интересным объектом, связанным с делителями натурального числа, является их сумма.

Определение. Функция, значение которой для натурального числа n равно сумме его делителей, называется *сигма-функцией* и обозначается $\sigma(n)$.

$$\sigma(n) = \sum_{d\,:\,n\vdots d} d.$$

Как посчитать значение сигма-функции, зная разложение числа на простые множители? Перед тем как дать ответ, рассмотрим графики сигма-функции. Для первых 60-ти натуральных чисел график представлен ниже.

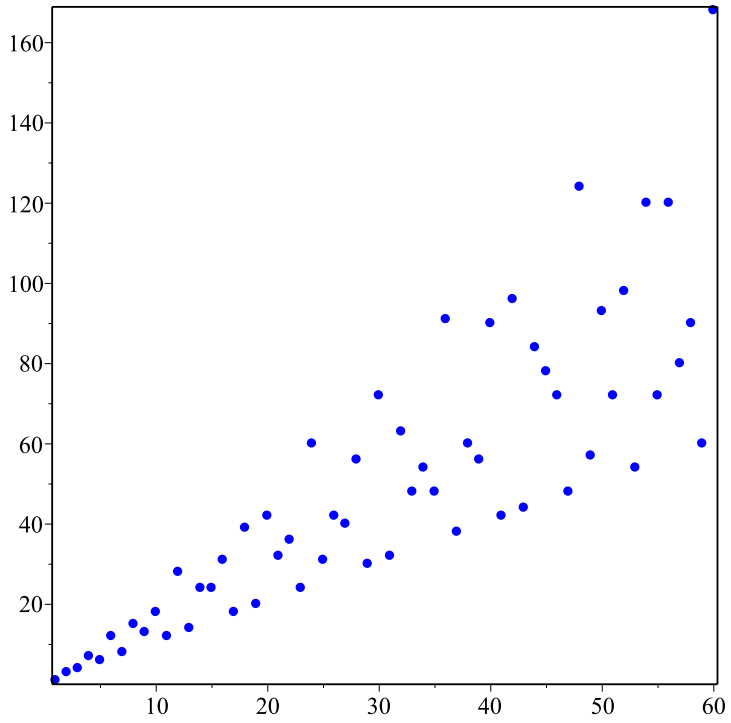


График $\sigma(n)$ на отрезке от 1 до 60.

Для наглядности соединим на втором рисунке соседние точки отрезками. Как мы уже могли наблюдать на других примерах, при рассмотрении большого отрезка натурального ряда картинка становится более регулярной.

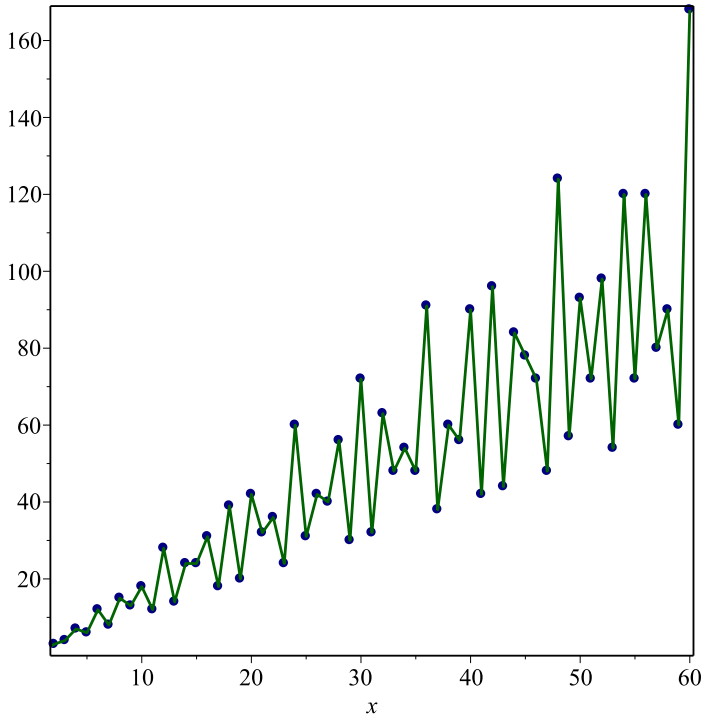


График $\sigma(n)$ на отрезке от 1 до 60.

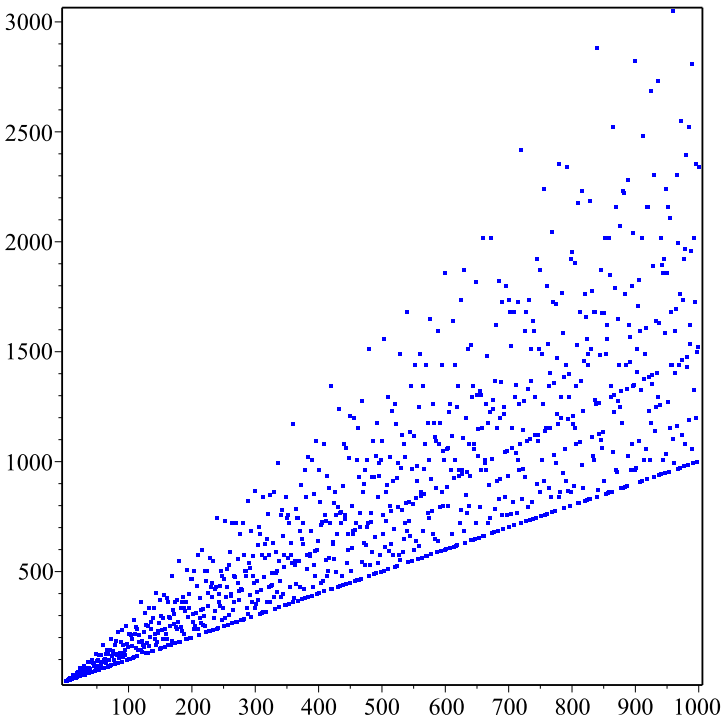


График $\sigma(n)$ на отрезке от 1 до 1000.

Теперь перейдем к выводу явной формулы для сигма-функции.

Теорема (О сумме делителей)**.** *Если*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

то

$$\sigma(n) = \frac{p_1^{\alpha_1+1} - 1}{p_1 - 1} \cdot \frac{p_2^{\alpha_2+1} - 1}{p_2 - 1} \cdot \ldots \cdot \frac{p_k^{\alpha_k+1} - 1}{p_k - 1}.$$

Доказательство. 1. Докажем, что функция σ является мультипликативной.

Очевидно, что $\sigma(1) = 1$. Если НОД$(m, n) = 1$, то по основной теореме арифметики любой делитель числа $m \cdot n$ может быть единственным образом представлен в виде произведения делителей $d_i \cdot c_j$ чисел m и n, и обратно, каждое такое произведение является делителем числа $m \cdot n$. Поэтому, если

$$\sigma(m) = \sum_i d_i \quad \text{и} \quad \sigma(n) = \sum_j c_j,$$

то

$$\sigma(m \cdot n) = \sum_{i,j} (d_i c_j) = \Big(\sum_i d_i\Big) \cdot \Big(\sum_j c_j\Big) = \sigma(m) \cdot \sigma(n).$$

2. Вычислим $\sigma(p^\alpha)$ для простого p. Как легко видеть, число p^α имеет в точности следующие делители:

$$1,\ p,\ p^2,\ \ldots,\ p^{\alpha-1},\ p^\alpha.$$

Таким образом,

$$\sigma(p^\alpha) = 1 + p + p^2 + \ldots + p^{\alpha-1} + p^\alpha = \frac{p^{\alpha+1} - 1}{p - 1}.$$

По лемме о мультипликативной функции

$$\sigma(n) = \frac{p_1^{\alpha_1+1} - 1}{p_1 - 1} \cdot \frac{p_2^{\alpha_2+1} - 1}{p_2 - 1} \cdot \ldots \cdot \frac{p_k^{\alpha_k+1} - 1}{p_k - 1}.$$

□

Замечание. Кроме τ и σ функций в теории чисел рассматриваются их обобщения — функции делителей:

$$\sigma_k(n) = \sum_{d\,:\,n\vdots d} d^k.$$

Применение этих функций не ограничивается теорией чисел. Например, они также появляются при исследовании эллиптических кривых, о которых мы скажем несколько слов в главе V.

6. Несократимые дроби и дзета-функция Римана

Основная теорема арифметики позволяет не только проводить конкретные вычисления, но и участвует в решении весьма нетривиальных задач.

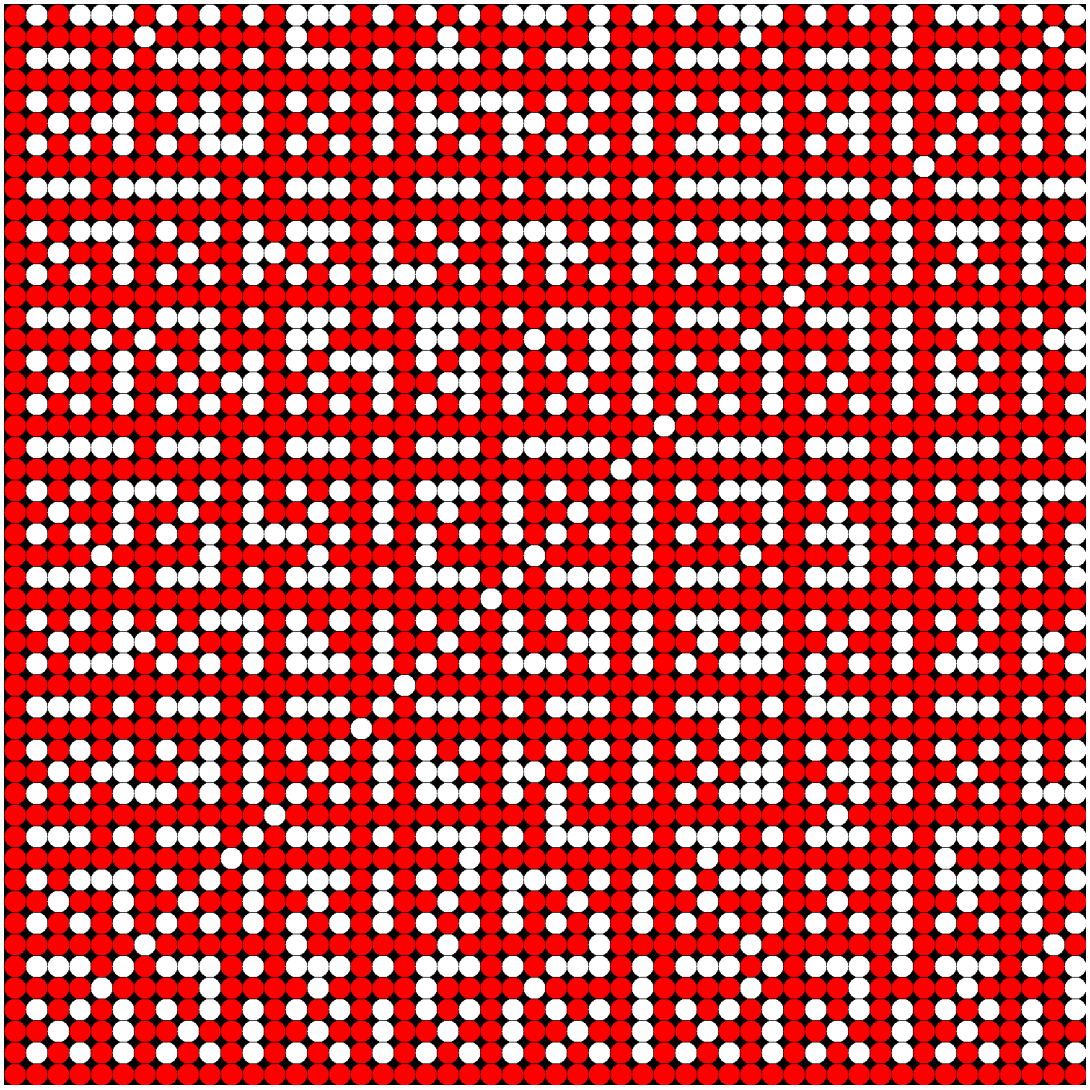
Обратимся к обыкновенным дробям. Некоторые из них сократимы, как $\frac{4}{6}$, другие — нет, как $\frac{5}{8}$.

Какова вероятность несократимости случайно выбранной дроби?

Дробь является несократимой тогда и только тогда, когда ее числитель и знаменатель взаимно просты. Рассмотрим точки плоскости, координаты которых являются натуральными числами. Назовем точку несократимой, если ее координаты — взаимно простые целые числа. Таким образом, несократимым дробям соответствуют несократимые точки плоскости, и наоборот. На рисунке ниже несократимые точки закрашены. Поэтому изначательный вопрос можно переформулировать так: какова доля закрашенных (несократимых) точек на плоскости?

Замечание. Несократимые точки на плоскости образуют замечательный узор. Можно отметить, что он симметричен относительно диагонали (почему?) и что он обладает фрактальной структурой, т.е. его составные части устроены, как целое.

Чтобы исследовать этот естественный и естественнонаучный вопрос, начнем с эксперимента и рассмотрим на плоскости квадрат со



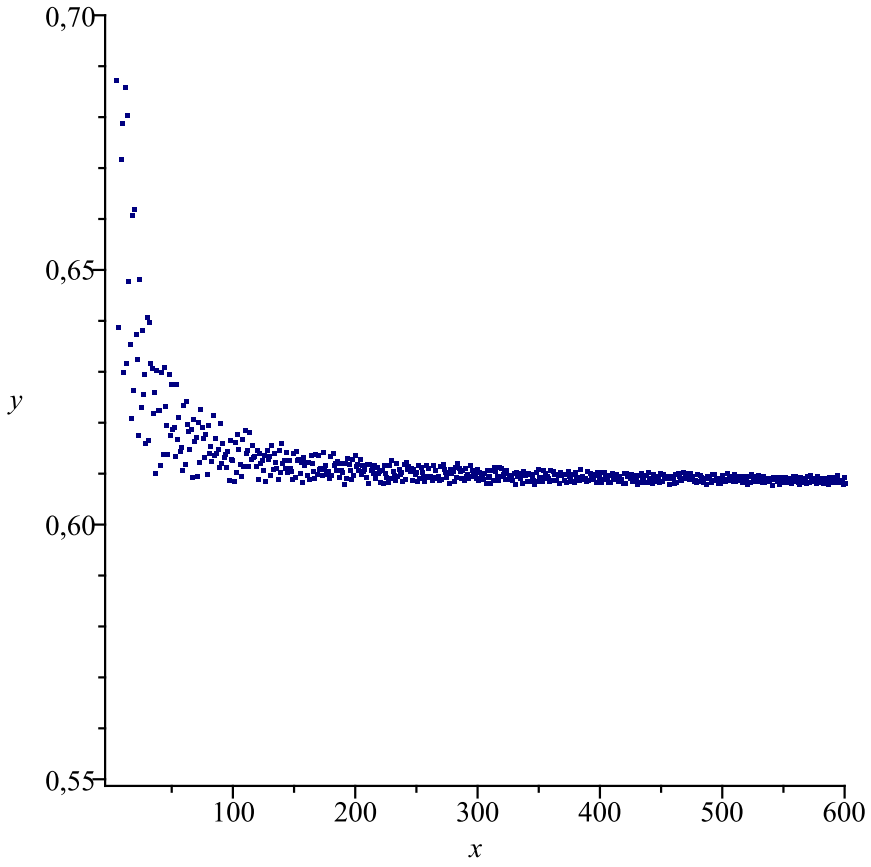
Узор пар взаимно простых чисел.

стороной 10. Число несократимых точек в нем составляет 63 (из общего числа ста целых точек в нем). Частота несократимости в данном квадрате получается равной 63/100, т.е. 63%. Если рассмотреть квадрат со стороной 30, то частота нескократимости получится равной $555/900 = 37/60$, т.е. примерно 61.7%.

Увеличивая сторону квадрата, можно убедиться, что частота несократимости стремится к некоторому фиксированному числу

$$P = 0.607927101\ldots$$

Эйлер смог найти его точное значение и ответить на поставленный вопрос. Пройдем и мы по его стопам.



Поведение доли нескатимых точек с ростом n.

Замечание. Прежде чем отыскать точный ответ, Эйлер экспериментально вычислил эту вероятность с точностью до 15-ого знака после запятой. По всей видимости, это позволило угадать правильный ответ.

Рассмотрим пару целых чисел (a, b). Их взаимная простота означает отсутствие общих делителей. Посчитаем вероятность этого отсутствия.

Какова вероятность того, что НОД(a, b) не делится на 2, т.е. среди общих делителей чисел a и b нет 2?

$$\text{НОД}(a, b) \vdots 2 \quad \Leftrightarrow \quad a \vdots 2, \;\; b \vdots 2.$$

Вероятность того, что a делится на 2 равна $1/2$. Аналогичное верно и для b. Поскольку делимость a на 2 не зависит от делимости b на 2,

вероятность того, что и a, и b одновременно делятся на 2, равна

$$\frac{1}{2}\cdot\frac{1}{2}=\frac{1}{2^2}.$$

Значит, вероятность того, что НОД(a,b) не делится на 2, равна $1-\frac{1}{2^2}$.

Вероятность того, что число делится на 3, равна 1/3. Поэтому аналогично предыдущему вероятность того, что НОД(a,b) не делится на 3, равна $1-\frac{1}{3^2}$. Для любого простого числа p вероятность того, что НОД(a,b) не делится на p, равна $1-\frac{1}{p^2}$.

Равенство НОД$(a,b)=1$ означает, что НОД(a,b) не делится ни на одно простое число p. Неделимости на разные простые числа независимы. Поэтому одновременная неделимость ни на одно из простых чисел имеет вероятность

$$P=\Big(1-\frac{1}{2^2}\Big)\cdot\Big(1-\frac{1}{3^2}\Big)\cdot\Big(1-\frac{1}{5^2}\Big)\cdot\ldots=\prod_{p-\text{простое}}\Big(1-\frac{1}{p^2}\Big).$$

Эйлер был первым, кто научился вычислять подобные *бесконечные* произведения. Чтобы и нам следовать за ним, понадобятся некоторые сведения о сумме геометрической прогрессии.

Замечание. Строгое обоснование дальнейших рассуждений будет дано вам в курсе математического анализа. Однако отсутствие этих обоснований не помешает нам получить искомый результат. В связи с этим уместно процитировать Арнольда: «Дальнейшие рассуждения можно строго обосновать и превратить в настоящее доказательство, но умение находить новые факты подобными нестрогими полуэмпирическими методами важнее последующих доказательств (которые могут появиться через сотни лет)».

Как вам хорошо известно, сумму геометрической прогресии можно вычислить следующим образом

$$1+q+q^2+\ldots+q^n=\frac{1-q^{n+1}}{1-q}=\frac{1}{1-q}-\frac{q^{n+1}}{1-q}.$$

Этого нам будет недостаточно. Что если рассмотреть *бесконечную* сумму

$$1+q+q^2+\ldots+q^n+\ldots$$

Можно ли отыскать ее точное значение?

Зафиксируем такое q, что $0 < q < 1$. Тогда с ростом n в разности

$$\frac{1}{1-q} - \frac{q^{n+1}}{1-q}$$

первое слагаемое не меняется, а второе стремится к 0. Неформально говоря, рассмотрение суммы бесконечной геометрической прогрессии означает, что мы «устремили» n к бесконечности и, тем самым, выражение $\frac{q^{n+1}}{1-q}$ можно считать равным нулю. В таком случае получаем равенство

$$1 + q + q^2 + \ldots + q^n + \ldots = \frac{1}{1-q},$$

где $0 < q < 1$.

Возвращаясь к вычислению эйлерова бесконечного произведения, нельзя не спросить, где же там «спрятана» сумма бесконечной геометрической прогресии?

Эйлер предложил считать не P, а P^{-1}! Имеем

$$P^{-1} = \prod_{p-\text{простое}} \frac{1}{1-\frac{1}{p^2}}.$$

Рассмотрим отдельный множитель этого произведения

$$\begin{aligned}\frac{1}{1-\frac{1}{p^2}} &= 1 + \frac{1}{p^2} + \left(\frac{1}{p^2}\right)^2 + \left(\frac{1}{p^2}\right)^3 + \ldots = \\ &= 1 + \frac{1}{p^2} + \frac{1}{p^4} + \frac{1}{p^6} + \ldots\end{aligned}$$

Вот же сумма бесконечной геометрической прогрессии, где в качестве q выступает $\frac{1}{p^2}$! Теперь мы имеем следующее бесконечное произведение бесконечных сумм

$$P^{-1} = \prod_{p-\text{простое}} \left(1 + \frac{1}{p^2} + \frac{1}{p^4} + \frac{1}{p^6} + \ldots\right)$$

Что с ним делать? Разумеется, раскрыть скобки!

Теорема (Эйлер)**.**

$$P^{-1} = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \ldots = \sum_{n\in\mathbb{N}} \frac{1}{n^2}$$

Доказательство. Чтобы получить слагаемое получившейся после раскрытия скобок суммы, мы должны выбрать по одному слагаемому из каждого множителя и перемножить их. Если в каждой скобке мы выберем по 1, то в результате получим 1. Если теперь мы в первой скобке выберем слагаемое $\frac{1}{2^2}$, а во всех остальных по 1, то получим в результирующей сумме $\frac{1}{2^2}$. Чтобы получить слагаемое $\frac{1}{3^2}$, необходимо во второй скобке выбрать $\frac{1}{3^2}$, а во всех остальных по 1 и так далее.

Раскрыв скобки, мы получим ряд, элементами которого являются конечные произведения квадратов различных простых чисел

$$\frac{1}{2^{2k_2}} \cdot \frac{1}{3^{2k_3}} \cdot \frac{1}{5^{2k_5}} \cdot \ldots = \frac{1}{n^2},$$

где $n = 2^{k_2} \cdot 3^{k_3} \cdot 5^{k_5} \cdot \ldots$. Эти произведения доставляют по разу в точности все натуральные числа в силу основной теоремы арифметики. □

Сумму обратных квадратов всех натуральных чисел Эйлер также умел считать. Ответ совершенно поразителен.

$$1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \ldots = \frac{\pi^2}{6}.$$

Замечание. Вычислять подобные суммы вы научитесь, изучая теорию рядов Фурье.

Итак, $P^{-1} = \pi^2/6$. Значит, искомая вероятность равна

$$P = \frac{6}{\pi^2} = 0.607927101\ldots$$

Эта работа привела Эйлера к открытию дзета-функции Римана

$$\zeta(s) = \sum_{n\in\mathbb{N}} \frac{1}{n^s}.$$

Таким образом получаем, что

$$\zeta(2) = 1 + \frac{1}{2^2} + \frac{1}{3^2} + \frac{1}{4^2} + \ldots = \frac{\pi^2}{6}.$$

Похожее выражение получится, если вычислить $\zeta(4)$:

$$\zeta(4) = 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \ldots = \frac{\pi^4}{90}.$$

Более того, можно вычислить значения дзета-функции во всех четных точках

$$\zeta(2k) = \gamma_k \cdot \pi^{2k},$$

где γ_k — некоторая рациональная константа, зависящая от k. А вот чему равно

$$\zeta(3) = 1 + \frac{1}{2^3} + \frac{1}{3^3} + \frac{1}{4^3} + \ldots$$

до сих пор неизвестно. Известно только, что число $\zeta(3)$ иррационально. Про остальные числа вида $\zeta(2k+1)$ известно только, что среди них бесконечно много иррациональных.

Об удивительных связях дзета-функции Римана с исследованиями перестановок и обобщений функций Эйлера вы можете прочитать в Добавлениях.

Из теоремы Эйлера также следует замечательное выражение для дзета-функции:

$$\zeta(s) = \prod_{p-\text{простое}} \frac{1}{1 - \frac{1}{p^s}}.$$

Если у вас сложилось впечатление, что значение дзета-функции можно вычислить для любого s, то это не так. Давайте убедимся, что вычислить $\zeta(1)$ не удастся. Значение дзета-функции — это сумма бесконечного ряда. Вообще говоря, такие ряды далеко не всегда можно просуммировать. Например, рассмотрим сумму бесконечного числа единиц: $1 + 1 + 1 + 1 + 1 + \ldots$. Очевидно, что эта сумма больше любого наперед заданного числа, потому не имеет конкретного значения. В таких случаях говорят, что рассматриваемый ряд *расходится*.

Покажем, что ряд

$$\zeta(1) = 1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \dots$$

расходится. Для этого разобъем в нем слагаемые на группы следующим образом

$$\begin{aligned} &1 + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8} + \dots = \\ =&1 + \left(\frac{1}{2}\right) + \left(\frac{1}{3} + \frac{1}{4}\right) + \left(\frac{1}{5} + \frac{1}{6} + \frac{1}{7} + \frac{1}{8}\right) + \dots > \\ >&1 + \left(\frac{1}{2}\right) + \left(\frac{1}{4} + \frac{1}{4}\right) + \left(\frac{1}{8} + \frac{1}{8} + \frac{1}{8} + \frac{1}{8}\right) + \dots = \\ =&1 + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \dots \end{aligned}$$

Откуда следует, что сумма обратных степеней натуральных чисел больше любого наперед заданного числа.

Замечание.

$$\zeta(1) = \prod_{p-\text{простое}} \frac{1}{1 - \frac{1}{p}}.$$

Из расходимости ряда $\zeta(1)$ следует бесконечность множества простых чисел. Ведь иначе произведение справа состояло бы из конечного числа слагаемых и имело бы конкретное значение.

В заключении отметим, что дзета-функция содержит в себе не только информацию о простых числах, она связана также с рядами Фурье и с подсчетом вероятностей, со статистикой и комплексным анализом, с квантовой физикой и теорией струн! (Некоторые из вас наверняка слышали о странной формуле

$$\zeta(-1) = 1 + 2 + 3 + 4 + \dots = -\frac{1}{12},$$

которая появляется в исследованиях по теории струн).

Особенно известна и примечательна задача описания нулей дзета-функции, т.е. таких s, что $\zeta(s) = 0$. Эта задача входит в список семи

так называемых «задач тысячелетия» — важных классических задач, решение которых не найдено вот уже в течение многих лет.

Решение этой труднейшей задачи повлекло бы за собой огромное количество потрясающих и глубоких результатов.

Глава IV

Кольца $\mathbb{Z}_m$ и их свойства

Зачастую когда речь идет о целых числах нет необходимости рассматривать сами числа. Иногда имеет смысл ограничиться лишь рассмотрением остатков. В чем преимущество рассмотрения именно остатков? Дело в том, что если мы фиксируем делитель m, то существует лишь *конечное* множество остатков от деления чисел на m: это 0, 1, 2, ..., $m-1$. Это множество обозначается через $\mathbb{Z}_m$. При этом каждый остаток получается при делении на m *бесконечного* набора чисел. Например, остаток 2 получится при делении на m чисел 2, $m+2$, $2m+2$, $3m+2$, ... Так появляется возможность свести задачу к конечному перебору. Эту чрезвычайно мощную идею демонстрирует, например, решение следующей задачи:

Найти все тройки простых чисел вида $p, p+2, p+4$.

Не имея никакой возможности перебрать все простые числа, мы обратимся к остаткам, которые дают эти числа при делении на 3. Таким образом, число p может иметь всего три остатка при делении на 3: 0, 1 или 2. Перебрав все три случая, мы убеждаемся, что из трех чисел $p, p+2, p+4$ одно обязательно делится на три.

p	$p+2$	$p+4$
0	2	1
1	0	2
2	1	0

Но среди простых чисел только одно делится на 3 без остатка — 3. Поэтому единственная тройка простых чисел вида $p, p+2, p+4$ — это 3, 5, 7.

С остатками можно работать, используя теорему о делении с остатком. Однако это неудобно. Зачастую достаточно рассматривать только остатки, потому возникает потребность выполнять с остатками стандартные арифметические операции так, чтобы опять получать остатки (т.е. не «выходить» из $\mathbb{Z}_m$).

Определение. *Суммой* (соответственно *произведением*) двух остатков $a, b \in \mathbb{Z}_m$ назовем остаток от деления на m обычной суммы (произведения) чисел a и b.

Замечание. Данное определение обладает некоторым недостатком. Чтобы определить операции в $\mathbb{Z}_m$, мы «вышли за его пределы». В XI.1 мы дадим другое определение множества $\mathbb{Z}_m$, лишенное данного недостатка.

В таблицах ниже приведен пример для $\mathbb{Z}_6$.

+	0	1	2	3	4	5
0	0	1	2	3	4	5
1	1	2	3	4	5	0
2	2	3	4	5	0	1
3	3	4	5	0	1	2
4	4	5	0	1	2	3
5	5	0	1	2	3	4

Таблица сложения по модулю 6

×	0	1	2	3	4	5
0	0	0	0	0	0	0
1	0	1	2	3	4	5
2	0	2	4	0	2	4
3	0	3	0	3	0	3
4	0	4	2	0	4	2
5	0	5	4	3	2	1

Таблица умножения по модулю 6

Для работы с остатками используется техника *сравнений*.

1. Сравнения и признаки делимости

Определение. Целые числа a и b *сравнимы по модулю* m, если они имеют один и тот же остаток при делении на m. Это обозначается следующим образом: $a \equiv b \pmod{m}$.

Сравнение $\equiv$ *похоже* на равенство $=$. Теперь мы попробуем перенести известные свойства равенства на сравнения. Чтобы это сделать нам потребуется

Лемма. $a \equiv b \pmod m$ *тогда и только тогда, когда* $(a-b) \vdots m$.

Доказательство. Если $a \equiv b \pmod m$, то

$$\begin{aligned} &a = k_1 m + r,\ b = k_2 m + r \quad \Rightarrow \\ &a - b = (k_1 m + r) - (k_2 m + r) = (k_1 - k_2)m \quad \Rightarrow \\ &(a-b) \vdots m. \end{aligned}$$

Если $(a-b) \vdots m$, то

$$\begin{aligned} &a = k_1 m + r_1,\ b = k_2 m + r_2 \quad \Rightarrow \\ &a - b = (k_1 - k_2)m + (r_1 - r_2) \vdots m. \end{aligned}$$

Значит, $(r_1 - r_2) \vdots m$, и так как $|r_1 - r_2| < m$, то $r_1 - r_2 = 0$. □

Основные свойства сравнений.

1) если $a \equiv b \pmod m$, то $a + c \equiv b + c \pmod m$;
2) если $a \equiv b \pmod m$ и $c \equiv d \pmod m$, то $a + c \equiv b + d \pmod m$;
3) если $a \equiv b \pmod m$, то $ac \equiv bc \pmod m$;
4) если $a \equiv b \pmod m$ и $c \equiv d \pmod m$, то $ac \equiv bd \pmod m$;
5) если $ac \equiv bc \pmod m$ и НОД$(c, m) = 1$, то $a \equiv b \pmod m$;
6) если $ac \equiv bc \pmod{cm}$, то $a \equiv b \pmod m$.

Доказательство всех основных свойств следует напрямую из леммы и основной теоремы арифметики.

Множество $\mathbb{Z}_m$ и теория сравнений играют одну из ключевых ролей в теории чисел и имеют многочисленные приложения, к некоторым из которых мы теперь обратимся.

Наверное, вы помните *признаки делимости* на 3 и на 9: число n делится на 3 (соответственно на 9), если и только если сумма его цифр делится на 3 (соответственно на 9). Оказывается, имеет место более сильное утверждение.

В дальнейшем будет полезно ввести следующее обозначение. Обозначим через $S(n)$ сумму цифр числа n.

Теорема (О сумме цифр)**.** *Имеют место следующие сравнения:*

$$n \equiv S(n) \pmod 3 \quad \textit{и} \quad n \equiv S(n) \pmod 9.$$

Доказательство. Запишем число n в десятичной записи:

$$n = \overline{a_k a_{k-1} a_{k-2} \dots a_1 a_0}, \quad \text{где } a_0, a_1, \dots, a_k \text{ — цифры от 0 до 9}$$

(горизонтальная черта над цифрами обозначает десятичную запись). Тогда

$$\begin{aligned} n &= 10^k \cdot a_k + 10^{k-1} \cdot a_{k-1} + 10^{k-2} \cdot a_{k-2} + \dots + 10^1 \cdot a_1 + 10^0 \cdot a_0 \equiv \\ &\equiv 1^k \cdot a_k + 1^{k-1} \cdot a_{k-1} + 1^{k-2} \cdot a_{k-2} + \dots + 1^1 \cdot a_1 + 1^0 \cdot a_0 = \\ &= a_k + a_{k-1} + a_{k-2} + \dots + a_1 + a_0 = S(n) \pmod 9. \end{aligned}$$

Сравнение по модулю 3 доказывается совершенно аналогично. □

Замечание. По существу, эта теорема — единственное утверждение, связанное с суммой цифр натурального числа. Так что если вам попадется задача, в которой присутствует сумма цифр, прежде всего нужно сравнить ее по модулям 3 или 9.

Как вы прекрасно знаете, число делится на 2 тогда и только тогда, когда его последняя цифра делится на 2. Что можно сказать о делимости на 4, 8, 16 и другие степени 2? Докажем соответствующую теорему.

Теорема (О делимости на 2^m)**.** *Если $n = \overline{a_k a_{k-1} a_{k-2} \dots a_1 a_0}$ — десятичная запись натурального числа n, то*

$$n \equiv \overline{a_{m-1} a_{m-2} \dots a_1 a_0} \pmod{2^m}.$$

Доказательство. Имеем

$$\begin{aligned} n &= \overline{a_k a_{k-1} a_{k-2} \dots a_m} \cdot 10^m + \overline{a_{m-1} a_{m-2} \dots a_1 a_0} \equiv \\ &\equiv \overline{a_{m-1} a_{m-2} \dots a_1 a_0} \pmod{2^m}. \end{aligned}$$

□

Перейдем теперь к признаку делимости на 11. Для этого заметим, что часто полезно рассматривать *отрицательные* остатки. Например, $-2 \equiv 3 \pmod 5$ или $-1 \equiv 10 \pmod{11}$. Смысл такого перехода заключается в уменьшении (по абсолютной величине) возникающих остатков (например, $|-2| = 2 < 3 = |3|$).

Теорема (О делимости на 11)**.** *Если* $n = \overline{a_k a_{k-1} a_{k-2} \dots a_1 a_0}$ *— десятичная запись натурального числа* n*, то*

$$n \equiv a_0 - a_1 + a_2 - \ldots + (-1)^{k-1} a_{k-1} + (-1)^k a_k \pmod{11}.$$

Доказательство. Имеем

$$\begin{aligned} n &= 10^k \cdot a_k + 10^{k-1} \cdot a_{k-1} + \ldots + 10^1 \cdot a_1 + 10^0 \cdot a_0 \equiv \\ &\equiv (-1)^k \cdot a_k + (-1)^{k-1} \cdot a_{k-1} + \ldots + (-1)^1 \cdot a_1 + (-1)^0 \cdot a_0 = \\ &= a_0 - a_1 + a_2 - \ldots + (-1)^{k-1} a_{k-1} + (-1)^k a_k \pmod{11}. \end{aligned}$$

□

Наконец, перейдем к признакам делимости на 7 и 13. Давайте попробуем по аналогии с признаками делимости на 9 и на 11 сравнить десятичную запись числа n, например, по модулю 7. Если вы попробуете сделать это, то убедитесь, что ничего хорошего не получается. Что же делать?

Оказывается, сделать нужно следующее. Давайте попробуем найти число, делящееся на 7 и отличающееся от какой-то степени 10 на 1. На эту роль подходит замечательное число

$$\mathbf{1001} = 7 \cdot 11 \cdot 13.$$

Очень полезно запомнить это «волшебное» число 1001 — оно очень часто возникает в разных задачах (в том числе и на разных олимпиадах).

Теперь мы готовы сформулировать признаки делимости на 7 и на 13.

Теорема. *Если* $n = \overline{a_k \dots a_5 a_4 a_3 a_2 a_1 a_0}$ *— десятичная запись натурального числа* n*, то*

$$n \equiv \overline{a_2 a_1 a_0} - \overline{a_5 a_4 a_3} + \ldots \pmod{7, 13}.$$

Замечание. Мы разбиваем цифры числа n на тройки (которые мы рассматриваем как трехзначные числа), начиная с конца, а потом

складываем их с чередующимися знаками. При этом последняя тройка может быть неполной и состоять из одной или двух цифр. Например,

$$1234567890 \equiv 890 - 567 + 234 - 1 = 556 \pmod{7, 13}.$$

Доказательство. Имеем

$$\begin{aligned} n &= \overline{a_2a_1a_0} + 10^3 \cdot \overline{a_5a_4a_3} + 10^6 \cdot \overline{a_8a_7a_6} + \ldots \equiv \\ &\equiv \overline{a_2a_1a_0} + (-1) \cdot \overline{a_5a_4a_3} + (-1)^2 \cdot \overline{a_8a_7a_6} + \ldots = \\ &= \overline{a_2a_1a_0} - \overline{a_5a_4a_3} + \overline{a_8a_7a_6} - \ldots \pmod{7, 13}. \end{aligned}$$

□

2. Делители нуля

Продолжим изучение множества $\mathbb{Z}_m$ и исследуем некоторые «странности», связанные с умножением остатков.

Давайте составим таблицы умножения остатков по модулям 5 и 6 (мы не будем включать в таблицу умножение остатков на 0, т.к. на что ни умножь 0, все равно 0 и получишь).

×	1	2	3	4
1	1	2	3	4
2	2	4	1	3
3	3	1	4	2
4	4	3	2	1

Таблица по модулю 5

×	1	2	3	4	5
1	1	2	3	4	5
2	2	4	0	2	4
3	3	0	3	0	3
4	4	2	0	4	2
5	5	4	3	2	1

Таблица по модулю 6

Что можно заметить, глядя на эти таблицы?

Во-первых, есть таблицы, в которых нет нулей, а есть таблицы, в которых нули есть. Например, имеет место следующее удивительное равенство:

$$2 \cdot 3 \equiv 0 \pmod{6}.$$

Опять-таки, если проводить параллель «остатки–числа», ничего подобного для целых чисел нет! Не существует двух ненулевых чисел, произведение которых равно 0. Более того, ни с чем подобным мы не сталкивались, рассматривая примеры многочленов или чисел $\mathbb{Z}[\sqrt{-k}]$. А вот для остатков такое иногда (правда, не всегда) возможно.

Определение. Ненулевой элемент $a \in \mathbb{Z}_m$ называется *делителем нуля*, если существует такой ненулевой элемент $b \in \mathbb{Z}_m$, что $ab \equiv 0 \pmod{m}$.

Замечание. Определения даются не только для того, чтобы подчеркнуть и зафиксировать аналогии, но и для того, чтобы зафиксировать различия. В дальнейшем мы увидим, насколько существенны различия между множествами остатков, в которых делители нуля есть и в которых они отсутствуют (таковыми, как мы вскоре покажем, являются $\mathbb{Z}_p$, где p — простое).

Как понять, является ли данный остаток делителем нуля? Не перебирать же всевозможные произведения, ища в них 0! Ответ на этот вопрос дают следующая теорема.

Теорема (О делителях нуля)**.** *Ненулевой элемент $a \in \mathbb{Z}_m$ является делителем нуля, если и только если НОД$(a, m) > 1$.*

Доказательство. 1. Пусть элемент $a \in \mathbb{Z}_m$ является делителем нуля, т.е. существует такой *ненулевой* элемент $b \in \mathbb{Z}_m$, что $ab \equiv 0 \pmod{m}$. Докажем, что НОД$(a, m) > 1$.

Предположим противное: пусть НОД$(a, m) = 1$. Тогда по основному свойству 5 сравнений имеем

$$ab \equiv 0 \pmod{m} \quad \Leftrightarrow \quad b \equiv 0 \pmod{m},$$

что противоречит условию.

2. Обратно, пусть НОД$(a, m) = d > 1$. Докажем, что существует такой ненулевой остаток $b \in \mathbb{Z}_m$, что $ab \equiv 0 \pmod{m}$.

Имеем

$$a = da', \quad m = dm'.$$

Тогда

$$am' = da'm' = a'(dm') = a'm \equiv 0 \pmod{m}.$$

Итак,

$$b = m' \not\equiv 0 \pmod{m}.$$

□

3. Делители единицы

Мы умеем складывать, вычитать, умножать остатки. Теперь наша цель — выяснить, когда можно делить остатки, т.е. когда $\frac{a}{b} = a \cdot b^{-1} \in \mathbb{Z}_m$, где $a, b \in \mathbb{Z}_m$. Иными словами, нам необходимо отыскать такие элементы $b \in \mathbb{Z}_m$, на которые всегда можно делить. Ими являются, как мы уже отмечали в предыдущих главах, в точности обратимые элементы или делители единицы в $\mathbb{Z}_m$. Напомним

Определение. Элемент $a \in \mathbb{Z}_m$ называется *делителем единицы* (или обратимым элементом, см. III.2), если существует такой элемент $b \in \mathbb{Z}_m$, что $ab \equiv 1 \pmod{m}$.

В ранее рассмотренных нами примерах многочленов от одной переменной и $\mathbb{Z}[\sqrt{-k}]$ множество делителей единицы были устроено довольно просто. В случае $\mathbb{Z}_m$ ситуация совершенно иная.

Давайте посмотрим на строчки наших таблиц. Видно, что иногда в них есть единицы, а иногда нет.

Вновь проведем параллель «остатки–числа». Рассмотрим уравнение $3x = 1$. У него есть (рациональный) корень $x = \frac{1}{3}$; это число именно так и определяется: $\frac{1}{3}$ — это такое число, которое, будучи умноженным на 3, дает 1.

А теперь давайте перейдем к остаткам. Рассмотрим таблицу умножения по модулю 5. Видно, что $3 \cdot 2 \equiv 1 \pmod{5}$. Значит, можно

написать, что

$$\frac{1}{3} \equiv 2 \pmod 5!$$

С другой стороны, по модулю 6 не существует остатка, который, будучи умноженным на 3, дает 1, так что делить можно не всегда. Рассмотрим еще несколько примеров

$$\frac{1}{3} \equiv 5 \pmod 7;$$
$$\frac{1}{3} \equiv 3 \pmod 8;$$
$$\frac{1}{4} \equiv 7 \pmod 9.$$

Теперь мы можем сказать, что если $b \in \mathbb{Z}_m$ обратим, то $\frac{a}{b} = a \cdot b^{-1} \in \mathbb{Z}_m$. Поскольку среди остатков встречаются делители нуля, для которых не существует обратного элемента по умножению (т.е. если $c \in \mathbb{Z}_m$ — делитель нуля, то c является необратимым элементом в $\mathbb{Z}_m$), то делить можно не всегда. Как же выяснить, на какие элементы $\mathbb{Z}_m$ можно делить?

Вот ответ и на этот вопрос.

Теорема (О делителях единицы)**.** *Элемент $a \in \mathbb{Z}_m$ является делителем единицы, если и только если НОД$(a, m) = 1$.*

Прежде чем приступить к ее доказательству, докажем следующее утверждение.

Лемма (О колоде карт)**.** *Если элемент $a \in \mathbb{Z}_m$ таков, что НОД$(a, m) = 1$, то*

$$\{1,\ 2,\ 3,\ \ldots,\ m-1\} = \{a \cdot 1,\ a \cdot 2,\ a \cdot 3,\ \ldots,\ a \cdot (m-1)\},$$

т.е. эти множества элементов $\mathbb{Z}_m$ совпадают.

Перед доказательством рассмотрим пример. Пусть $a = 7,\ m = 12$.

$$\{1,\ 2,\ 3,\ 4,\ 5,\ 6,\ 7,\ 8,\ 9,\ 10,\ 11\} = \{7,\ 2,\ 9,\ 4,\ 11,\ 6,\ 1,\ 8,\ 3,\ 10,\ 5\}$$

Доказательство. Рассмотрим m произведений

$$a \cdot 1,\ a \cdot 2,\ a \cdot 3,\ \ldots,\ a \cdot (m-1).$$

Докажем, что все эти произведения различны. В самом деле, если $a \cdot k \equiv a \cdot l \pmod{m}$ для некоторых k и l, то, используя основное свойство 5 сравнений, получаем $k \equiv l \pmod{m}$, а т.к. $1 \leqslant k, l \leqslant m-1$, то $k = l$. На самом деле мы умножили a на один и тот же остаток. □

Теперь мы готовы дать доказательство теоремы о делителях единицы.

Доказательство. 1. Пусть элемент $a \in \mathbb{Z}_m$ является делителем единицы, т.е. существует такой элемент $b \in \mathbb{Z}_m$, что $ab \equiv 1 \pmod{m}$. Докажем, что НОД$(a, m) = 1$.

Из того, что $ab \equiv 1 \pmod{m}$ следует равенство $ab = 1 + km$ для некоторого целого k. Тогда НОД$(ab, km) = 1$, откуда следует, что НОД$(a, m) = 1$.

2. Обратно, пусть НОД$(a, m) = 1$. Докажем, что существует такой остаток $b \in \mathbb{Z}_m$, что $ab \equiv 1 \pmod{m}$.

По лемме о колоде карт все произведения

$$a \cdot 1,\ a \cdot 2,\ a \cdot 3,\ \ldots,\ a \cdot (m-1)$$

различны, и их ровно m штук. Значит, среди них есть произведение, равное 1. Тем самым мы доказали, что существует такой остаток b, что $ab \equiv 1 \pmod{m}$. Значит, элемент a является делителем единицы, что и требовалось доказать. □

Замечание. В $\mathbb{Z}_m$ любой элемент является либо делителем 0, либо делителем 1. В общем случае это не так. Скажем, в гауссовых числах $\mathbb{Z}[\sqrt{-1}]$ вовсе нет делителей нуля и всего четыре делителя единицы.

4. Решение сравнений

Теорема о делителях единицы является типичным примером *теоремы существования*. Она гарантирует, что если НОД$(a, m) = 1$, то

в $\mathbb{Z}_m$ *существует* a^{-1}. Однако, теорема ничего не говорит о том, как его отыскать. Или более общо, если нам даны $a, b \in \mathbb{Z}_m$ и a — делитель единицы, как *вычислить* $\frac{b}{a} = b \cdot a^{-1} \in \mathbb{Z}_m$? Эта задача равносильна решению следующего сравнения

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

Покажем, как это сделать, на конкретных примерах. Попробуем решить сравнение

$$32x \equiv 7 \pmod{37}.$$

Действовать будем следующим образом: будем *уменьшать модуль коэффициента при* x.

$$\begin{aligned} 32x &\equiv 7 \pmod{37} \quad \Leftrightarrow \\ -5x &\equiv 7 \pmod{37} \quad \mid \cdot 7 \end{aligned}$$

Здесь $7 = \left[\frac{37}{5}\right]$, т.е. мы выбираем такой множитель, чтобы «поближе подобраться» к числу, кратному 37 (в данном случае к -37).

$$\begin{aligned} -35x &\equiv 12 \pmod{37} \quad \Leftrightarrow \\ 2x &\equiv 12 \pmod{37} \quad \Leftrightarrow \\ x &\equiv 6 \pmod{37}. \end{aligned}$$

Таким образом, можно утверждать, что

$$\frac{7}{32} \equiv 6 \pmod{37}.$$

Впрочем, вычисления не всегда можно провести так легко. Как вы могли заметить, на последнем шаге нам удалось поделить обе части сравнения на 2. Мы имели на это право, поскольку $\text{НОД}(2, 37) = 1$. Вообще говоря, так будет не всегда. Иногда приходится *делить сравнение на делители нуля.* Поясним, о чем идет речь на конкретном примере, и решим сравнение

$$13x \equiv 1 \pmod{48}.$$

Имеем

$$13x \equiv 1 \pmod{48} \quad | \cdot 4$$

Здесь $4 = \left[\frac{48}{13}\right] + 1$, т.к. $13 \cdot 4$ «ближе» к 48, чем $13 \cdot 3$.

$$\begin{aligned} &52x \equiv 4 \pmod{48} \quad \Leftrightarrow \\ &4x \equiv 4 \pmod{48}. \end{aligned}$$

Поскольку НОД$(4, 48) > 1$, мы не можем разделить обе части сравнения на 4! Что же делать? Воспользуемся свойством сравнений 6 и сократим на 4 не только обе части сравнения, но и модуль. Получаем

$$x \equiv 1 \pmod{12} \quad \Leftrightarrow \quad x = 1 + 12n,\ n \in \mathbb{Z}$$

Получили ли мы решение? Конечно, нет! Например, при $n = 0$ получаем $x = 1$. Но

$$13 \not\equiv 1 \pmod{48}.$$

Это означает, что мы получили слишком много решений, и некоторые нужно отбросить. Для этого проверим найденные x, подставив их в исходное сравнение:

$$\begin{aligned} &13(1 + 12n) \equiv 1 \pmod{48} \quad \Leftrightarrow \\ &13 \cdot 12n + 13 \equiv 1 \pmod{48} \quad \Leftrightarrow \\ &13 \cdot 12n \equiv -12 \pmod{48} \quad \text{Снова свойство 6!} \quad \Leftrightarrow \\ &13n \equiv -1 \pmod{4} \quad \Leftrightarrow \\ &n \equiv -1 \pmod{4} \quad \Leftrightarrow \quad n = -1 + 4k,\ k \in \mathbb{Z} \end{aligned}$$

Окончательно имеем

$$\begin{aligned} &x = 1 + 12(-1 + 4k) = -11 + 48k \quad \Leftrightarrow \\ &x \equiv -11 \equiv 37 \pmod{48}. \end{aligned}$$

Тот факт, что в процессе вычислений мы смогли вернуться к исходному модулю, кажется очень удачным совпадением. Однако, в данном примере это следует из того, что $x \equiv 1 \pmod{12}$ и очевидно, так будет происходить всегда.

Суммируя наши наблюдения, отметим, что множество остатков $\mathbb{Z}_m$ *похоже* на множество целых чисел $\mathbb{Z}$: элементы этих множеств можно складывать, вычитать, умножать, а вот делить можно не всегда. У нас уже накопилась целая коллекия таких множеств. Среди них множества многочленов $\mathbb{R}[x]$ и $\mathbb{Q}[x]$, множество $\mathbb{Z}[\sqrt{-1}]$ гауссовых чисел и другие $\mathbb{Z}[\sqrt{-k}]$.

Как мы уже отмечали в предисловии, ***похожесть этих множеств фиксируется определением*: они называются *кольцами***.

Если число $m = p$ — простое, то во множестве $\mathbb{Z}_p$ можно еще и делить (на ненулевые элементы). Этим множество $\mathbb{Z}_p$ похоже на множества рациональных чисел $\mathbb{Q}$ и вещественных чисел $\mathbb{R}$.

Эта похожесть также фиксируется определением*: эти множества называют *полями.

В будущем вы познакомитесь с множеством других колец и полей, которые, на первый взгляд, будут казаться совершенно непохожими на разбираемые нами примеры. Однако их похожесть, зафиксированная в определении, позволит использовать сформированную ранее интуицию в совсем новых областях.

5. Китайская теорема об остатках

Несмотря на похожесть всех $\mathbb{Z}_m$, на уровне конкретных вычислений есть и различие: так $2 + 1 \equiv 0$ в $\mathbb{Z}_3$, а в $\mathbb{Z}_4$ имеем $2 + 1 \equiv 3$. Если 2 является делителем нуля в $\mathbb{Z}_6$, то в $\mathbb{Z}_7$ 2 является делителем единицы. При решении сравнений мы тщательно следили за тем, чтобы в конце концов оказаться в исходном модуле, что естественно, так как результат зависит от модуля кардинальным образом. Например, выпишем несколько решений сравнения $3x \equiv 1$ для разных модулей:

$$\frac{1}{3} \equiv 2 \pmod 5;$$
$$\frac{1}{3} \equiv 5 \pmod 7;$$
$$\frac{1}{3} \equiv 3 \pmod 8;$$
$$\frac{1}{3} \equiv 7 \pmod{10}.$$

И все-таки некоторая связь между кольцами $\mathbb{Z}_m$ для различных m имеет место. Рассмотрим, к примеру, $\mathbb{Z}_2, \mathbb{Z}_3$ и $\mathbb{Z}_6$ и вернемся к исходному определению этих колец как остатков от деления на 2, 3 и 6 соответственно. Зная остаток от деления числа на 6, можно очевидным образом определить остатки от деления этого же числа на 2 и 3 ($2 \cdot 3 = 6$). А можно ли, наоборот, зная остатки от деления на 2 и 3, определить остаток от деления на 6? Оказывается, можно, причем единственным образом!

Упоминания о соответсвующей теореме впервые было найдено в трактате китайского математика Сунь Цзы «Сунь Цзы Суань Цзин», предположительно датируемом III веком н.э. и затем в книге Цинь Цзюшао «Математические рассуждения в 9 главах» датируемой 1247 годом, где было приведено точное доказательство. Именно поэтому обсуждаемый результат принято называть китайской теоремой об остатках.

Теорема (Китайская теорема об остатках)**.** *1. Пусть даны n попарно взаимно простых чисел $m_1, m_2, \ldots, m_n$. Тогда система*

$$\begin{cases} x \equiv r_1 \pmod{m_1} \\ x \equiv r_2 \pmod{m_2} \\ \quad \vdots \\ x \equiv r_n \pmod{m_n} \end{cases}$$

имеет решение x_0.

2. Если x_0 — решение, то все решения имеют вид

$$x_0 + (m_1 \cdot m_2 \cdot \ldots \cdot m_n) \cdot k,$$

где $k \in \mathbb{Z}$.

Доказательство. Приведем *конструктивное* доказательство.
1. Будем искать решение в следующем виде:

$$x = x_1 \cdot (\widehat{m}_1 \cdot m_2 \cdot \ldots \cdot m_n) + x_2 \cdot (m_1 \cdot \widehat{m}_2 \cdot \ldots \cdot m_n) + \ldots + \\ + x_k \cdot (m_1 \cdot \ldots \cdot \widehat{m}_k \cdot \ldots \cdot m_n) + \ldots + x_n \cdot (m_1 \cdot m_2 \cdot \ldots \cdot \widehat{m}_n),$$

где $x_1, \ldots, x_n$ — неизвестные величины, а $\widehat{m}_i$ означает, что в произведении слагаемое m_i *пропущено*. Переходя поочередно к сравнениям по модулям $m_1, \ldots, m_n$, получаем систему

$$\begin{cases} x_1 \cdot (\widehat{m}_1 \cdot m_2 \cdot \ldots \cdot m_n) \equiv r_1 \pmod{m_1} \\ x_2 \cdot (m_1 \cdot \widehat{m}_2 \cdot \ldots \cdot m_n) \equiv r_2 \pmod{m_2} \\ \quad \vdots \\ x_k \cdot (m_1 \cdot \ldots \cdot \widehat{m}_k \cdot \ldots \cdot m_n) \equiv r_k \pmod{m_k} \\ \quad \vdots \\ x_n \cdot (m_1 \cdot m_2 \cdot \ldots \cdot \widehat{m}_n) \equiv r_n \pmod{m_n} \end{cases}$$

Замечание. Исходняя система *разделилась* на n независимых сравнений.

Поскольку $m_1, m_2, \ldots, m_n$ попарно взаимно просты, то по теореме о делителях единицы (которая здесь играет ключевую роль) данные сравнения имеют решение:

$$\begin{cases} x_1 \equiv l_1 \pmod{m_1} \\ x_2 \equiv l_2 \pmod{m_2} \\ \quad \vdots \\ x_k \equiv l_k \pmod{m_k} \\ \quad \vdots \\ x_n \equiv l_n \pmod{m_n} \end{cases}$$

Таким образом, мы получили решение исходной системы:

$$x_0 = \sum_{k=1}^{n} l_k \cdot (m_1 \cdot \ldots \cdot \widehat{m}_k \cdot \ldots \cdot m_n).$$

2. Пусть y_0 — другое решение системы. Имеем

$$\begin{aligned} &x_0 \equiv y_0 \equiv r_i \pmod{m_i} \quad \text{при } i = 1, 2, \ldots n \Rightarrow \\ &x_0 - y_0 \equiv 0 \pmod{m_i}) \quad \text{при } i = 1, 2, \ldots n \Rightarrow \\ &x_0 - y_0 \equiv 0 \pmod{m_1 \cdot m_2 \cdot \ldots \cdot m_n}, \end{aligned}$$

поскольку в силу попарной взаимной простоты чисел $m_1, m_2, \ldots, m_n$ разность $x_0 - y_0$ будет делиться на их произведение, если она делится на каждое из них. Следовательно $y_0 = x_0 + (m_1 \cdot m_2 \cdot \ldots \cdot m_n) \cdot k$. □

Замечание. Условие попарной взаимной простоты чисел $m_1, m_2, \ldots, m_n$ является необходимым для доказательства теоремы и не может быть опущено. Например, система

$$\begin{cases} x \equiv 1 \pmod{2} \\ x \equiv 2 \pmod{4} \end{cases}$$

вообще не имеет решений по очевидным соображениям четности.

Китайская теорема об остатках дает возможность установить *взаимно однозначное* соответствие (т.е. изоморфизм) между элементами кольца $\mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n}$ и наборами остатков $(r_1, r_2, \ldots, r_n)$, где $r_i \in \mathbb{Z}_{m_i}$. Множество таких наборов обозначается следующим образом:

$$\mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n}$$

и называется *прямой суммой* $\mathbb{Z}_{m_1}, \mathbb{Z}_{m_2}, \ldots, \mathbb{Z}_{m_n}$. Действительно, в силу теоремы любой набор $r \in \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n}$ однозначным образом задает $x \in \mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n}$ как решение соответствующей системы сравнений. Наоборот, любой $x \in \mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n}$ имеет однозначно определенный набор остатков $r \in \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n}$ при делении на попарно взаимно простые $m_1, m_2, \ldots, m_n$.

Более того, если

$$x \longleftrightarrow r = (r_1, r_2, \ldots, r_n),$$
$$y \longleftrightarrow s = (s_1, s_2, \ldots, s_n),$$

то сумме $x + y \in \mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n}$ соответствует сумма

$$r + s = (r_1 + s_1, r_2 + s_2, \ldots, r_n + s_n) \in \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n},$$

а произведению $x \cdot y \in \mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n}$ — произведение

$$r \cdot s = (r_1 \cdot s_1, r_2 \cdot s_2, \ldots, r_n \cdot s_n) \in \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n}.$$

В таком случае говорят, что между $\mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n}$ и $\mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n}$ установлен *естественный изоморфизм*, что обозначается так:

$$\mathbb{Z}_{m_1 \cdot m_2 \cdot \ldots \cdot m_n} \simeq \mathbb{Z}_{m_1} \oplus \mathbb{Z}_{m_2} \oplus \ldots \oplus \mathbb{Z}_{m_n}.$$

Замечание. Обратите внимание, что при отсутствии взаимной простоты чисел $m_1, m_2, \ldots, m_n$ отсутствует указанный нами изоморфизм. Рассмотрим, например, кольца $\mathbb{Z}_4$ и $\mathbb{Z}_2 \oplus \mathbb{Z}_2$. И первое, и второе состоят из четырех элементов, но они *существенно различны.* Например, во втором сумма любого элемента с самим собой равна $(0, 0)$, т.е. нулю, в первом это, очевидно, неверно.

Замечание. В этой главе мы заложили основы теории сравнений. В следующих главах мы разовьем эту теорию и обсудим множество важных и интересных результатов, как классических (доказанных такими великими математиками как П. Ферма, Л. Эйлер, К. Гаусс. . .), так и сравнительно недавних (полученных, в частности, нашими учениками). Однако перед тем как двигаться дальше, вспомним еще раз аналогию между кольцами целых чисел $\mathbb{Z}$ и многочленов $\mathbb{R}[x]$, о которой мы много раз говорили.

Кольцо остатков $\mathbb{Z}_m$ — это по определению множество остатков от деления целых чисел на фиксированное целое ненулевое число m с естественными операциями сложения и умножения. Можно спросить, что получится, если вместо целых чисел мы возьмем многочлены?

Иначе говоря, что собой представляет множество остатков от деления всевозможных многочленов на фиксированный многочлен f (это множество обозначается через $\mathbb{R}[x]/(f)$)? Как выглядят в нем операции сложения и умножения? Что представляют собой делители нуля и делители единицы?...

Ответы на все эти вопросы оказываются неожиданно тесно связанными с самыми разными областями математики (в частности, с теорией полей и алгебраической геометрией, а также с уже знакомыми нам кольцами $\mathbb{Z}[\sqrt{-k}]$). Позднее в разделе XI.6 мы дадим на них ответы.

Глава V

Диофантовы уравнения

Кольцо остатков $\mathbb{Z}_m$ и теория сравнений, как мы уже отмечали, играют одну из ключевых ролей в теории чисел. Среди их многочисленных приложений методы решения уравнений в целых числах, к которым мы теперь переходим.

1. Линейные диофантовы уравнения

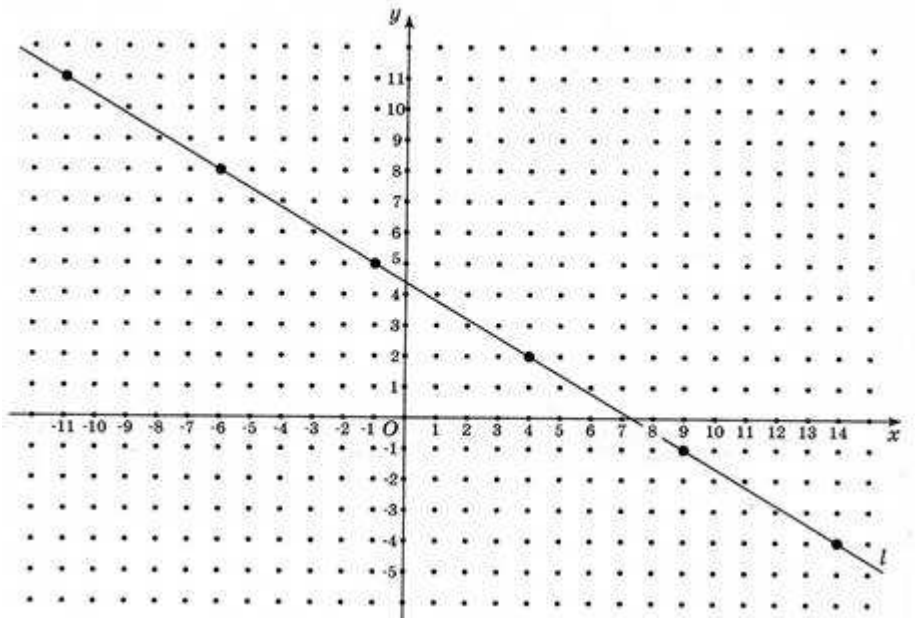
Эффективным методом решения некоторых задач является составление и решение соответствующего уравнения или системы уравнений. В некоторых случаях по самому смыслу задачи приходится рассматривать исключительно целые числа. Таким образом, возникает необходимость исследовать целочисленные уравнения, т.е. уравнения, в которых и коэффициенты, и переменные принимают только целые значения. Такие уравнения называются диофантовыми в честь уже упоминавшегося нами древнегреческого ученого Диофанта Александрийского, жившего в III веке н.э. Именно Диофант в своем труде под названием «Арифметика» заложил основы изучения таких уравнений. Более того, его работы, переведенные на латинский, послужили стимулом развития теории чисел в Новое время.

Неявно мы уже сталкивались с такими уравнениями. Действительно, рассмотрим, например, сравнение $3x \equiv 2 \pmod 5$. Оно означает, что $3x = 2 + 5y$, где $y \in \mathbb{Z}$. Т.е. $3x - 5y = 2$, где $x, y \in \mathbb{Z}$. Таким образом, решение сравнения тесно связано с решением уравнения.

Определение. Уравнение вида $ax+by = c$, где $a, b, c \in \mathbb{Z}$ называется *линейным диофантовым.*

Решением линейного диофантова уравнения называется пара чисел $x, y \in \mathbb{Z}$, которая удовлетворяет этому уравнению.

Рассмотрим линейное диофантово уравнение с геометрической точки зрения. Уравнение $ax + bx = c$ задает прямую на плоскости. Задача решения диофантова уравнения равносильна задаче нахождения всех точек с целыми координатами на этой прямой.



Прямая $3x + 5y = 22$
с отмеченными на ней целыми точками

Как решать такие уравнения? С помощью сравнений! Наиболее удобен следующий алгоритм.

1. Сначала надо сократить уравнение на НОД(a, b). Если $c \not\vdots$ НОД(a, b), то решений нет.
   Действительно, пусть $d =$ НОД(a, b). Тогда имеем

   $$ax + by = c \quad \Leftrightarrow \quad d(a'x + b'y) = c.$$

   Если бы уравнение имело решение, то можно было бы утверждать, что $c \vdots d$. Противоречие.
2. После сокращения можно считать, что НОД$(a, b) = 1$. Тогда надо перейти к сравнению. По какому модулю?

   a. Если a — простое и не слишком больше b, то надо сравнивать по модулю a.

б. Если a и b — большие и неясно, простые или нет, лучше сравнивать по меньшему модулю.

3. Сравнение необходимо решить и подставить решение в исходное уравнение, чтобы найти вторую переменную.

Покажем, как это сделать, на конкретном примере.

$$183x - 144y = 6 \quad \Leftrightarrow \quad 61x - 48y = 1$$

Переходим к сравнению по модулю 48:

$$13x \equiv 1 \pmod{48}.$$

Линейные сравнения мы уже научились решать в главе IV. Получаем, что

$$x = 1 + 12(-1 + 4k) = -11 + 48k, \text{ где } k \in \mathbb{Z}.$$

Теперь подставим полученный результат в исходное уравнение и найдем y:

$$y = \frac{61x - 1}{48} = \frac{61(-11 + 48k) - 1}{48} = -14 + 61k.$$

Мы видим, что линейное диофантово уравнение либо не имеет решений вовсе, либо имеет бесконечно много решений (что не столь неожиданно, ведь мы рассматриваем *одно* уравнение с *двумя* переменными). Возвращаясь к геометрической интерпретации, можно сказать, что либо прямая на координатной плоскости не проходит ни через одну целую точку, либо на ней бесконечное количество таких точек.

Зафиксируем наши наблюдения.

Теорема. *1. Уравнение $ax + by = c$, где $a, b, c, \in \mathbb{Z}$, имеет решение в целых числах тогда и только тогда, когда $c \vdots НОД(a, b)$.*

2. Если $НОД(a, b) = 1$ и (x_0, y_0) — решение, то все решения задаются формулами:

$$\begin{cases} x = x_0 + bn \\ y = y_0 - an, \end{cases}$$

где $n \in \mathbb{Z}$.

Доказательство. 1. Пусть уравнение имеет решение (x_0, y_0) и пусть $d = \text{НОД}(a, b)$. Тогда имеем $a = a_1 \cdot d,\ b = b_1 \cdot d$ и

$$ax_0 + by_0 = c \ \Leftrightarrow\ d \cdot (a_1 x_0 + b_1 y_0) = c \ \Rightarrow\ c \vdots d.$$

Пусть теперь $c \vdots d = (a, b)$. Тогда $c = c_1 \cdot d$ и уравнение принимает вид

$$ax + by = c \ \Leftrightarrow\ a_1 x + b_1 y = c_1, \quad \text{где } (a_1, b_1) = 1.$$

Из теоремы о делителях единицы следует, что сравнение $a_1 x \equiv c_1 \pmod{b_1}$ имеет *единственное* решение $x_0 \in \mathbb{Z}_{b_1}$.
Значит, все решения сравнения имеют вид $x = x_0 + b_1 n$, где $n \in \mathbb{Z}$. Подставляя x в исходное диофантово уравнение, находим

$$y = \frac{c_1 - a_1(x_0 + b_1 n)}{b_1} = \frac{c_1 - a_1 x_0}{b_1} - a_1 n = y_0 - a_1 n,$$

где $y_0 = \frac{c_1 - a_1 x_0}{b_1} \in \mathbb{Z}$. Намоним, что y_0 является целым, поскольку $a_1 x_0 \equiv c_1 \pmod{b_1}$ и следовательно $(c_1 - a_1 x_0) \vdots b_1$.

2. Этот пункт следует из доказательства пункта 1. □

Доказанная теорема подсказывает следующий способ решения линейных диофантовых уравнений. Если коэффициенты не очень велики, то *частное решение* (x_0, y_0) можно попробовать угадать. Зная (x_0, y_0), общее решение выписывается мгновенно.

Отметим, что из доказанной нами теоремы также вытекает уже знакомый (см. главу II) нам факт.

Следствие (представление НОД)**.** *Для любых $a, b \in \mathbb{Z}$ найдутся такие $x, y \in \mathbb{Z}$, что $ax + by = НОД(a, b)$.*

Оказывается, что именно это свойство является ключевым при доказательстве основной теоремы арифметики. А именно, теорема утверждает два факта: *существование* разложения на простые множители и его *единственность* (с точностью до перестановки и умножения на обратимые элементы). Чтобы доказать существование, необходимо использовать понятие нормы, которое нам уже

встречалось на примере гауссовых чисел и других колец вида $\mathbb{Z}[\sqrt{-k}]$. Возможность представить НОД двух элементов в виде $ax + by = \text{НОД}(a, b)$ позволяет доказать единственность разложения на простые множители.

2. Нелинейные уравнения

Исследовав случай линейных уравнений, возникает естественный вопрос, а что можно сказать про уравнений степени 2 и выше? Существует ли общий метод решения таких уравнений? 8 августа 1900 года на II международном конгрессе математиков Давид Гильберт представил список из 23 проблем, решение которых, по его мнению, сыграло бы ключевую роль в дальнейшем развитии математики. С тех пор эти задачи известны как «проблемы Гильберта». В частности, 10-ая проблема была сформулирована следующим образом: существует ли алгоритм, который позволяет узнать, разрешимо ли диофантово уравнение вида

$$P(x_1, x_2, \dots x_m) = 0,$$

где P — многочлен с целыми коэффициентами? В случае линейного многочлена от двух переменных

$$P(x_1, x_2) = ax_1 + bx_2 + c = 0$$

получается в точности разобранный нами случай. Однако, в общем случае ответ оказался отрицательным! Этот ответ был получен в 1970 году, и одним из ключевых продвижений были результаты российского математика Юрия Матиясевича.

Несмотря на отсутствие общего алгоритма, существует несколько приемов, позволяющих решать некоторые диофантовы уравнения высших степеней. Рассмотрим их на конкретных примерах.

Разложение на множители

Идея этого приема чрезвычайно проста — разложить одну из частей уравнения на множители и ограничить перебор. С этой идеей мы уже сталкивались, когда искали простые и составные числа в $\mathbb{Z}[\sqrt{-k}]$. А именно, мы получали уравнение вида

$$(a^2 + 3b^2) \cdot (c^2 + 3d^2) = 4,$$

которое решали, используя тот факт, что левая часть уравнения разложена на множители.

В качестве еще одного примера решим уравнение

$$x^2 - 3xy + 2y^2 = 7.$$

Вообще говоря, неясно, есть ли вообще решения у этого уравнения, не говоря уже об их поиске. Внимательный взгляд позволяет заметить, что левую часть можно разложить на множители

$$x^2 - 3xy + 2y^2 = 7 \quad \Leftrightarrow \quad (x - y) \cdot (x - 2y) = 7.$$

Поскольку разложить число 7 на два *целых* сомножителя можно лишь двумя способами

$$7 = 1 \cdot 7 = (-1) \cdot (-7),$$

то у исходного уравнения существуют целые решения, если они удовлетворяют одной из следующих систем уравнений

$$\begin{cases} x - y = 1 \\ x - 2y = 7, \end{cases} \quad \begin{cases} x - y = 7 \\ x - 2y = 1, \end{cases} \quad \begin{cases} x - y = -1 \\ x - 2y = -7, \end{cases} \quad \begin{cases} x - y = -7 \\ x - 2y = -1. \end{cases}$$

Прямая проверка показывает, что каждая из систем имеет решение в целых числах. Таким образом, исходное уравнение имеет следующие решения: $(-5, -6), (13, 6), (5, 6), (-13, -6)$.

Отметим, что возможность разложения на множители накладывает жесткие условия на левую часть уравнения, поэтому этот прием работает далеко не всегда.

Квадраты и кубы по модулю

Чтобы в полной мере осознать следующий метод решения нелинейных диофантовых уравнений, необходимо сделать несколько предварительных наблюдений.

Рассмотрим произвольное натуральное число a. Какие остатки оно может давать по модулю 3? Конечно же, только три: 0, 1 и 2. Давайте будем использовать отрицательные остатки, тогда вместо 2 можно написать -1. Итак, $a \equiv 0, 1, -1 \pmod 3$.

Что произойдет, если мы возведем число a в квадрат? Согласно свойствам сравнений, остатки тоже нужно возвести в квадрат. Давайте составим таблицу остатков:

a	0	1	-1
a^2	0	1	1

Таблица остатков по модулю 3.

Таким образом, мы доказали следующее

Утверждение. *Квадрат целого числа не может давать остатка* -1 *по модулю 3.*

Совершенно аналогично можно доказать

Утверждение. *Квадрат целого числа может давать только остатки 0 и* 1 *по модулю 4.*

a	0	1	2	-1
a^2	0	1	0	1

Таблица остатков по модулю 4.

Таким образом, очень полезным оказывается сравнение квадратов целых чисел по модулям 3 и 4. Однако иногда требуется сравнение и по другим модулям.

Например,какие остатки может давать квадрат целого числа по модулю 5?

a	0	1	2	-2	-1
a^2	0	1	-1	-1	1

Таблица остатков по модулю 5.

Замечание. Отметим следующую удивительную вещь:

$$2^2 \equiv -1 \pmod 5.$$

Как мы видели ранее, сравнение $\equiv$ очень похоже на равенство $=$. Но не существует такого (вещественного) числа, квадрат которого был бы равен -1! А по модулю 5 такое число существует! Легко видеть, что такое число не единственно:

$$3^2 \equiv (-2)^2 \equiv -1 \pmod 5.$$

Таким образом, имеем:

$$\sqrt{-1} \equiv \pm 2 \pmod 5$$

В заключение отметим, что и для кубов существуют «удачные» модули, по которым очень полезно сравнивать. Например, посмотрим, какие остатки дает куб по модулю 7.

a	0	1	2	3	4	5	6
a^3	0	1	1	-1	1	-1	-1

Таблица остатков по модулю 7.

И по модулю 9.

a	0	1	2	3	4	-4	-3	-2	-1
a^3	0	1	-1	0	1	-1	0	1	-1

Таблица остатков по модулю 9.

Итак, теперь мы готовы перейти к следующему приему.

Сравнение левой и правой частей уравнения по модулю

Данный прием имеет более широкую (хотя все равно ограниченную) область применения, чем простое разложение на множители. В качестве примера рассмотрим уравнение

$$x^2 - 4xy + y^2 = 3.$$

Разложить на множители левую часть нам не удастся. Перейдем к сравнению по модулю 4 (таким образом мы «убиваем» слагаемое $4xy$)

$$x^2 - 4xy + y^2 = 3 \quad \Rightarrow \quad x^2 + y^2 \equiv 3 \pmod 4.$$

Поскульку квадраты целых чисел могут давать только остатки 0 или 1 при делении на 4, данное сравнение не имеет решений, а значит, не имеет решений и исходное уравнение.

Замечание. Абсолютно аналогичное рассуждение показывает, что натуральное число вида $4k + 3$ не представимо в виде суммы двух квадратов. О том, какие числа представимы в таком виде, мы подробно поговорим в VIII.1.

Комбинирование этих приемов позволяет решать уравнения, которые содержат другие функции, помимо многочленов. Например, найдем решения уравнения

$$3^x + 1 = 2^y.$$

Очевидно, что при $x, y < 0$ оно не имеет решений. Переходя к сравнению по модулю 3, получаем

$$2^y \equiv 1 \pmod 3 \quad \Leftrightarrow \quad (-1)^y \equiv 1 \pmod 3,$$

если $x > 0$. Откуда следует, что $y = 2k$. Подставляя в исходное уравнение, имеем

$$3^x = 2^{2k} - 1 = (2^k - 1) \cdot (2^k + 1).$$

Откуда

$$\begin{cases} 2^k - 1 = 3^m \\ 2^k + 1 = 3^n \\ m + n = x. \end{cases}$$

Но единственные степени тройки, которые отличаются на 2, это 1 и 3. Таким образом, $2^k - 1 = 1 = 3^0$, и мы получаем решение исходного уравнения $(1, 2)$.

Осталось проверить единственный случай $x = 0$. Имеем $2^y = 1$, откуда получаем еще одно решение $(0, 1)$. Итак, уравнение $3^x + 1 = 2^y$ имеет следующие решения: $(0, 1), (1, 2)$.

Во всех рассмотренных примерах уравнение имело лишь конечное число решений, что неудивительно, ведь мы сводили задачу к конечному перебору. Отсутствие общего алгоритма решения приводит к тому, что каждый случай приходится рассматривать отдельно и совсем необязательно удастся свести задачу к конечному перебору. Более того, заранее нельзя выяснить, сведется ли задача к такому перебору, и вообще насколько трудной она может оказаться. Например, уравнения вида $x^n + y^n = z^n$ при $n > 2$ не имеют решений в целых числах (кроме тривиальных вида $(1, 0, 1)$), что является утверждением «великой теоремы Ферма». Как мы уже упоминали, потребовалось около 350 лет, чтобы найти доказательство. Случай $n = 2$ отличается от указанных выше, и теперь мы переходим к его детальному изучению.

3. Пифагоровы тройки и рациональные кривые

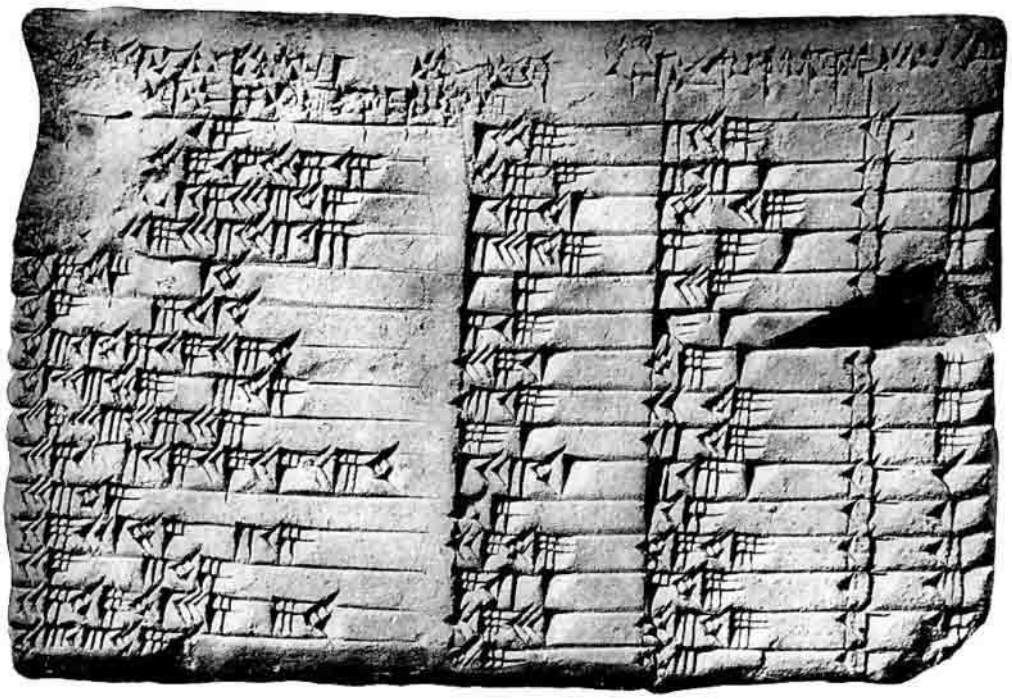
Определение. *Пифагоровой* называется тройка целых чисел (a, b, c), которая удовлетворяет уравнению

$$a^2 + b^2 = c^2.$$

Замечание. Очевидно, что если (a, b, c) — пифагорова тройка, то (b, a, c) — также пифагорова тройка. Поэтому будем искать решения с точностью до перестановки первых двух чисел.

Пифагоровыми эти тройки называются потому, что по теореме Пифагора длины катетов прямоугольного треугольника a и b связаны с длиной гипотенузы c в точности соотношением $a^2 + b^2 = c^2$,

и потому, что были известны задолго до Пифагора еще в Древнем Вавилоне.



Вавилонская глиняная табличка, которой почти 4 тысячи лет, содержащая список из пятнадцати пифагоровых троек.

Полное описание пифагоровых троек можно отыскать уже в работах древнегреческих ученых — Евклида и Диофанта.

Первым делом заметим, что если (a, b, c) — пифагорова тройка, то (ka, kb, kc), где $k \in \mathbb{Z}$, также пифагорова тройка, поскольку

$$a^2 + b^2 = c^2 \quad \Leftrightarrow \quad (ka)^2 + (kb)^2 = (kc)^2.$$

Поэтому будем искать тройки (a, b, c), которые состоят из *попарно взаимно простых* чисел. Такие пифагоровы тройки называются *примитивными*. Именно такие тройки мы теперь опишем.

Пользуясь простыми соображениями теории сравнений, можно показать, что из трех чисел a, b и c два являются нечетными, а одно — четным, причем c будет обязательно нечетным. Будем считать, что a — нечетное, а b — четное. Имеем

$$a^2 + b^2 = c^2 \quad \Leftrightarrow \quad b^2 = c^2 - a^2 = (c - a)(c + a).$$

Числа $c - a$ и $c + a$ являются четными, однако других общих делителей у них быть не может, то есть числа $\frac{c-a}{2}$ и $\frac{c+a}{2}$ являются взаимно

простыми. Действительно, если d — общий делитель $\frac{c-a}{2}$ и $\frac{c+a}{2}$, то

$$\frac{c-a}{2}+\frac{c+a}{2}=c\vdots d,\quad \frac{c+a}{2}-\frac{c-a}{2}=a\vdots d,$$

откуда следует, что $d=1$. Тогда рассмотрим следующее равенство

$$\left(\frac{b}{2}\right)^2=\frac{b^2}{4}=\frac{c-a}{2}\cdot\frac{c+a}{2}.$$

Число $\frac{b}{2}$ является целым, следовательно произведение $\frac{c-a}{2}\cdot\frac{c+a}{2}$ взаимно простых чисел является точным квадратом. Из основной теоремы арифметики (следствие 10) тогда получаем, что $\frac{c-a}{2}$ и $\frac{c+a}{2}$ сами являются точными квадратами

$$\begin{cases}\frac{c-a}{2}=v^2\\ \frac{c+a}{2}=u^2\\ \left(\frac{b}{2}\right)^2=v^2u^2\end{cases}\Leftrightarrow\begin{cases}a=u^2-v^2\\ b=2uv\\ c=u^2+v^2,\end{cases}$$

где u и v — взаимно простые числа. Более того, u и v имеют разную четность, поскольку число $a=u^2-v^2$ является нечетным. Полученная формула доставляет описание всех примитивных троек, а тем самым и всех пифагоровых троек.

Приведенное доказательство, однако, обладает существенными недостатками: оно не вскрывает сути явления и не позволяет решать совершенно аналогичные задачи. Например, имеет ли уравнение

$$2a^2+3b^2=5c^2$$

решения в целых числах? А если и имеет, то можно ли их явно описать? Примененный нами алгебраический подход не дает ответов на эти вопросы. Оказывается, что для полноценного ответа необходима геометрия!

Имеем

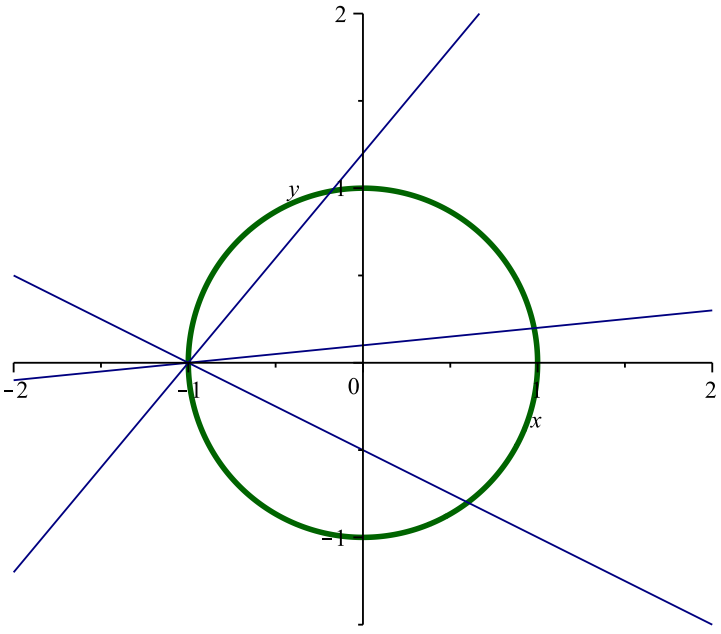
$$a^2+b^2=c^2\quad\Leftrightarrow\quad\left(\frac{a}{c}\right)^2+\left(\frac{b}{c}\right)^2=1.$$

Если тройка (a, b, c) примитивная, то $\left(\frac{a}{c}, \frac{b}{c}\right)$ — пара несократимых дробей, и наоборот. Если ввести обозначения $x = \frac{a}{c}$ и $y = \frac{b}{c}$, то можно утверждать, что задача описания всех примитивных пифагоровых троек равносильна задаче описания всех *рациональных* решений уравнения $x^2 + y^2 = 1$. Последнюю задачу мы сейчас и решим.

Кажется, что мы только усложнили себе работу. Как вообще искать рациональные решения? Удивительно, но это возможно сделать. Причиной тому служат глубокие математические факты, которых мы еще коснемся. А сейчас заметим, что уравнение $x^2+y^2 = 1$ задает единичную окружность на плоскости x, y.

Таким образом, задача отыскания примитивных пифагоровых троек равносильна нахождению всех рациональных точек на окружности!

Как же отыскать все рациональные точки на окружности? Используем следующее соображение. Рассмотрим точку $(-1, 0)$, принадлежащую окружности, и проведем через нее пучок прямых вида $y = t(x + 1)$. Когда меняется параметр t, прямая вращается вокруг точки $(-1, 0)$.



Любая прямая из пучка пересекает окружность еще в одной точке.

Вычислим ее координаты. Для этого необходимо решить следующую систему относительно x и y:

$$\begin{cases} x^2 + y^2 = 1 \\ y = t(x+1) \end{cases} \Leftrightarrow \begin{cases} x^2 + t^2(x+1)^2 = 1 \\ y = t(x+1) \end{cases} \Leftrightarrow$$

$$\begin{cases} (t^2+1)x^2 + 2t^2 x + (t^2 - 1) = 0 \\ y = t(x+1). \end{cases}$$

Первое уравнение — квадратное относительно x. Один из его корней нам известен: это $x = -1$ (так как прямая проведена через точку $(-1, 0)$). По теореме Виета находим второй корень

$$x = \frac{1 - t^2}{1 + t^2},$$

откуда вытекает, что

$$y = t(x+1) = \frac{2t}{1+t^2}.$$

Итак, координаты второй точки перечечения прямой и окружности есть

$$\begin{cases} x = \frac{1-t^2}{1+t^2} \\ y = \frac{2t}{1+t^2}. \end{cases}$$

Когда параметр t принимает всевозможные значения, иными словами, пробегает прямую, точка (x, y) пробегает единичную окружность на плоскости (кроме одной выделенной точки $(-1, 0)$). Обратное соответствие задается формулой

$$t = \frac{y}{x+1}.$$

Полученная параметризация устанавливает *взаимно однозначное* соответствие между точками прямой и окружности с одной выколотой точкой. Более того, мы получили *рациональную* параметризацию окружности, поскольку она задана рациональными функциями (т.е. отношениями двух многочленов) $\frac{1-t^2}{1+t^2}$ и $\frac{2t}{1+t^2}$. Из явного вида этой

параметризации следует, что точка на единичной окружности имеет рациональные координаты тогда и только тогда, когда число t является рациональным!

Подставим в полученную параметризацию $t = \frac{p}{q} \in \mathbb{Q}$

$$\begin{cases} x = \frac{1-t^2}{1+t^2} \\ y = \frac{2t}{1+t^2} \end{cases} \Leftrightarrow \begin{cases} x = \frac{1-(p/q)^2}{1+(p/q)^2} \\ y = \frac{2(p/q)}{1+(p/q)^2} \end{cases} \Leftrightarrow \begin{cases} x = \frac{q^2-p^2}{q^2+p^2} \\ y = \frac{2pq}{q^2+p^2}. \end{cases}$$

Таким образом, в точности все рациональные точки на единичной окружности имеют вид

$$\left(\frac{q^2-p^2}{q^2+p^2}, \frac{2pq}{q^2+p^2} \right),$$

где $\text{НОД}(p,q) = 1$, поскольку дробь $\frac{p}{q}$ несократима. А значит, в точности все примитивные пифагоровы тройки имеют вид

$$(q^2-p^2, 2pq, q^2+p^2).$$

Необходимо только правильно указать необходимые условия на p и q. Дело в том, что, как вы помните, по определению тройка является примитивной, если состоит из попарно взаимно простых чисел. Однако, условие $\text{НОД}(p,q) = 1$ не может этого гарантировать. Пусть, например, $t = \frac{7}{3}$. Такому значению t соответствует точка окружности

$$\left(\frac{q^2-p^2}{q^2+p^2}, \frac{2pq}{q^2+p^2} \right) = \left(\frac{40}{58}, \frac{42}{58} \right) = \left(\frac{20}{29}, \frac{21}{29} \right).$$

Соответствующая тройка

$$(q^2-p^2, 2pq, q^2+p^2) = (40, 42, 58)$$

не является примитивной.

Эту трудность совсем несложно преодолеть. Необходимо потребовать, чтобы числа p и q были не только взаимно простыми, но и имели бы разную четность. В таком случае все примитивные пифагоровы тройки (a,b,c) с точностью до перестановки первых двух членов имеют вид

$$(a,b,c) = (q^2-p^2, 2pq, q^2+p^2),$$

где p и q — взаимно простые целые числа, имеющие разную четность.

Итак, нами доказана

Теорема (О пифагоровых тройках)**.** *Диофантово уравнение*

$$a^2 + b^2 = c^2$$

имеет следующие решения (с точностью до перестановки a и b)

$$a = k \cdot (q^2 - p^2), \quad b = k \cdot (2pq), \quad c = k \cdot (q^2 + p^2),$$

где p и q — взаимно простые целые числа, имеющие разную четность, k — произвольное целое число.

Замечание. Некоторое количество вычислений часто необходимо для получения окончательного ответа, но настоящая идея приведенного доказательства формулы для пифагоровых троек не в них, а в геометрической конструкции, приводящей к рациональности координат точки пересечения окружности с прямой с рациональным наклоном t.

Если мы теперь вернемся к уравнению

$$2a^2 + 3b^2 = 5c^2,$$

то нам потребуется искать рациональные точки на следующей кривой:

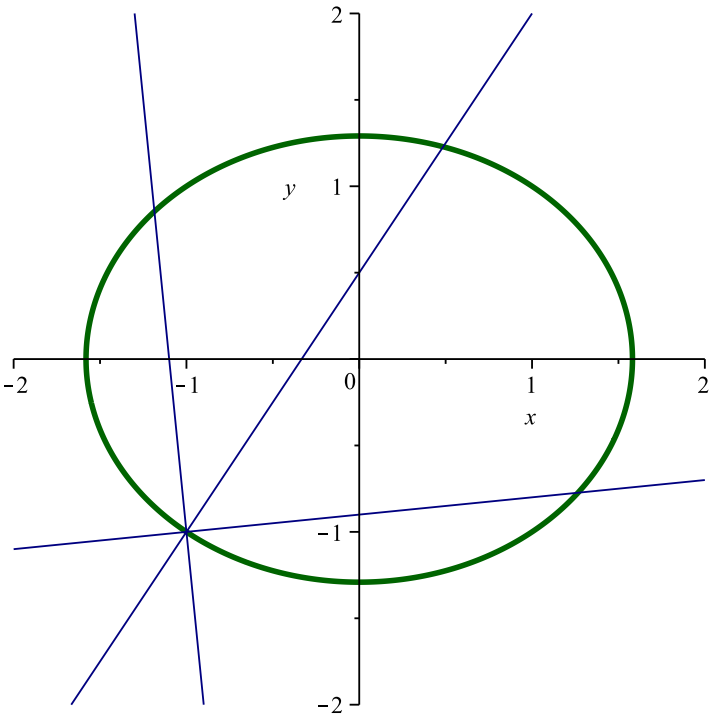
$$2x^2 + 3y^2 = 5,$$

которая является эллипсом.

Легко найти на данном эллипсе рациональную точку. Рассмотрим, например, $(-1, -1)$ и проведем через нее пучок прямых вида $y = t(x + 1) - 1$. Когда меняется параметр t, прямая вращается вокруг точки $(-1, -1)$.

Проделав аналогичную процедуру, мы получаем параметризацию эллипса

$$\begin{cases} x = \frac{2+6t-3t^2}{2+3t^2} \\ y = \frac{4t+6t^2}{2+3t^2}. \end{cases}$$



Значит, все целые решения уравнения $2a^2 + 3b^2 = 5c^2$ задаются формулами

$$\begin{cases} a = k \cdot (2q^2 + 6pq - 3p^2) \\ b = k \cdot (4pq + 6p^2) \\ c = k \cdot (2q^2 + 3p^2). \end{cases}$$

В приведенном решении важную роль играет первый шаг — выбор рациональной точки на кривой, поэтому возникает вопрос: когда на кривой $\alpha x^2 + \beta y^2 = \gamma$ найдется хотя бы одна рациональная точка? Это отнюдь не простой вопрос, ответ на который мы приведем в разделе VIII.2. Оказывается, что ключевую роль здесь играет теория сравнений и квадратичные вычеты, которым и посвящена глава VIII.

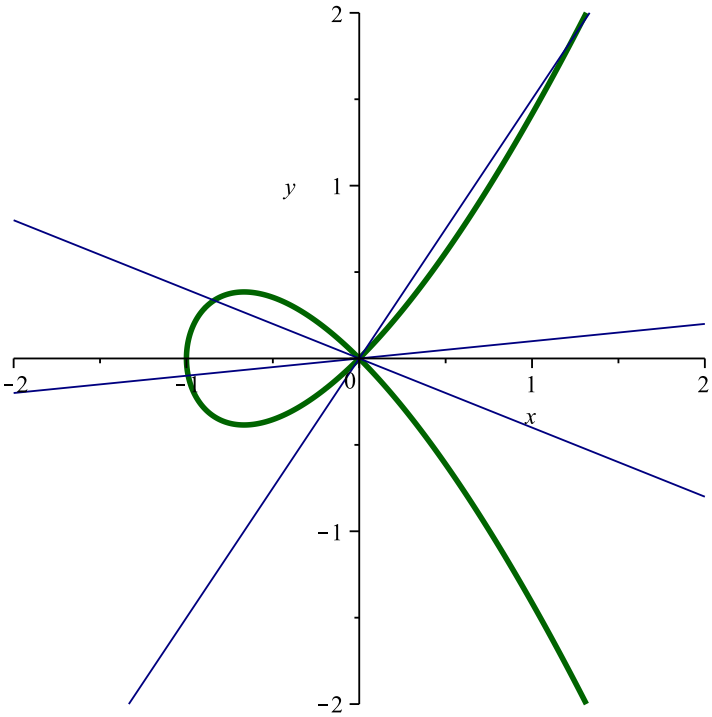
4. Эллиптические кривые и эллиптические интегралы

Идея поиска рациональных точек на кривой, заданной данным диофантовым уравнением, оказывается плодотворной и в других случаях, что мы сейчас продемонстрируем.

Попробуем отыскать решения уравнения

$$y^2 = x^3 + x^2 = x^2(x+1).$$

Будем действовать по аналогии с рассмотренным случаем пифагоровых троек. Нарисуем на плоскости (x, y) кривую, заданную данным уравнением, и проведем через точку $(0, 0)$, принадлежащую кривой, пучок прямых вида $y = tx$.



Любая прямая из пучка пересекает кривую еще в одной точке. Вычислим ее координаты. Для этого необходимо решить следующую систему относительно x и y

$$\begin{cases} y^2 = x^3 + x^2 \\ y = tx \end{cases} \Leftrightarrow \begin{cases} t^2x^2 = x^3 + x^2 \\ y = tx \end{cases} \Leftrightarrow \begin{cases} x^2(x - t^2 + 1) = 0 \\ y = tx. \end{cases}$$

Первое уравнение имеет известный нам корень $x = 0$ (так как прямая проведена через точку $(0, 0)$). Интересующий нас нетривиальный корень имеет вид

$$x = t^2 - 1,$$

откуда вытекает, что

$$y = tx = t(t^2 - 1).$$

Итак, искомые координаты есть

$$\begin{cases} x = t^2 - 1 \\ y = t(t^2 - 1). \end{cases}$$

Мы не будем заниматься описанием всех решений исходного уравнения, как мы сделали в случае пифагоровых троек. Однако, полученная параметризация позволяет мгновенно указать бесконечное множество решений уравнения

$$y^2 = x^3 + x^2$$

— любое $t \in \mathbb{Z}$ дает нам такое решение, что уже очень немало, учитывая сложность рассматриваемых вопросов.

Класс диофантовых уравнений (кривых на плоскости), для которых аналогичные соображения позволяют находит решения не так велик, как может показаться. Стоит лишь немного изменить условие...

Рассмотрим уравнение, очень похожее на предыдущее, а именно

$$y^2 = x^3 - x = x(x-1)(x+1).$$

Попробуем действовать по аналогии. Нарисуем на плоскости (x, y) кривую, заданную данным уравнением, и проведем через точку $(0, 0)$, принадлежащую кривой, пучок прямых вида $y = tx$.

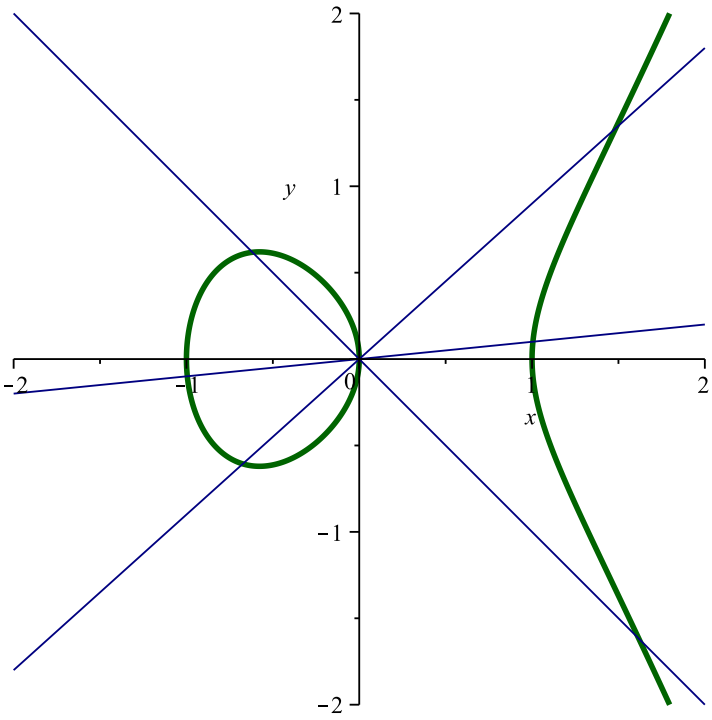
Видно, что проведенные прямые пересекают кривую в двух точках, не считая точку $(0, 0)$. Попробуем найти их координаты, решив следующую систему

$$\begin{cases} y^2 = x^3 - x \\ y = tx \end{cases} \Leftrightarrow \begin{cases} t^2x^2 = x^3 - x \\ y = tx \end{cases} \Leftrightarrow \begin{cases} x(x^2 - t^2x - 1) = 0 \\ y = tx. \end{cases}$$

Попытка же отыскать рациональные по t решения уравнения

$$x^2 - t^2x - 1 = 0,$$

не приводит к успеху. В чем же дело? Возможно, мы неудачно выбрали начальную точку? Или провели не тот пучок прямых?



Замечание. В этом случае в отличие от предыдущих прямые из нашего пучка пересекают кривую не в одной, а в двух точках, не считая точки $(0,0)$. Можно было бы предположить, что именно этот факт является препятствием к рациональной параметризации кривой. Однако это не так. Если бы мы ограничились, например, левым овалом, который прямые из пучка пересекают только в одной точке, не считая $(0,0)$, то все равно не смогли бы получить его рациональную параметризацию.

В действительности, какой бы мы пучок ни провели и какую бы точку ни выбрали, ситуация была бы аналогичной. Невозможно установить взаимно однозначное рациональное соответствие между прямой (с параметром t) и кривой $y^2 = x^3 - x$ (или ее связными компонентами), наподобие тех, что мы установили в предыдущих случаях. Причиной тому служит фундаментальное отличие между прямой и кривой $y^2 = x^3 - x$, о котором мы скажем несколько слов.

В чем состоит истинная причина рациональности окружности? Можно ли как-то по виду уравнения понять, возможна ли рациональная параметризация соответствующей кривой? Оказывается, что да.

Для этого нам необходимо ввести новое понятие.

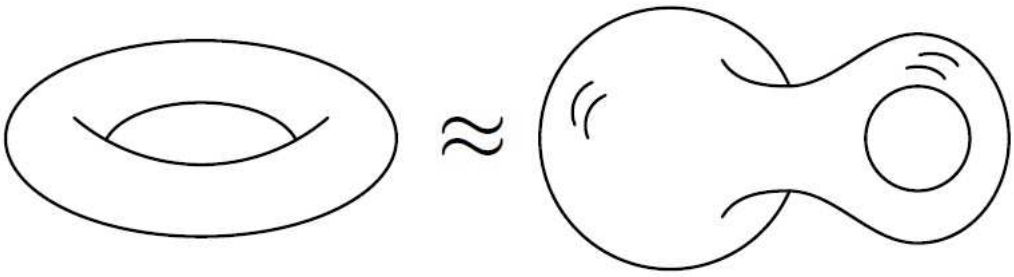
С любым уравнением

$$P(x, y) = 0,$$

где P — многочлен от двух переменных степени d, можно связать не только кривую, но и некоторую поверхность, называемую римановой.

Замечание (для технически подготовленного читателя). Для этого необходимо рассматривать кривую не над полем вещественных, а над полем комплексных чисел. Тогда уравнение будет задавать одномерную комплексную кривую в комплексной проективной плоскости $\mathbb{CP}^2$, т.е. двумерный объект с вещественной точки зрения — искомую поверхность.

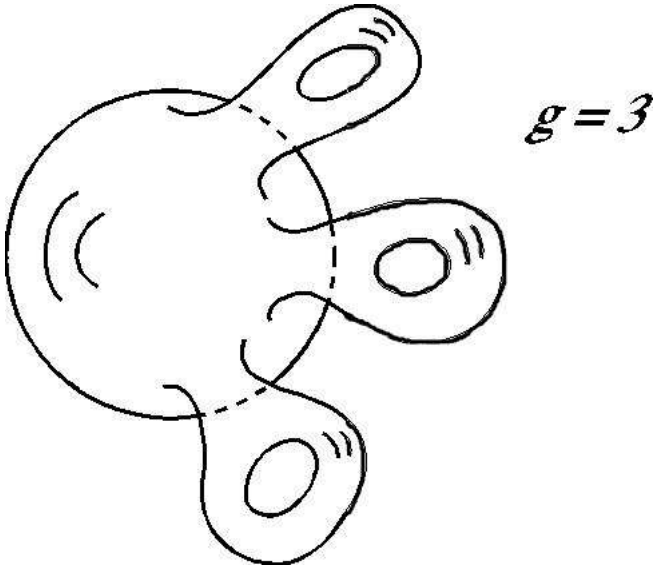
С уравнением $x^2 + y^2 = 1$ и окружностью связана сфера. А кривые, заданные уравнением более высокой степени, приводят к более сложным поверхностям. Например, в случае уравнения $y^2 = x^3 - x$ получится тор или сфера с одной ручкой!



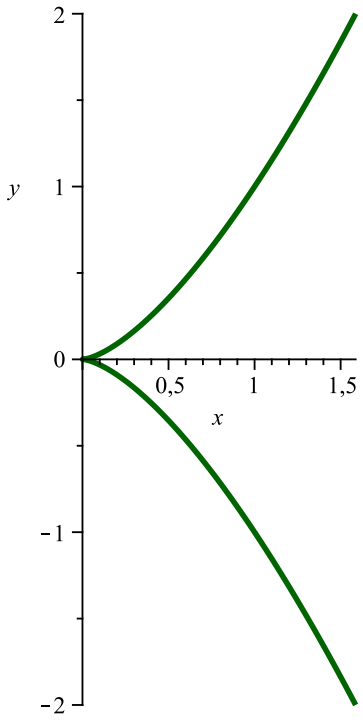
В общем случае с уравнением $P(x, y) = 0$ оказывается связана сфера с некоторым количеством ручек.

Определение. Число ручек g называется *родом поверхности.*

Зная степень d рассматриваемого многочлена, довольно легко найти род соответствующей поверхности. Уравнение $P(x, y) = 0$, где P — многочлен от двух переменных степени d, задает на плоскости (x, y) кривую. Такие кривые называются плоскими алгебраическими кривыми. Будем также считать, что кривая не имеет особенностей, т.е. «плохих» точек наподобие точек самопересечения (как у кривой



$y^2 = x^3 + x^2$ в точке $(0, 0)$) или «клювов» (как у кривой $y^2 = x^3$ в точке $(0, 0)$).



Кривая $y^2 = x^3$ с особенностью в виде клюва.

Тогда род соответствующей поверхности можно отыскать по простой формуле Римана-Гурвица

$$g = \frac{(d-1)(d-2)}{2}.$$

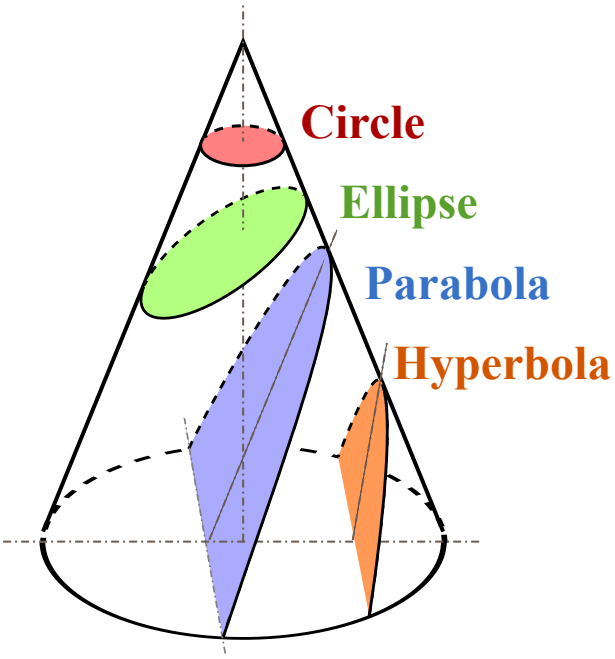
Оказывается, что только кривые, которым соответствуют поверх-

ности рода 0 (сферы), допускают возможность рациональной параметризации, поэтому они называются *рациональными кривыми*. Если поверхность имеет род 1 (тор), то соответствующие кривые представляют собой замечательный математический объект, встречающийся в различных областях математики. Они называются *эллиптическими кривыми*.

Эллиптические кривые без особенностей не могут быть рационально параметризованы, поэтому вопрос наличия на них рациональных точек представляет большую сложность. Именно наличие особенностей является причиной существования рациональной параметризации у кривых $y^2 = x^3 + x^2$ и $y^2 = x^3$.

Замечание. Несмотря на свое название, эллиптические кривые имеют к эллипсам весьма отдаленное отношение.

Еще в III в. до н.э. древнегреческий ученый Аполлоний занимался исследованием конических сечений, которые представляют из себя эллипс, параболу или гиперболу в зависимости от угла наклона секущей плоскости.



Эллипс, парабола и гипербола как сечения кругового конуса.

Несмотря на глубокие результаты, полученные Аполлонием, некоторые вопросы остались без ответа. Одним из таких вопросов было

вычисление длины дуги эллипса (для частного случая эллипса — окружности ответ был хорошо известен древним грекам).

Спустя почти две тысячи лет в 1609 году Иоганн Кеплер публикует знаменитые законы движения планет, первый из которых гласит: «Каждая планета Солнечной системы обращается по эллипсу, в одном из фокусов которого находится Солнце». Таким образом, задача вычисления длины дуги эллипса приобретает важнейшее значение — подсчет длины орбиты планет Солнечной системы.

Усилиями Ньютона во второй половине XVII века было развито дифференциально-интегральное исчисление. Благодаря этой технике, вычисление длины дуги эллипса сводится к подсчету интеграла

$$L = \int\limits_{x_0}^{x_1} \sqrt{\frac{a^2 - k^2x^2}{a^2 - x^2}}dx,$$

который поэтому называется эллиптическим интегралом. К сожалению, данный интеграл нельзя вычислить с использованием элементарных функций.

Подобные интегралы исследовались такими великими математиками, как Лежандр, Абель и Якоби, что привело их к открытию эллиптических функций. Речь идет о функциях, обратных эллиптическим интегралам, потому они и называются эллиптическими. Самая известная из таких функций, которая встречается в комплексном анализе, алгебраической геометрии, математической физике и многих других областях математики — это так называемая пи-функция Вейерштрасса $\wp$. Каким образом такие функции связаны с кривыми?!

Если эллиптическую кривую нельзя параметризовать с помощью рациональных функций, то с помощью каких можно? Оказывается, существует замечательная парметризация неособых эллиптических кривых

$$\begin{cases} x = \wp(t) \\ y = \wp'(t), \end{cases}$$

где $\wp$ — эллиптическая функция Вейерштрасса, а $\wp'$ — ее производная. Данная параметризация имеет место, поскольку функции $\wp$ и $\wp'$ связаны соотношением

$$(\wp')^2 = 4\wp^3 - g_2\wp - g_3,$$

где g_2 и g_3 — некоторые константы. Именно поэтому кривые, заданные уравнениями вида (от слагаемого x^2 легко избавиться с помощью замены координат)

$$y^2 = x^3 + ax + b,$$

и называются эллиптическими.

Что касается кривых, которым соответствуют поверхности рода $g > 1$, то в 1922 году английским математиком Луисом Морделлом была выдвинута гипотеза о том, что на них содержится лишь конечное число рациональных точек. Эта гипотеза стала теоремой благодаря немецкому математику Герду Фальтингсу, который доказал ее в 1983 году. Из этого результата, в частности, следовало, что уравнение $a^n + b^n = c^n$ при $n > 4$ имеет лишь конечное число взаимно простых решений (так называемая слабая форма великой теоремы Ферма), поскольку род поверхности, связанной с кривой $x^n + y^n = 1$, больше 1 при $n > 4$.

Род поверхности оказывается связан со многими удивительными явлениями. В курсе математического анализа вы столкнетесь с интегралами вдоль алгебраической кривой от рациональной функции

$$I = \int\limits_{f(x,y)=0} R(x,y)dx,$$

где f — многочлен от переменных x, y, а R — рациональнная функция, т.е. отношение многочленов от переменных x, y. Такие интегралы называют абелевыми. Например,

$$\int \sqrt{x^2+1}dx,$$
$$\int \frac{1}{\sqrt{x^2+1}}dx,$$
$$\int \frac{1}{\sqrt{x^3+x+1}}dx,$$
$$\int \frac{1}{\sqrt{x^4-1}}dx.$$

Некоторые из этих интегралов можно «взять явно», т.е. выразить с помощью элементарных функций. Другие — нет. В чем причина этого явления? В курсах математического анализа вас будут учить многочисленным приемам и формулам, которые позволяют «взять» интегралы вида

$$\int R(x, \sqrt{ax^2+bx+c})dx.$$

Одним из таких приемов будут подстановки Эйлера, которые сводят такие интегралы к элементарным. Так вот эти подстановки не более чем рациональная парметризация римановой поверхности кривой $y^2 = ax^2 + bx + c$! Риманова поверхность будет сферой, поэтому и будет допускать рациональную параметризацию. Поэтому абелевы интегралы сводятся к интегралам от рациональных функций и «берутся явно». А если соответствующая риманова поверхность имеет род больше 0, то проинтегрировать в элементарных функциях не удастся. Например, закон колебаний нелинейного маятника дается следующим выражением:

$$t(X) = \int_{X_0}^{X} \frac{dx}{\sqrt{x^3+ax+b}}.$$

$t(X)$ не удастся явно выразить в элементарных функциях переменной X, поскольку риманова поверхность кривой $Y^2 = X^3 + aX + b$ имеет

род, равный 1. По этой же причине трудно вычислить длину орбиты планеты Солнечной системы.

Поразительная идея связать решение диофантовых уравнений с поиском рациональных точек на различных кривых, которая также приводит к изучению поверхностей различного рода, оказала решающее влияние не только на теорию чисел, она также объединила вместе теорию чисел, алгебру, геометрию и другие математические дисциплины, подстегнув их дальнейшее развитие. Род поверхностей, в свою очередь, отвечает не только за решение диофантовых уравнений, но и за элементарность абелевых интегралов, за топологию поверхностей и многое другое. Вновь и вновь мы наблюдаем, что математика едина как наука и что основные ее достижения связаны с ее единством.

Глава VI

Геометрические прогрессии и функция Эйлера

1. Периодичность остатков

Еще одним интересным объектом, связанным с остатками, является *геометрическая прогрессия* $\{a^k\}$ по модулю m. Начнем с рассмотрения нескольких примеров.

k	0	1	2	3	4	5
$2^k \pmod 7$	[1	2	4]	1	2	4

k	0	1	2	3	4	5	6
$2^k \pmod{10}$	1	[2	4	8	6]	2	4

k	0	1	2	3	4	5	6	7
$2^k \pmod{48}$	1	2	4	8	[16	32]	16	32

k	0	1	2	3	4	5
$6^k \pmod{10}$	1	[6]	6	6	6	6

k	0	1	2	3	4	5
$6^k \pmod{16}$	1	6	4	8	[0]	0

k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
$3^k \pmod{17}$	[1	3	9	10	13	5	15	11	16	14	8	7	4	12	2	6]	1

Видно, что, начиная с некоторого места, остатки начинают повторяться: возникает «период» (в таблицах он выделен квадратными скобками), который повторяется бесконечно много раз. При этом период может начинаться не с первого остатка, и длина этого периода может быть как очень маленькой, так и очень большой.

Великий математик Владимир Игоревич Арнольд часто говорил, что математика — это раздел теоретической физики, в котором эксперименты стоят копейки. Как физик, так и маетматик, ставят эксперименты и наблюдают явления. Теория чисел особенно хороша тем, что дает возможность наблюдать множество явлений. Например, периодичность геометрических прогрессий по разным модулям.

Эти наблюдения формализуются следующей теоремой.

Теорема (О периодичности остатков)**.** *Для каждой геометрической прогрессии* $\{a^k\}$ *существует такое натуральное число* $T \in \mathbb{N}$ (*называемое длиной периода*)*, что*

$$a^{k+T} \equiv a^k \pmod{m} \quad \text{для всех } k \geqslant k_0,$$

причем все

$$a^{k_0}, \quad a^{k_0+1}, \quad a^{k_0+2} \quad \ldots \quad a^{k_0+(T-1)}$$

различны по модулю m (*множество этих остатков называется периодом*)*.*

Доказательство. Рассмотрим элементы нашей прогрессии по модулю m:

$$a^0, \quad a^1, \quad a^2, \quad a^3 \quad \ldots$$

Т.к. наша прогрессия бесконечна, а остатков не более m, найдутся такие номера $k < l$, что $a^k \equiv a^l \pmod{m}$.

Выберем наименьшее такое k_0, что $a^{k_0} \equiv a^l \pmod{m}$ для некоторого l. Теперь зафиксируем k_0 и выберем наименьшее $l_0 > k_0$, для которого $a^{k_0} \equiv a^{l_0} \pmod{m}$ (продумайте этот момент! Это довольно сложный этап доказательства!). Положим теперь $T = l_0 - k_0$. Докажем, что число T — это наш искомый период.

Очевидно, что $a^{k_0+T} \equiv a^{k_0} \pmod{m}$. Пусть теперь $k \geqslant k_0$ — произвольный номер. Домножим наше сравнение на a^{k-k_0}. Получим, что $a^{k+T} \equiv a^k \pmod{m}$.

Осталось доказать, что все

$$a^{k_0}, \quad a^{k_0+1}, \quad a^{k_0+2} \quad \ldots \quad a^{k_0+(T-1)}$$

различны. Это следует из минимальности l_0: мы выбрали его так, чтобы все

$$a^{k_0}, \quad a^{k_0+1}, \quad a^{k_0+2} \quad \dots \quad a^{l_0-1} = a^{k_0+(T-1)}$$

были бы различны.

Тем самым наша теорема доказана. □

Возвращаясь к наблюдению примеров, можно сформулировать несколько вопросов:

- Почему в некоторых прогрессиях встречается 1, а в других нет?
- Когда $k_0 = 1$?
- Как вычислить длину периода T, зная a и m?

Дадим ответ на первые два из поставленных вопросов.

Утверждение. *В геометрической прогрессии $\{a^k\}$ по модулю m встречается 1 тогда и только тогда, когда НОД$(a, m) = 1$.*

Доказательство. Следует из теоремы о делителях единицы в $\mathbb{Z}_m$. Действительно, по теореме о периодичности остатков существует такое $T \in \mathbb{N}$, что

$$a^{k+T} \equiv a^k \pmod{m} \quad \text{для всех } k \geqslant k_0.$$

По теореме о делителях единицы

$$a^{k+T} \equiv a^k \pmod{m} \quad \Leftrightarrow \quad a^T \equiv 1 \pmod{m},$$

так как НОД$(a^k, m) = 1$. □

Следствие. $k_0 = 1$ *тогда и только тогда, когда НОД$(a, m) = 1$.*

Что касается вопроса о вычислении длины периода, то ответ на него неизвестен.

Заметим, что похожая ситуация с возникновением периода появляется при обращении обыкновенных дробей в десятичные. Чтобы получить десятичную запись для числа $\frac{m}{n}$, надо разделить уголком m на n. При такой операции будет всегда получаться десятичная

дробь, обладающая следующим свойством: начиная с некоторого места, в ней повторяется одна и та же группа цифр, т.е. десятичная дробь будет периодической. Например,

$$\frac{1}{5} = 0,200000\ldots = 0,2;$$
$$\frac{1}{3} = 0,333333\ldots = 0,(3);$$
$$\frac{1}{7} = 0,142857142\ldots = 0,(142857);$$
$$\frac{5}{14} = 0,357142857\ldots = 0,3(571428).$$

Докажем это утверждение.

Теорема. *Бесконечная десятичная дробь, соответствующая рациональному числу* $\frac{m}{n}$*, является периодической.*

Доказательство. При делении m на n может получиться не более чем n различных остатков: $0, 1, \ldots, n-1$. Поэтому среди n идущих подряд остатков должны быть хотя бы два одинаковых. Но повторение остатков означает и повторение соответствующих цифр в частном, т.е. периодичность получающейся десятичной дроби. □

Замечание. Доказанная теорема мгновенно доставляет примеры чисел, не являющихся рациональными. Таковым, например, является число

$$0,101001000100001\ldots.$$

Замечание. Оказывается, что верно и обратное утверждение. Каждой периодической десятичной дроби соответствует рациональное число. Причем это число можно явно предъявить! Продемонстрируем это на конкретном примере и найдем рациональное число, которое равно дроби

$$0,78(246).$$

Имеем

$$0,78(246) = 0,78 + 0,00(246).$$

Как преобразовать $0,00(246)$? Воспользуемся следующим соображением. Пусть $x = 0,00(246)$. Тогда

$$x = 0,00(246) \Leftrightarrow 100x = 0,(246) \Leftrightarrow 100000x = 246,(246) \Leftrightarrow$$
$$100000x - 100x = 246 \Leftrightarrow 99900x = 246 \Leftrightarrow x = \frac{246}{99900}.$$

Таким образом,

$$0,78(246) = 0,78 + 0,00(246) = \frac{78}{100} + \frac{246}{99900} = \frac{6514}{8325}.$$

Существует ли какая-то связь между периодичностью геометрических прогрессий по модулю и периодичностью десятичных дробей? Имеем

$$\frac{m}{n} = m \cdot \frac{1}{n}.$$

Обратите внимание, что при делении 1 на n уголком вы получаете последовательность остатков $\{10^k\}$ по модулю n (мы пользуемся десятичной системой счисления для записи чисел). Таким образом, поведение прогрессии $\{10^k\}$ по модулю n напрямую связано с поведением цифр десятичной дроби $\frac{1}{n}$, тем самым, и $\frac{m}{n}$. Из проведенной аналогии, например, мгновенно следует ответ на вопрос: каким обыкновенным дробям соответствуют *конечные* десятичные дроби (т.е. с периодом, состоящим из нулей), а каким — *бесконечные*?

Утверждение. *Рациональному числу $\frac{m}{n}$ соответсвует конечная десятичная дробь тогда и только тогда, когда $n = 2^k 5^l$.*

Доказательство. Десятичная дробь будет конечной тогда и только тогда, когда найдется такое N, что

$$10^N \equiv 0 \pmod{n},$$

то есть соответствующий остаток равен 0 и деление на этом останавливается. Откуда следует, что $10^N \vdots n$. Значит, $n = 2^k 5^l$. □

Можно ли указать какую-то более детальную информацию о длине периода соответствующей геометрической прогрессии или десятичной дроби? Оказывается, что да. Начнем со случая простого модуля.

2. Малая теорема Ферма

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

$$a^{p-1} \equiv 1 \pmod{p}.$$

Данный результат был им сформулирован в письме в 1640 году. Доказательство Ферма не было опубликовано. Впервые доказательство опубликовали Лейбниц, а затем Эйлер.

Теорема (Малая теорема Ферма)**.** *Для любого $a \in \mathbb{Z}_p$, $a \neq 0$ имеет место сравнение*

$$a^{p-1} \equiv 1 \pmod{p}.$$

Приведем пример. Пусть $a = 4, p = 11$. Рассмотрим прогрессию $\{4^n\}$ по модулю 11. Имеем

$$1 \to 4 \to 5 \to 9 \to 3 \to 1.$$

Таким образом,

$$4^5 \equiv 1 \pmod{11} \quad \Rightarrow \quad 4^{10} \equiv 1 \pmod{11}.$$

Заметим, что длина периода прогрессии $\{4^n\}$ по модулю 11, равная 5, оказалась делителем числа $10 = 11 - 1$. Посмотрим, что происходит с другими остатками. Выберем остаток a_1, которого нет среди членов прогрессии, и рассмотрим последовательность $\{a_1 \cdot 4^{n-1}\}$. Выбирая $a_1 = 2$, получаем

$$2 \to 8 \to 10 \to 7 \to 6 \to 2.$$

Запишем полученный результат в виде таблицы.

1	4	5	9	3
2	8	10	7	6

Обратите внимание, что все ненулевые элементы $\mathbb{Z}_{11}$, которых ровно $11 - 1 = 10$, содержатся в данной таблице. Тот факт, что

она является прямоугольной, и означает, что длина ее строки (длина периода прогрессии $\{4^n\}$) является делителем общего количества элементов в ней.

Приведем еще один пример. Пусть $a = 3, p = 13$. Получаем таблицу

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

И вновь длина периода прогрессии $\{3^n\}$, равная 3, оказывается делителем числа $13 - 1 = 12$ ненулевых элементов $\mathbb{Z}_{13}$.

Таким образом,

$$3^3 \equiv 1 \pmod{13} \quad \Rightarrow \quad 3^{12} \equiv 1 \pmod{13}.$$

Может оказаться и так, что таблица будет состоять из одной строки. Например, в случае $a = 3, p = 7$.

1	3	2	6	4	5

Откуда сразу следует, что $3^6 \equiv 1 \pmod{7}$.

Докажем малую теорему Ферма.

Доказательство. Рассмотрим прогрессию $\{a^n\}$ по модулю p. НОД$(a, p) = 1$. В силу доказанного нами утверждения в ней встретится 1

$$1 \to a \to a^2 \to \ldots \to a^{T-1} \to a^T \equiv 1.$$

Если $T = p - 1$, то теорема доказана. Иначе выберем ненулевой $a_1 \in \mathbb{Z}_p$, которого нет в прогрессии $\{a^n\}$, и рассмотрим последовательность $\{a_1 \cdot a^n\}$

$$a_1 \to a_1 \cdot a \to a_1 \cdot a^2 \to \ldots \to a_1 \cdot a^{T-1} \to a_1.$$

Ее длина очевидно равна T.

Продолжим этот процесс, пока не исчерпаем все ненулевые элементы $\mathbb{Z}_p$. Получаем следующую таблицу

1	a	a^2	$\ldots\ldots$	a^{T-1}
a_1	a_1a	a_1a^2	$\ldots\ldots$	a_1a^{T-1}
$\ldots$	$\ldots$	$\ldots$	$\ldots\ldots$	$\ldots$
a_k	a_ka	a_ka^2	$\ldots\ldots$	a_ka^{T-1}

В таблице $p-1$ элементов. В каждой строке T элементов, так как любая строка получается из первой умножением на a_i. Поэтому таблица является прямоугольной размера $T \times k$. Откуда следует, что

$$a^T \equiv 1 \pmod{p} \quad \Rightarrow \quad a^{T\cdot k} = a^{p-1} \equiv 1 \pmod{p}.$$

□

Замечание. Существует множество других доказательств малой теоремы Ферма. Например, ее можно доказать с использованием леммы о колоде карт. В самом деле, в силу этой леммы имеем

$$(a\cdot 1)\cdot(a\cdot 2)\cdot\ldots\cdot(a\cdot(p-1)) \equiv 1\cdot 2\cdot\ldots\cdot(p-1) \pmod{p} \Leftrightarrow$$
$$a^{p-1}\cdot(1\cdot 2\cdot\ldots\cdot(p-1)) \equiv 1\cdot 2\cdot\ldots\cdot(p-1) \pmod{p}.$$

Поскольку p — простое, по теореме о делителях единицы получаем

$$a^{p-1} \equiv 1 \pmod{p}.$$

Однако, приведенное нами доказательство вскрывает причину явления, которая носит комбинаторный характер: построенные нами таблицы являются прямоугольными.

Замечание (для технически подготовленного читателя). Умножению на $a \in \mathbb{Z}_p$ соответствует перестановка элементов $\mathbb{Z}_p$. Остатку b ставится в соответствие остаток $b\cdot a$. Рассмотренные нами таблицы есть не что иное, как диаграммы Юнга соответствующих перестановок. Об исследовании диаграмм Юнга перестановок вы можете

прочитать в добавлении «О геометрии диаграмм Юнга перестановок Арнольда».

Замечание. Иногда полезно вместо сравнения $a^{p-1} \equiv 1 \pmod{p}$ использовать его следствие, которое верно для любого остатка, включая 0: $a^p \equiv a \pmod{p}$.

Следствие. *Длина периода прогрессии* $\{a^k\}$ *по простому модулю* p *является делителем числа* $p-1$.

Замечание. На основе малой теоремы Ферма может быть осуществлен так называемый тест простоты Ферма. А именно, рассмотрим натуральное число n. Если найдется хотя бы одно $a \in \mathbb{Z}_n$ такое, что $a^{n-1} \not\equiv 1 \pmod{n}$, то число n — составное. Однако, в противном случае нельзя утверждать, что n будет простым, поскольку выполнение сравнения $a^{n-1} \equiv 1 \pmod{n}$ для всех $a \in \mathbb{Z}_n$ является необходимым, но не достаточным условием простоты числа n. Число n, для которого выполнено сравнение $a^{n-1} \equiv 1 \pmod{n}$, называется *псевдопростым по основанию* a. Например, существуют составные числа, для которых сравнение $a^{n-1} \equiv 1 \pmod{n}$ выполняется для всех a, взаимно простых с n. эти числа называются числами Кармайкла. Существует бесконечное множество таких чисел. Вот несколько примеров:

$$561 = 3 \cdot 11 \cdot 17$$
$$1105 = 5 \cdot 13 \cdot 17$$
$$1729 = 7 \cdot 13 \cdot 19$$
$$2465 = 5 \cdot 17 \cdot 29.$$

3. Функция Эйлера

Возникает естественный вопрос: можно ли сформулировать аналог малой теоремы Ферма для случая *составного* модуля и, тем самым, получить дополнительную информацию о длине периода?

Обратим внимание, что в малой теореме Ферма появляется число $p-1$. Заметим также, что в доказательстве теорем ключевую роль

играли делители единицы и лемма о колоде карт. Как это число связано с полем $\mathbb{Z}_p$? Правильный ответ на поставленный вопрос состоит в том, что $p-1$ — *это количество обратимых элементов (делителей единицы) в* $\mathbb{Z}_p$.

Значит, в случае произвольного модуля $m \in \mathbb{Z}$ нужно рассматривать только обратимые элементы в $\mathbb{Z}_m$.

Определение. Множество

$$\begin{aligned}\mathbb{Z}_m^* &= \{a \in \mathbb{Z}_m : a \text{ — обратимый элемент}\} = \\ &= \{a \in \mathbb{Z}_m : a \text{ — делитель единицы}\} = \\ &= \{a \in \mathbb{Z}_m : \text{НОД}(a, m) = 1\}\end{aligned}$$

называется *множеством обратимых элементов.*

Как правило, какое бы то ни было множество само по себе не очень интересно. Содержательная теория появляется, когда на множестве задана какая-либо операция или операции. Скажем, элементы можно складывать и/или перемножать. Если проанализировать все, что мы к этому моменту обсуждали, то ключевой оказывается именно возможность складывать и перемножать числа, многочлены, остатки и прочее. Теперь появилось новое множество $\mathbb{Z}_m^*$. Какие операции оно «выдерживает»?

Утверждение. $\mathbb{Z}_m^*$ *замкнуто относительно умножения, т.е. если* $a, b \in \mathbb{Z}_m^*$, *то* $a \cdot b \in \mathbb{Z}_m^*$.

Доказательство. Если $a, b \in \mathbb{Z}_m^*$, то по определению a, b — делители единицы в $\mathbb{Z}_m$. Покажем, что $a \cdot b$ также будет делителем единицы. Действительно, существуют $a^{-1}, b^{-1} \in \mathbb{Z}_m$. Как легко видеть,

$$(a \cdot b) \cdot (b^{-1} \cdot a^{-1}) \equiv 1 \pmod{m},$$

т.е. $(a \cdot b)^{-1} = b^{-1} \cdot a^{-1}$. □

При этом множество $\mathbb{Z}_m^*$ не выдерживает сложения! Т.е. *сумма* двух элементов $\mathbb{Z}_m^*$, вообще говоря, не принадлежит $\mathbb{Z}_m^*$. Например, $3 \in \mathbb{Z}_8^*$, однако, $3 + 3 = 6 \notin \mathbb{Z}_8^*$.

Замечание. Вы уже знакомы с *кольцами* $\mathbb{Z}$, $\mathbb{Z}_m$, $\mathbb{R}[x]$, $\mathbb{Z}[\sqrt{-k}]$, которые выдерживают две операции — сложение и умножение. Такие множества как $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{Z}_p$ в случае простого p, в которых можно также делить, называются, как вы знаете, *полями*.

Рассмотрим теперь такие примеры множеств, которые выдерживают только одну операцию:

- Множество $\mathbb{Z}_m^*$ с операцией *умножения*;
- Множество перестановок S_n множества из n элементов с операцией *композиции*;
- Множество *движений* плоскости (т.е. преобразований плоскости, сохраняющих расстояние между точками) $\mathrm{Isom}(\mathbb{E}^2)$ с операцией *композиции*.
- *Симметрии физических законов*, т.е. преобразования координат, при которых закон сохраняется, с операцией *композиции*.
  Так, законы механики, должны сохраняться при переходе от одной инерциальной системы отсчета к другой.

Вышеперечисленные множества похожи — они замкнуты относительно некоторой операции, и у каждого элемента есть обратный. Данная *похожесть также фиксируется определением*: такие множества называются *группами*.

Возникает естественный вопрос, который оказывается ключевым для дальнейших исследований: *сколько элементов в* $\mathbb{Z}_m^*$*?*
Иными словами, сколько существует натуральных чисел, меньших m и взаимно простых с ним?

Определение. Функция $\varphi(m) = |\mathbb{Z}_m^*|$ называется *функцией Эйлера*.

Замечание. Обычно функция Эйлера $\varphi(m)$ вводится как количество натуральных чисел, меньших m и взаимно простых с m. Однако такое определение не мотивировано и не позволяет обобщать функцию Эйлера на другие кольца. О таких обобщениях смотрите в Добавлениях.

Естественно спросить, каким образом связаны $\varphi(m)$ и m. У нас уже были примеры функций, заданных на множестве натуральных чисел, — τ-функция и σ-функция. Чтобы вычислить их значения, мы доказывали, что они обладают свойством мультипликативности. Таким же образом мы вычислим значение функии Эйлера.

Теорема (о мультипликативности функции Эйлера)**.** *Для любых* $a, b \in \mathbb{N}$ *с условием* $НОД(a, b) = 1$ *имееет место равенство*

$$\varphi(a \cdot b) = \varphi(a) \cdot \varphi(b).$$

Перед доказательством рассмотрим пример. Пусть $a = 9,\ b = 4$. Расположим числа от 1 до 36 в таблице следующим образом и выделим числа, которые взаимно просты с 36.

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

Заметим, что во всех столбцах одинаковое количество выделенных чисел, как и во всех строках.

Доказательство. Расположим числа от 1 до ab в прямоугольную таблицу размером $a \times b$: в первой строке числа от 1 до a, во второй — от $a + 1$ до $2a$ и т.д.

Количество чисел, взаимно простых с ab, есть $\varphi(ab)$. С другой стороны, оно равно количеству чисел, взаимно простых и с a, и с b. Посчитаем это число другим способом.

Числа из таблицы представим в виде $ka + r$, где $r = 1, \ldots, a$ и $k = 0, \ldots, b - 1$.

1. Зафиксируем k. Таким образом, мы рассматриваем строку и числа, взаимно простые с a. Так как $НОД(ka + r, a) = НОД(r, a) = 1$, то количество таких чисел в строке равно $\varphi(a)$.

2. Зафиксируем r. Таким образом, мы рассматриваем столбец и числа, взаимно простые с a. Так как НОД$(a, b) = 1$, то среди

$$0 \cdot a + r,\ 1 \cdot a + r,\ \ldots,\ (b-1) \cdot a + r$$

представлены все остатки по модулю b.

Почему мы можем это утверждать?

Поскольку НОД$(a, b) = 1$, по лемме о колоде карт все остатки $0 \cdot a,\ 1 \cdot a,\ \ldots,\ (b-1) \cdot a$ различны. Прибавление одного и того же числа r к каждому из них не меняет картины. Значит, при фиксированном r чисел вида $ka + r$, взаимно простых с b, ровно $\varphi(b)$.

Таким образом, имеется $\varphi(a)$ подходящих столбцов, в каждом из которых $\varphi(b)$ подходящих чисел. Следовательно, всего подходящих чисел $\varphi(a) \cdot \varphi(b)$. □

Утверждение. *Если p — простое число, то*

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

Доказательство. Среди чисел от 1 до p^α в точности $p^{\alpha-1}$ штук, кратных p. Оставшиеся взаимно просты с p^α. □

Следствие. *Если $m = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k}$, то*

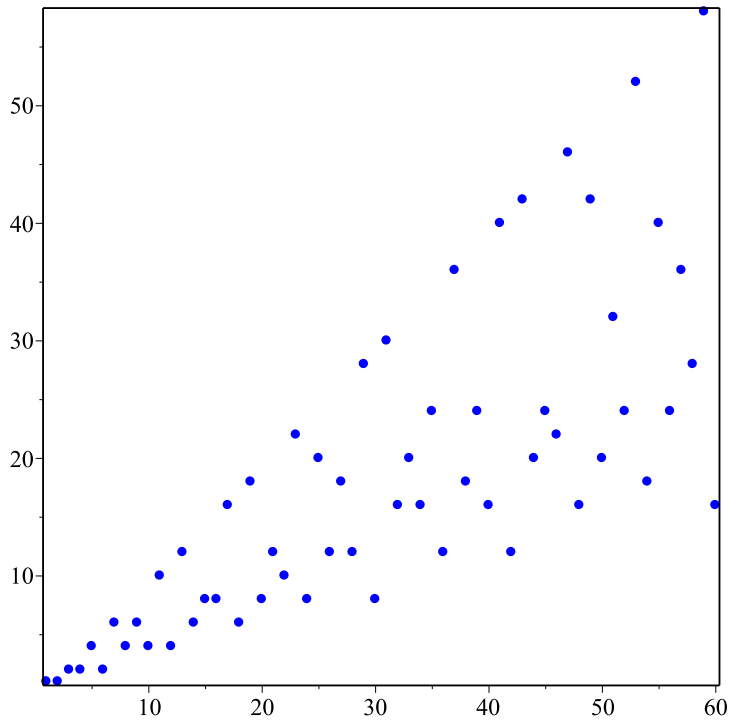
$$\varphi(m) = m \cdot \left(1 - \frac{1}{p_1}\right) \cdot \ldots \cdot \left(1 - \frac{1}{p_k}\right).$$

Доказательство.

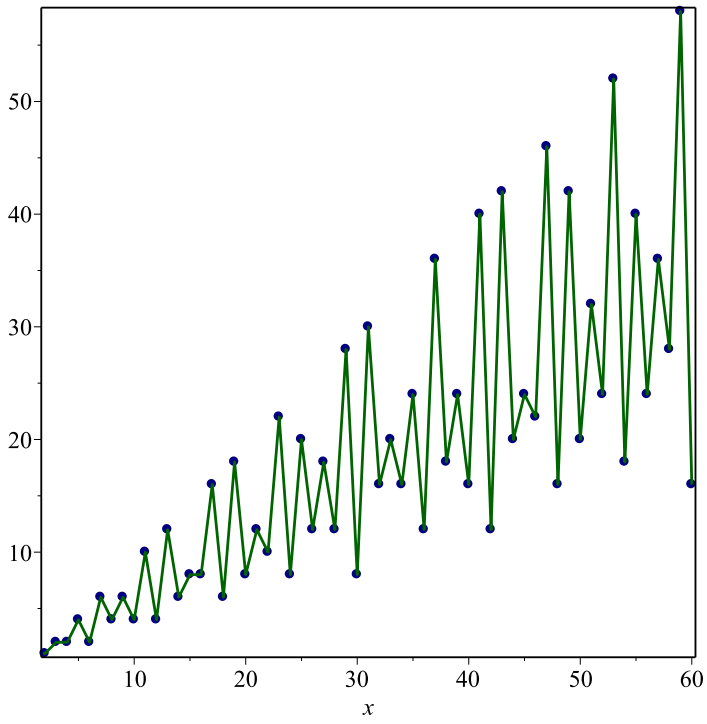
$$\begin{aligned}
\varphi(m) &= \varphi(p_1^{\alpha_1} \cdot \ldots \cdot p_k^{\alpha_k}) = \varphi(p_1^{\alpha_1}) \cdot \ldots \cdot \varphi(p_k^{\alpha_k}) \\
&= (p_1^{\alpha_1} - p_1^{\alpha_1 - 1}) \cdot \ldots \cdot (p_k^{\alpha_k} - p_k^{\alpha_k - 1}) \\
&= p_1^{\alpha_1} \cdot \left(1 - \frac{1}{p_1}\right) \cdot \ldots \cdot p_k^{\alpha_k} \cdot \left(1 - \frac{1}{p_k}\right) \\
&= m \cdot \left(1 - \frac{1}{p_1}\right) \cdot \ldots \cdot \left(1 - \frac{1}{p_k}\right).
\end{aligned}$$

□

Что мы можем сказать о множестве *значений* функции Эйлера?

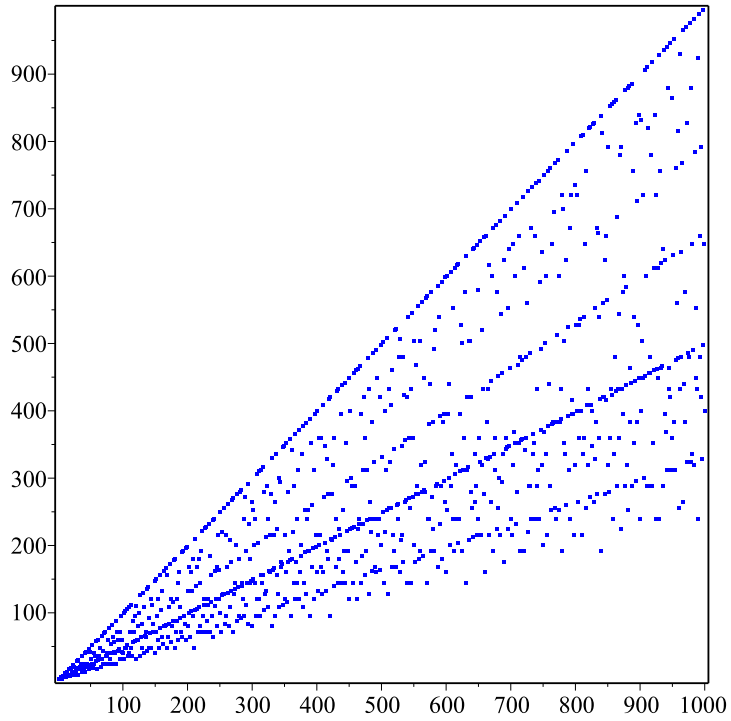


Значения $\varphi(n)$ на отрезке от 1 до 60.



Значения $\varphi(n)$ на отрезке от 1 до 60.

Если теперь рассмотреть график на больших масштабах — для первой тысячи натуральных чисел, то можно заметить некоторые регулярные детали.



Значения $\varphi(n)$ на отрезке от 1 до 1000.

Исследование структуры множества значений функции Эйлера представляет собой очень сложную задачу, которая содержит множество открытых вопросов (например, не каждое число является значением функции Эйлера и нет явного описания множества значений: если вам дано четное число, не существует эффективного способа проверить, является ли оно одним из значений функции Эйлера). Попробуйте доказать, что, например, число 14 не является значением функции Эйлера. Почему мы говорим именно о четных числах? Дело в том, что имеет место следующая теорема

Теорема. *$\varphi(m)$ четно при $m \geqslant 3$.*

Доказательство. $(a, m) = (m - a, m)$. Следовательно, элементы $\mathbb{Z}_m^*$ разбиваются на пары. Если же $a = m - a$, то $(a, m) = (a, 2a) > 1$ при $m > 2$. □

Теперь приведем доказательство теоремы, обобщающей результат

Ферма на случай составного модуля.

Теорема (Эйлер)**.** *Для любого* $a \in \mathbb{Z}_m^*$ *выполнено*

$$a^{\varphi(m)} \equiv 1 \pmod{m}.$$

Как и в малой теореме Ферма, начнем с примера. Пусть $a = 5, m = 36$ и $\varphi(36) = 12$. Рассматривая умножение на 5, получаем следующую таблицу.

1	5	25	17	13	29
7	35	31	11	19	23

Таким образом,

$$5^6 \equiv 1 \pmod{36} \quad \Rightarrow \quad 5^{12} \equiv 1 \pmod{36}.$$

Доказательство. Рассмотрим прогрессию $\{a^n\}$ по модулю m. НОД$(a, m) = 1$, поэтому в силу доказанного нами утверждения в ней встретится 1

$$1 \to a \to a^2 \to \ldots \to a^{T-1} \to a^T \equiv 1.$$

Если $T = \varphi(m)$, то теорема доказана. Иначе выберем $a_1 \in \mathbb{Z}_m^*$, которого нет в прогрессии $\{a^n\}$, и рассмотрим последовательность $\{a_1 \cdot a^n\}$

$$a_1 \to a_1 \cdot a \to a_1 \cdot a^2 \to \ldots \to a_1 \cdot a^{T-1} \to a_1.$$

Ее длина очевидно равна T.

Продолжим этот процесс, пока не исчерпаем все элементы $\mathbb{Z}_m^*$. Получаем следующую таблицу

1	a	a^2	$\ldots\ldots$	a^{T-1}
a_1	$a_1 a$	$a_1 a^2$	$\ldots\ldots$	$a_1 a^{T-1}$
$\ldots$	$\ldots$	$\ldots$	$\ldots\ldots$	$\ldots$
a_k	$a_k a$	$a_k a^2$	$\ldots\ldots$	$a_k a^{T-1}$

В таблице $\varphi(m)$ элементов. В каждой строке T элементов, так как любая строка получается из первой умножением на a_i. Поэтому таблица является прямоугольной размера $T \times k$. Откуда следует, что

$$a^T \equiv 1 \pmod{m} \quad \Rightarrow \quad a^{T \cdot k} = a^{\varphi(m)} \equiv 1 \pmod{m}.$$

□

Замечание. По аналогии с малой теоремой Ферма теорему Эйлера можно также доказать с привлечением леммы о колоде карт.

Пусть $\mathbb{Z}_m^* = \{a_1, \ldots, a_{\varphi(m)}\}$. Тогда по лемме о колоде карт для $\mathbb{Z}_m^*$ имеем

$$\{a \cdot a_1, \ldots, a \cdot a_{\varphi(m)}\} = \mathbb{Z}_m^*.$$

Значит,

$$\begin{aligned} & a_1 \cdot \ldots \cdot a_{\varphi(m)} \equiv (a \cdot a_1) \cdot \ldots \cdot (a \cdot a_{\varphi(m)}) \equiv \\ & \equiv a^{\varphi(m)} \cdot (a_1 \cdot \ldots \cdot a_{\varphi(m)}) \pmod{m} \quad \Leftrightarrow \\ & a^{\varphi(m)} \equiv 1 \pmod{m} \end{aligned}$$

Впрочем, данное доказательство, как и в случае простого модуля, не вскрывает причины явления.

Следствие. *Длина периода прогрессии $\{a^k\}$ по модулю m при условии, что НОД$(a, m) = 1$, является делителем числа $\varphi(m)$.*

Возвращаясь к третьему из поставленных в начале главы вопросов, ответ на который неизвестен, приведем имеющиеся экспериментальные данные (см. Арнольд В.И. «Группы Эйлера и арифметика геометрических прогрессий»). Рассмотрим прогрессию $\{2^k\}$ по нечетным модулям m. Ниже представлены данные о количестве строк N в таблицах и длине периода T.

m	3	5	7	9	11	13	15	17	19	21	23	25	27	29	31	33	35	37	39
N	1	1	2	1	1	1	2	2	1	2	2	1	1	1	6	2	2	1	2
T	2	4	3	6	10	12	4	8	18	6	11	20	18	28	5	10	12	36	12

m	41	43	45	47	49	51	53	55	57	59	61	63	65	67	69	71	73	75
N	2	3	2	2	2	4	1	2	2	1	1	6	4	1	2	2	8	2
T	20	14	12	23	21	8	52	20	18	58	60	6	12	66	22	35	9	20

m	77	79	81	83	85	87	89	91	93	95	97	99
N	2	2	1	1	8	2	8	6	6	2	2	2
T	30	39	54	82	8	28	11	12	10	36	48	30

4. Рост в среднем функции Эйлера

Значения функции Эйлера $\varphi(m)$ ведут себя при росте m нерегулярно. Чтобы исследовать это поведение, его нужно каким-то образом «регуляризовать». Одним из способов является рассмотрение усредненной функции

$$\widehat{\varphi}(N) = \frac{\varphi(1) + \varphi(2) + \ldots + \varphi(N)}{N}.$$

«Скачки» значений функции Эйлера компенсирую друг друга при подсчете среднего арифметического, в результате чего рост усредненной функции будет исследовать легче. Из графика $\widehat{\varphi}$ видно, что усредненная функция растет линейно.

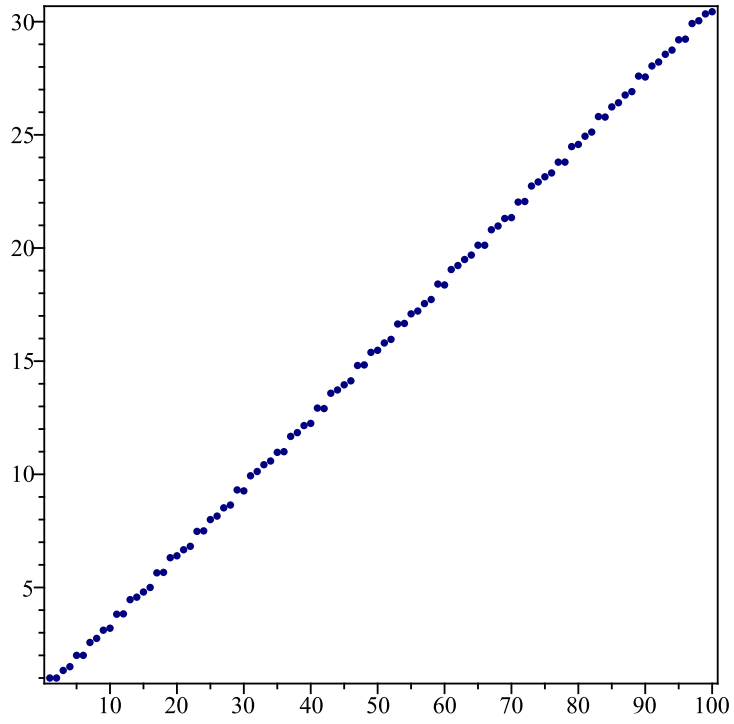
Определение. Функции f и g растут в среднем одинаково, если отношение

$$\frac{f(1) + f(2) + \ldots + f(N)}{g(1) + g(2) + \ldots + g(N)} = \frac{\sum\limits_{n \leqslant N} f(n)}{\sum\limits_{n \leqslant N} g(n)}$$

стремится к 1 с ростом N.

Будем обозначать такие функции $f \widehat{\sim} g$. Условие $f \widehat{\sim} g$ означает, что средние арифметические

$$\widehat{f}(N) = \frac{f(1) + f(2) + \ldots + f(N)}{N} \quad \text{и} \quad \widehat{g}(N) = \frac{g(1) + g(2) + \ldots + g(N)}{N}$$



Значения $\widehat{\varphi}(n)$ на отрезке от 1 до 100.

стремятся к равенству с ростом N. Именно поэтому и говорят, что функции растут *в среднем* одинаково.

Замечание. Строгое определение слова «стремится» будет дано вам в курсе математического анализа. Наши рассуждения будут носить «полуэмпирический» характер, что нисколько не помешает получить желаемый результат и не затмит основные идеи разбираемых конструкций.

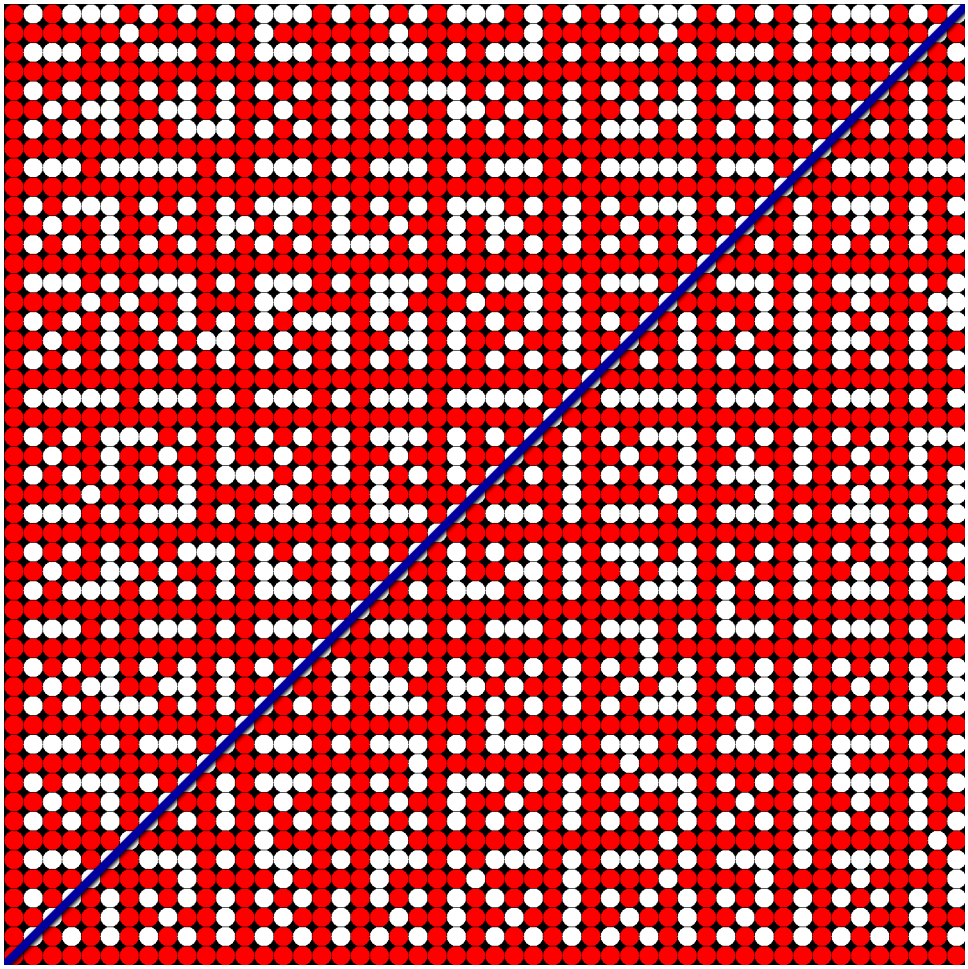
Если для «сложной» функции f удастся найти «простую» g, которая растет в среднем так же, то это даст информацию о поведении функции f. Исследуем таким образом фунцию Эйлера.

Имеет место удивительная

Теорема. *Отношение*

$$\frac{\varphi(1)+\varphi(2)+\ldots+\varphi(N)}{1+2+\ldots+N}$$

стремится к $1/\zeta(2)=6/\pi^2$ *с ростом* N.



Доказательство. Рассмотрим квадрат размером $N \times N$ в первой четверти координатной плоскости. Этот график мы уже видели в III.6. Целые точки с взаимно простыми координатами закрашены. Посчитаем, сколько их в этом квадрате.

Рассмотрим треугольник под главной диагональю и его столбцы. Во втором столбце $\varphi(2)$ закрашенных точек, в третьем — $\varphi(3)$ и так далее вплоть до N-ого, в котором $\varphi(N)$ закрашенных точек. Легко видеть, что график симметричен относительно главной диагонали, поэтому в нем всего

$$1 + 2 \cdot (\varphi(2) + \ldots + \varphi(N))$$

закрашенных точек (+1 потому что необходимо учесть точку (1,1)).

Из раздела III.6 мы знаем, что доля закрашенных точек (вероятность несократимости случайной обыкновенной дроби) стремится к

$1/\zeta(2) = 6/\pi^2$ с ростом N. Всего точек в квадрате N^2, поэтому

$$\frac{1 + 2 \cdot (\varphi(2) + \ldots + \varphi(N))}{N^2} \longrightarrow 1/\zeta(2) = 6/\pi^2.$$

Имеем

$$\frac{2 \cdot (\varphi(1) + \varphi(2) + \ldots + \varphi(N))}{N^2} - \frac{1 + 2 \cdot (\varphi(2) + \ldots + \varphi(N))}{N^2} = \frac{1}{N^2}.$$

При стремлении N к бесконечности $1/N^2$ стремится к 0, и им можно пренебречь.

Итак, мы имеем

$$\frac{\varphi(1) + \varphi(2) + \ldots + \varphi(N)}{\frac{N^2}{2}} \longrightarrow 1/\zeta(2) = 6/\pi^2.$$

Осталось заметить, что

$$1 + 2 + \ldots + N = \frac{N(N+1)}{2} = \frac{N^2}{2} + \frac{N}{2}.$$

Из-за слагаемого $N/2$ может показаться, что мы получили неподходящую функцию. Но при больших N слагаемое $N^2/2$ намного превосходит $N/2$, поэтому последним можно пренебречь. □

Из всего вышесказанного следует, что

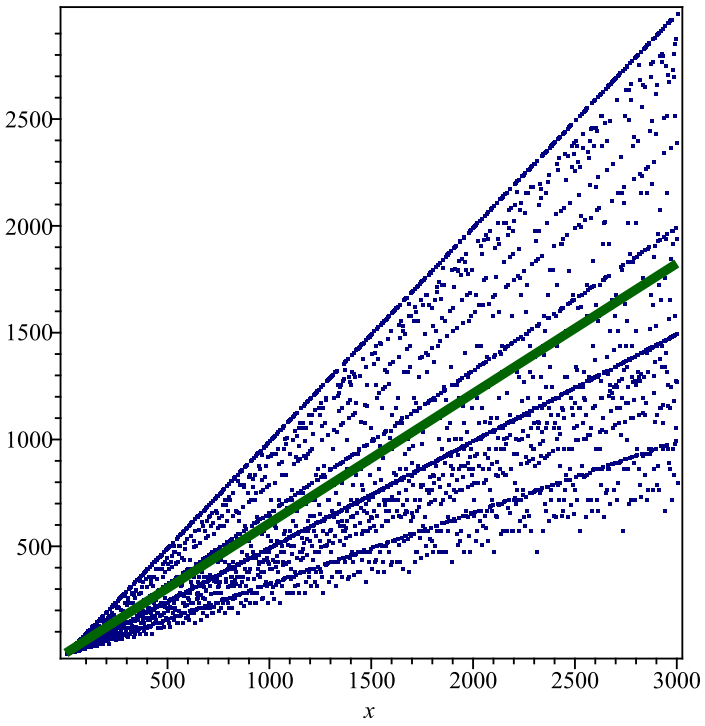
$$\varphi(n) \widehat{\sim} \frac{6}{\pi^2} n.$$

Аналогично можно иисследовать рост в среднем других арифметических функций. Например, сигма-функция имеет рост в среднем

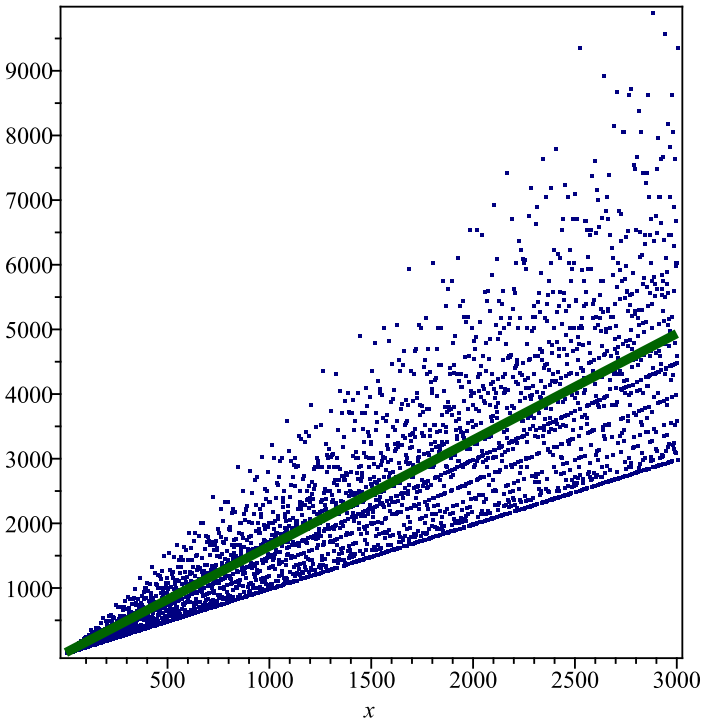
$$\sigma(n) \widehat{\sim} \frac{\pi^2}{6} n.$$

А вот в случае тау-функции рост в среднем устроен неожиданным образом (определение функции ln смотри в I.2)

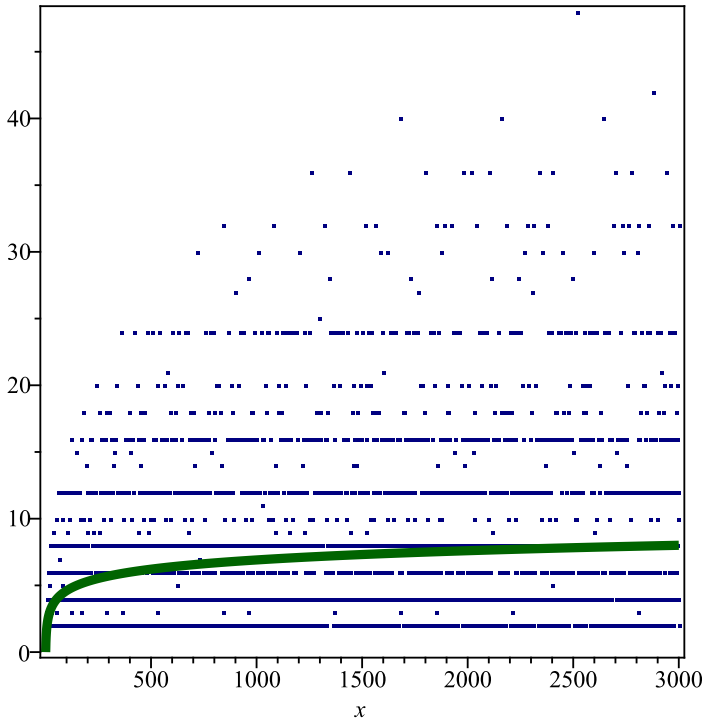
$$\tau(n) \widehat{\sim} \ln n.$$



Функция Эйлера и ее рост в среднем.



Сигма-функция и ее рост в среднем.



Тау-функция и ее рост в среднем.

Об обобщениях функции Эйлера и исследовании их поведения вы можете прочитать в добавлениях «О матричных аналогах функции Эйлера» и «О функции Эйлера алгебраических расширений колец вычетов».

Глава VII

Почему многочлен не функция

В большинстве школьных учебников и пособий дается примерно такое «определение».

Многочленом от одной переменной называется выражение вида

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n.$$

Вообще говоря, эта фраза определением не является. Здесь просто одно слово — «многочлен», заменено другим — «выражение». А что такое выражение? Даже если нам аккуратно разъяснят, что это такое, непонятно, зачем такие сложности? Почему нельзя просто сказать, что многочлен — это функция, которая числу c ставит в соответствие число

$$a_0 + a_1c + a_2c^2 + \ldots + a_nc^n,$$

где n — степень многочлена?

И еще. Нетрудно заметить, что определение многочлена как «выражения вида» в дальнейшем попросту игнорируется, например, когда мы ищем корни многочлена, т.е. решаем уравнение или, когда строим график квадратичной функции. В связи с этим возникает вопрос. Если определение не работает, если мы, пусть и неявно, пользуемся другим определением, то, может быть, следует отказаться от такого определения? Может быть, многочлен это все-таки функция, а не «выражение вида»?

1. Что есть многочлен?

Давайте разберемся, в чем тут дело. Прежде всего, заметим, что, как мы уже многократно отмечали, *множество многочленов очень похоже на множество целых чисел.* Многочлены, так же как и числа,

можно складывать и перемножать, иногда делить и даже раскладывать на множители, т.е. они являются кольцами. Кроме того, в множестве многочленов есть неприводимые многочлены, очень похожие на простые числа. Кроме многочленов и целых чисел у нас уже накопилась целая коллекция колец (вообще полезно коллекционировать кольца, поля и другие математические объекты, объединенные некоторой далеко не всегда очевидными аналогиями). Вспомним про кольца $\mathbb{Z}[\sqrt{-k}]$, среди которых особенно выделим случай $k = 1$ гауссовых чисел, и кольцо остатков $\mathbb{Z}_m$. Но и это еще не все. И в множестве целых чисел, и в множестве многочленов существует деление с остатком! Только в множестве чисел остаток меньше модуля делителя, а в множестве многочленов степень остатка меньше степени делителя.

Замечание. Отметим также, что деление с остатком существует и в кольце гауссовых чисел. Что будет для них аналогом модуля целого числа и степени многочлена? Как вы догадываетесь, таким аналогом будет норма гауссова числа. Напомним, что нормой числа $a+b\sqrt{-1} \in \mathbb{Z}[\sqrt{-1}]$ называется целое неотрицательное число $a^2 + b^2$.

Итак, степень многочлена является аналогом модуля числа, и, более того, благодаря степени, многочлены становятся еще больше похожими на числа. Имея это в виду, говорят, что и числа, и многочлены являются не просто кольцами, а *евклидовыми кольцами*, и там, и там имеет место деление с остатком и выполнена соответствующая теорема.

Однако аналогия между модулем числа и степенью многочлена (если многочлен — это функция) исчезает, если рассматривать многочлены с коэффициентами в поле остатков по простому модулю. Кольцо многочленов с коэффициентами из поля $\mathbb{Z}_p$ обозначается $\mathbb{Z}_p[x]$. В самом деле, в этом случае, как известно, справедлива малая теорема Ферма. Для всякого $a \in \mathbb{Z}$ и простого p имеет место сравнеие

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

Иными словами, малая теорема Ферма утверждает, что на множестве $\mathbb{Z}_p$ многочлен-функция x^p совпадает с многочленом-функцией x. Легко понять, что в этом случае нет и аналога теоремы о делении многочлена на многочлен с остатком, то есть, степень многочлена не «работает». А степень хорошо бы «спасти», иначе «полетит» наша аналогия. Спасение оказывается довольно неожиданным: мы отказываемся от определения многочлена как функции, а даем новое

Определение. *Многочленом* называется бесконечная последовательность

$$(a_0, a_1, \ldots, a_n, \ldots)$$

в которой почти все члены, т.е. все, за исключением конечного числа, равны нулю (номер, начиная с которого стоят нули у разных последовательностей может быть разным). Условимся нумеровать члены последовательностей с нуля.

Если начиная с номера $d+1$, стоят нули, то *степенью многочлена* q называют теперь число d — номер последнего ненулевого члена последовательности. Обозначается это следующим образом:

$$\deg q = d.$$

Итак, для того, чтобы «спасти» степень, пришлось изменить определение многочлена. Правда, мы еще должны доказать, что эта «новая» степень, действительно является аналогом модуля. Заметим сначала, что каждый многочлен-последовательность $(a_0, a_1, \ldots, a_n, \ldots)$ индуцирует функцию

$$f(x) = a_0 + a_1 x + \ldots + a_n x^n + \ldots = \sum_{k \geqslant 0} a_k x^k.$$

Эта функция числу c ставит в соответствие число

$$a_0 + a_1 c + a_2 c^2 + \ldots + a_d c^d,$$

где d — степень соответствующего многочлена. Так последовательность $(0, \ldots, 0, 1, 0, \ldots)$ с единицей на p-ом месте индуцирует функцию x^p, а последовательность $(0, 1, 0, \ldots, 0, \ldots)$, индуцирует функцию

x, то есть разные многочлены-последовательности в данном случае, индуцируют в силу малой теоремы Ферма одну и ту же функцию на $\mathbb{Z}_p$.

Продолжая «спасение» степени, отметим, что на языке последовательностей очень просто определить сумму и произведение многочленов.

Определение. *Суммой* многочлена $(a_0, a_1, \ldots, a_n, \ldots)$ и многочлена $(b_0, b_1, \ldots, b_n, \ldots)$ называется многочлен $(a_0 + b_0, a_1 + b_1, \ldots, a_n + b_n, \ldots)$, а *произведением* — многочлен $(c_0, c_1, \ldots, c_n)$, в котором

$$c_k = \sum_{s+t=k} a_s b_t$$

(суммирование в последней формуле ведется по всем неотрицательным индексам s и t).

Если вы обратитесь к главе II, где обсуждалось деление с остатком для многочленов, то легко увидите, что процедура деления с остатком работает для многочленов-последовательностей и с коэффициентами из $\mathbb{Z}_p$. Для них также справедлива теорема деления с остатком.

Теорема. *Пусть a и b — многочлены от одной переменной. Тогда существуют однозначно определенные многочлены q и r такие, что $a = bq + r$ и*

$$\deg r < \deg b \quad \text{(строго меньше!)}.$$

Как следует из предыдущих рассуждений, можно и нужно рассматривать многочлены с коэффициентами в различных *полях*. Нам известны следующие примеры полей: поле вещественных чисел $\mathbb{R}$, рациональных чисел $\mathbb{Q}$, остатков по простому модулю $\mathbb{Z}_p$. Соответствующие кольца многочленов от одной переменной обозначаются $\mathbb{R}[x], \mathbb{Q}[x]$ и $\mathbb{Z}_p[x]$. Все эти кольца являются евклидовыми (таковым является кольцо многочленов над любым полем).

Разумеется, возникает резонный вопрос, а что будет, если рассматривать многочлены с коэффициентами не в поле, а в кольце.

Например, $\mathbb{Z}[x]$. Очевидно, их можно складывать и перемножать — в этом смысле $\mathbb{Z}[x]$ не отличается от $\mathbb{R}[x]$ и также является кольцом. Вопрос в том, является ли $\mathbb{Z}[x]$ евклидовым кольцом, т.е. существует ли в этом кольце возможность деления с остатком и выполняется ли алгоритм Евклида? Оказывается, что нет! Дело в том, что целые числа, вообще говоря, не делятся друг на друга, а в процессе деления с остатком двух многочленов необходимо делить друг на друга их коэффициенты. В случае коэффициентов из поля таких проблем не возникает потому, что определяющим свойством поля является как раз возможность деления на любые ненулевые элементы.

2. Многочлены с коэффициентами в поле

Опишем теперь некоторые общие свойства, которыми обладают многочлены с коэффициентами в поле. Будем обозначать их $\Bbbk[x]$ (известыне нам примеры — $\Bbbk = \mathbb{R}, \mathbb{Q}$ или $\mathbb{Z}_p$). Удивительным образом теорема о делении с остатком дает некоторую информацию об устройстве *множества корней* таких многочленов. Особое значение при этом имеет деление с остатком на многочлен $x - c$, где $c \in \Bbbk$.

Утверждение (Теорема Безу)**.** *Остаток от деления многочлена q на $x - c$ равен $q(c)$ (т.е. значению многочлена q при $x = c$.)*

Доказательство. Действительно, разделим многочлен q на $x - c$ с остатком

$$q(x) = (x - c) \cdot s(x) + r(x),$$

где $\deg r < \deg(x - c)$, а значит, $\deg r = 0$, т.е. r — число. При $x = c$ имеем

$$q(c) = r.$$

□

Из теоремы Безу очевидным образом получаем

Следствие. *Число c является корнем многочлена q тогда и только тогда, когда $q \vdots (x - c)$.*

Этим мы сейчас воспользуемся для доказательства следующей теоремы.

Теорема (О числе корней многочлена)**.** *Число корней ненулевого многочлена не превосходит его степени.*

Замечание. Напомним, что мы уже пользовались данной теоремой в первой главе, когда доказывали невозможность построения многочлена от одной переменной, значения которого при всех натуральных x — простые числа.

Доказательство. Пусть c_1 — корень многочлена q. Тогда в силу следствия из теоремы Безу

$$q = (x - c_1)q_1.$$

Пусть c_2 — корень многочлена q_1. Тогда

$$q_1 = (x - c_2)q_2$$

и, значит,

$$q = (x - c_1)(x - c_2)q_2.$$

Продолжая этот процесс далее, мы в конце концов получим

$$q = (x - c_1)(x - c_2)\dots(x - c_m)s,$$

где многочлен s не имеет корней.
Таким образом, числа $c_1, c_2, \dots, c_m$ — все корни q (среди них могут быть одинаковые). Очевидно, что $m \leqslant \deg q$. □

Покажем, насколько существенно рассмотрение многочленов с коэффициентами именно над полем. Для этого рассмотрим несколько примеров.

- $2x - 4 \in \mathbb{Z}_6[x]$. Поиск корней данного многочлена равносилен решению сравнения $2x - 4 \equiv 0 \pmod 6$. Как легко видеть, оно имеет два решения: 2 и 5. Значит, многочлен первой степени имеет два корня!

- $x^2 - 1 \in \mathbb{Z}_8[x]$. Сравнение $x^2 - 1 \equiv 0 \pmod 8$ имеет четыре решения: 1, 3, 5, 7. То есть многочлен второй степени имеет 4 корня.

Заметим, что в школьном курсе алгебры, где обычно не рассматриваются многочлены над $\mathbb{Z}_p$, можно было бы сразу определить многочлен как функцию над $\mathbb{R}$, так как в этом случае определения многочлена как функции и как последовательности эквивалентны. Это вытекает из следующего утверждения.

Теорема. *Коэффициенты равных многочленов-функций p и q над $\mathbb{R}$ равны.*

Доказательство. Предположим, что многочлены-последовательности p и q различны. Рассмотрим их разность $s = p - q$ — многочлен-последовательность, который в силу условия индуцирует нулевую функцию. Следовательно, все элементы поля $\mathbb{R}$ являются корнями ненулевого многочлена-последовательности s, что противоречит теореме о том, что количество корней многочлена не превосходит его степени, поскольку поле $\mathbb{R}$ содержит *бесконечное* число элементов. □

Замечание. Именно *бесконечность* поля $\mathbb{R}$ являлась ключевым фактом при доказательстве данной теоремы. Действительно, в случае *конечного* поля $\mathbb{Z}_p$ мы приводили два различных многочлена x^p и x, которые индуцировали одну и ту же функцию на $\mathbb{Z}_p$. Все элементы поля $\mathbb{Z}_p$ является корнями многочлена $s = x^p - x$ степени p. Поскольку их всего p (поле конечное!), это никак не противоречит теореме о том, что количество корней многочлена не превосходит его степени.

Интересно было бы узнать, а какие вообще бывают конечные поля? Или хотя бы сколько элементов может содержать конечное поле? Исчерпываются ли конечные поля известными нам примерами остатков по простому модулю $\mathbb{Z}_p$? Оказывается, что нет. Существуют еще конечные поля, содержащие p^n элементов, где p — простое, а n —

произвольное натуральное число. Однако других конечных полей не существует. Эти поля оказываются тесно связанными с кольцом многочленов $\mathbb{Z}_p[x]$, о чем мы будем подробнее говорить в разделе XI.6.

В заключение отметим, что определение многочлена как «выражения вида», которое «старше» определения многочлена как последовательности, также «работает» и для многочленов над $\mathbb{Z}_p$. По всей видимости, именно по этой причине оно и появилось в свое время в различных учебниках алгебры. Потом до остатков не «дошли руки», а определение осталось, вызывая у многих вопрос: «Почему же многочлен не функция?»

Глава VIII

Квадратичные вычеты

В предыдущих главах мы научились решать линейные сравнения вида $ax \equiv b \pmod{m}$ и использовали эту технику при исследовании колец остатков и решении линейных диофантовых уравнений. Возникает вопрос, а можно ли сделать следующий шаг и научиться решать квадратные сравнения? Какие интересные явления можно наблюдать, исследуя их?

Если рассматривать случай простого модуля p, сравнения $ax \equiv b \pmod{p}$ устроены простым образом, поскольку в $\mathbb{Z}_p$ все ненулевые элементы является делителями единицы. Однако, изучать уже квадратные сравнения в $\mathbb{Z}_p$ очень интересно. Этим мы теперь и займемся!

1. Суммы квадратов и теорема о $\sqrt{-1}$

Мы уже встречались с парадоксальными равенствами в остатках, наподобие $\frac{1}{3} \equiv 2 \pmod{5}$. Вот еще один пример:

$$2^2 \equiv -1 \pmod{5}.$$

Таким образом, сравнение

$$x^2 + 1 \equiv 0 \pmod{5}$$

имеет решение, в то время как аналогичное ему уравнение в действительных числах

$$x^2 + 1 = 0$$

решений не имеет.

Что такое $\frac{1}{3}$ в $\mathbb{Z}_5$? Как мы с вами знаем, это решение сравнения $3x \equiv 1 \pmod{5}$.

Что такое $\sqrt{-1}$? Это решение уравнения $x^2+1=0$. Таким образом, в $\mathbb{Z}_5$ существует $\sqrt{-1}$!

Возникает естественный вопрос: *для каких простых p в $\mathbb{Z}_p$ существует $\sqrt{-1}$?*

Прежде чем пытаться ответить на этот вопрос, необходимо провесьти *эксперимент*, который позволяет сформулировать *гипотезу*. Математика — экспериментальная наука!

Чтобы доказать теорему о $\sqrt{-1}$, нам понадобится следующее утверждение, которое выполнено для любого простого модуля.

Теорема (Вильсон)**.** *Если число p — простое, то имеет место сравнение*

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

Мы приведем два доказательства этого утверждения. Первое не использует ничего, кроме стандартных соображений теории сравнений. Второе носит более идейный характер и вскрывает причину явления.

Доказательство I. Пусть p — простое число. Рассмотрим $a \in \mathbb{Z}_p$ такой, что $a \neq 0, \pm 1$. Тогда по теореме о делителях единицы существует $a^{-1} \in \mathbb{Z}_p$. Более того, в силу леммы о колоде карт такой a^{-1} *единственный.*

Кроме того, $a^{-1} \not\equiv a \pmod{p}$. Действительно, предположим, что это не так. Тогда

$$\begin{aligned} &a^{-1} \equiv a \pmod{p} \Rightarrow a^2 \equiv 1 \pmod{p} \Rightarrow \\ &(a-1)(a+1) \equiv 0 \pmod{p} \end{aligned}$$

Откуда следует, что либо $a-1 \equiv 0 \pmod{p}$, либо $a+1 \equiv 0 \pmod{p}$, поскольку p — простое.

В таком случае можно разбить остатки на пары:

$$\begin{aligned} (p-1)! &\equiv 1 \cdot 2 \cdot \ldots \cdot (p-2) \cdot (p-1) \equiv \\ &\equiv 1 \cdot (2 \cdot 2^{-1}) \cdot \ldots \cdot \left(\frac{p-1}{2}\right) \cdot \left(\frac{p-1}{2}\right)^{-1} \cdot (p-1) \equiv \\ &\equiv 1 \cdot 1 \cdot \ldots \cdot 1 \cdot (p-1) \equiv -1 \pmod{p}. \end{aligned}$$

□

Доказательство II. Рассмотрим многочлен $x^{p-1}-1$ как элемент $\mathbb{Z}_p[x]$ кольца многочленов с коэффициентами из поля $\mathbb{Z}_p$. В силу теоремы о количестве корней многочлена, доказанной в предыдущей главе, сравнение

$$x^{p-1} - 1 \equiv 0 \pmod{p}$$

имеет не более $p-1$ решений. В силу малой теоремы Ферма этими решениями являются $1, 2, \ldots, (p-1) \in \mathbb{Z}_p$. По теореме Виета

$$1 \cdot 2 \cdot \ldots \cdot (p-1) = (p-1)! \equiv -1 \pmod{p}.$$

□

Замечание. Оказывается, что верно и обратное утверждение. Если $(p-1)! \equiv -1 \pmod{p}$, то p — простое. Доказательство следует непосредственно из леммы о простом делителе.

Теперь мы готовы доказать теорему.

Теорема (О $\sqrt{-1}$)**.** *Сравнение*

$$x^2 + 1 \equiv 0 \pmod{p}$$

имеет решение тогда и только тогда, когда

$$p \equiv 1 \pmod{4}.$$

Доказательство. Пусть сравнение $x^2+1 \equiv 0 \pmod{p}$ имеет решение a. Докажем, что $p = 4k+1$.

Предположим противное: пусть $p = 4k+3$. Тогда по малой теореме Ферма имеем

$$1 \equiv a^{p-1} = a^{4k+2} = (a^2)^{2k+1} \equiv (-1)^{2k+1} \equiv -1 \pmod{p}.$$

Противоречие.

Пусть теперь $p = 4k+1$. Найдем решение сравнения $x^2 \equiv -1 \pmod{p}$. По теореме Вильсона имеем

$$\begin{aligned} -1 \equiv (p-1)! \equiv 1 \cdot 2 \cdot \ldots \cdot (2k) \cdot (-2k) \cdot \ldots \cdot (-2) \cdot (-1) = \\ = (-1)^{2k} \cdot (1 \cdot 2 \cdot \ldots \cdot 2k)^2 = ((2k)!)^2 \pmod{p} \end{aligned}$$

□

Доказанная теорема не только примечательна сама по себе, но и позволяет дать решение одной классической задачи. Речь идет о следующем вопросе: *какие числа представимы в виде суммы двух квадратов целых чисел?*

Еще Диофантом было открыто следующее тождество, носящее его имя:

$$(a^2+b^2)\cdot(c^2+d^2)=(ac-bd)^2+(ad+bc)^2.$$

Интересующая нас трактовка этого тождества такова: *произведение чисел, представимых в виде суммы двух квадратов, само представимо в виде суммы двух квадратов.* А раз так, то в силу основной теоремы арифметики вопрос сводится к описанию таких *простых* чисел, которые представимы в виде суммы двух квадратов.

Замечание. В III.2 мы доказали тождество Эйлера, которое утверждает, что норма чисел вида $a+b\sqrt{-k}$ обладает свойством мультипликативности. В случае $k=1$ тождество Эйлера превращается в тождество Диофанта:

$$\begin{gathered}(a^2+b^2)\cdot(c^2+d^2)=N(a+b\sqrt{-1})\cdot N(c+d\sqrt{-1})=\\ =N((a+b\sqrt{-1})\cdot(c+d\sqrt{-1}))=(ac-bd)^2+(ad+bc)^2,\end{gathered}$$

где $N(m+n\sqrt{-1})=m^2+n^2$ — норма гауссова числа.

Напомним, что, обсуждая методы решения нелинейных диофантовых уравнений, мы доказали, что уравнение

$$x^2+y^2=4k+3$$

не имеет решений, откуда мгновенно следует, что любое целое число вида $4k+3$ не представимо в виде суммы двух квадратов.

Разумеется, возникает вопрос, а что можно сказать про простые числа вида $4k+1$? Про те самые, заметим, которые появились в

теореме о $\sqrt{-1}$. Приведем несколько примеров

$$5 = 1^2 + 2^2,$$
$$13 = 2^2 + 3^2,$$
$$17 = 1^2 + 4^2,$$
$$29 = 2^2 + 5^2, \ldots$$

Имея в виду такие наблюдения, Ферма выдвинул гипотезу о том, что *любое* простое число вида $4k+1$ представимо в виде суммы двух квадратов. По своему обыкновению Ферма указал в письме в 1640 году, что доказательство ему известно, но его так и не привел. Первое известное доказательство было получено спустя 100 лет Эйлером. Впрочем, оно носит довольно технический характер и можно даже сказать, не вскрывает причины явления. Поэтому мы приведем доказательство, лишенное таких недостатков, которое было опубликовано немецким математиком Дедекиндом в 1894 году.

Теорема. *Любое простое число вида* $4k+1$ *представимо в виде суммы квадратов двух натуральных чисел.*

Доказательство. Рассмотрим простое число p вида $4k+1$. Из теоремы о $\sqrt{-1}$ следует, что сравнение

$$x^2 + 1 \equiv 0 \pmod{p}$$

имеет решение $x = a$. Значит, можно утверждать, что

$$(a^2 + 1) \vdots p.$$

Рассмотрим число a^2+1 как элемент кольца $\mathbb{Z}[\sqrt{-1}]$ гауссовых чисел и разложим его на множители (!!!).

$$a^2 + 1 = (a + \sqrt{-1}) \cdot (a - \sqrt{-1}) \vdots p.$$

Докажем теперь, что число p, как элемент $\mathbb{Z}[\sqrt{-1}]$, является *составным*. Действительно, в кольце гауссовых чисел выполнена основная теорема арифметики и ее следствия, в частности, лемма Евклида.

Таким образом, если бы p было простым, оно бы делило одно из чисел $(a+\sqrt{-1})$ или $(a-\sqrt{-1})$, что, очевидно, неверно. Откуда следует, что число p является *составным* в $\mathbb{Z}[\sqrt{-1}]$ и

$$p = z \cdot w,$$

где числа $z, w \in \mathbb{Z}[\sqrt{-1}]$ должны быть сопряжены, поскольку их произведение есть целое число. Тогда

$$p = (m + n\sqrt{-1}) \cdot (m - n\sqrt{-1}) = m^2 + n^2.$$

□

Из доказанной теоремы и тождества Диофанта вытекает результат для натуральных чисел (в одну сторону доказательство напрямую следует из тождества Диофанта, в другую — требует некоторых несложных технических рассуждений).

Следствие. *Натуральное число представимо в виде суммы квадратов двух целых чисел тогда и только тогда, когда любое простое число вида $4k+3$ входит в его разложение на простые множители в четной степени.*

Замечание. Спустя полвека после публикации доказательства Эйлера Лежандр доказал теорему о представимости натуральных чисел в виде суммы трех квадратов. Оказывается, что все натуральные числа, кроме имеющих вид $4^k(8m + 7)$, представимы в таком виде.

А какие натуральные числа можно представить в виде четырех квадратов? Оказывается, что любые! Соответствующую теорему доказал Лагранж в 1770 году.

2. Квадратичные вычеты и символы Лежандра

После того как мы исследовали сравнение $x^2 \equiv -1 \pmod{p}$, возникает естественная задача исследовать сравнение $x^2 \equiv a \pmod{p}$ в случае произвольного $a \in \mathbb{Z}_p$.

Определение. Если $a \in \mathbb{Z}_p$ таков, что сравнение $x^2 \equiv a \pmod{p}$ имеет решение, то a называется *квадратичный вычетом.*

В противном случае a называется *квадратичный невычетом.*

Замечание. Случай $a = 0$ мы не будем рассматривать в дальнейшем в силу его тривиальности.

Обратим внимание, что решение сравнения

$$x^2 - a \equiv 0 \pmod{p}$$

эквивалентно нахождению корней многочлена $x^2 - a$, заданного над *полем* $\mathbb{Z}_p$. В предыдущей главе мы доказали, что количество корней многочлена, рассматриваемого над полем, не превосходит его степени. Таким образом, можно утверждать, что рассматриваемое сравнение будет иметь не более двух решений.

Сколько в $\mathbb{Z}_p$ квадратичных вычетов?

Утверждение. *Среди ненулевых элементов $\mathbb{Z}_p$ ровно половина квадратичных вычетов и ровно половина квадратичных невычетов.*

Доказательство. Рассмотрим множество ненулевых остатков в $\mathbb{Z}_p$, т.е. $\mathbb{Z}_p^*$, и следующее отображение на этом множестве: $x \longmapsto x^2$. Его образ в силу определения состоит из множества квадратичных вычетов. Осталось вычислить, из какого количества элементов состоит образ.

Во-первых, очевидно, что для любого $x \in \mathbb{Z}_p^*$ x и $-x$ отображаются в один и тот же элемент $\mathbb{Z}_p^*$. Значит, в образе не более $\frac{p-1}{2}$ элементов.

Во-вторых, возможно ли, чтобы в какой-то элемент отобразилось бы более двух остатков? Предположим, что, например, в элемент q отобразилось больше двух элементов. Тогда сравнение $x^2 \equiv q \pmod{p}$ должно иметь больше двух решений, что невозможно. Значит, в образе в точности $\frac{p-1}{2}$ элементов. □

Как узнать, является ли данный элемент квадратичным вычетом? Иными словами, разрешимо ли сравнение $x^2 \equiv a \pmod{p}$? На данный вопрос дает ответ следующая теорема, доказанная Эйлером.

Теорема (критерий Эйлера)**.** *Пусть $p > 2$ — простое число. Ненулевой $a \in \mathbb{Z}_p$ является квадратичным вычетом тогда и только тогда, когда*

$$a^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Доказательство. 1. Докажем, что если a — квадратичный вычет, то выполнено сравнение $a^{\frac{p-1}{2}} \equiv 1 \pmod{p}$.
Действительно, найдется такой x_0, что $x_0^2 \equiv a \pmod{p}$. Тогда имеем

$$a^{\frac{p-1}{2}} = x_0^{p-1} \equiv 1 \pmod{p}$$

в силу малой теоремы Ферма.

2. Из первого пункта следует, что любой квадратичный вычет является решением сравнения

$$x^{\frac{p-1}{2}} \equiv 1 \pmod{p}.$$

Как мы знаем, квадратичных вычетов в точности $\frac{p-1}{2}$ штук. Таким образом, сравнение $x^{\frac{p-1}{2}} \equiv 1 \pmod{p}$ имеет по крайней мере $\frac{p-1}{2}$ решений, а по теореме о количестве корней оно имеет не более $\frac{p-1}{2}$ решений. Значит, решения — в точности $\frac{p-1}{2}$ квадратичных вычетов. □

В V.3, исследуя возможность решать диофантовы уравнения вида

$$\alpha a^2 + \beta b^2 = \gamma c^2,$$

мы пришли к необходимости ответить на следующий вопрос: существует ли у кривой

$$\alpha x^2 + \beta y^2 = \gamma$$

хотя бы одна рациональная точка? Ответ на него дает

Теорема (Лежандр)**.** *Кривая, заданная уравнением*

$$\alpha x^2 + \beta y^2 = \gamma,$$

где α, β и γ — попарно взаимно простые числа, имеет хотя бы одну рациональную точку тогда и только тогда, когда число $(-\alpha\beta)$

является квадратичным вычетом по модулю γ, число $\alpha\gamma$ — квадратичным вычетом по модулю β, а число $\beta\gamma$ — квадратичным вычетом по модулю α.

Таким образом, мы приходим к необходимости научиться эффективно вычислять, является ли данное число квадратичным вычетом по данному модулю.

Определение. Пусть a — целое число, p — простое число. Положим

$$\left(\frac{a}{p}\right) = \begin{cases} 0, & \text{если } a \vdots p; \\ 1, & \text{если } a \text{ — квадратичный вычет по модулю } p; \\ -1, & \text{если } a \text{ — квадратичный невычет по модулю } p. \end{cases}$$

Число $\left(\frac{a}{p}\right)$ называется символом Лежандра.

Замечание. Из теоремы о $\sqrt{-1}$ следует равенство

$$\left(\frac{-1}{p}\right) = \begin{cases} 1, & \text{если } \ p = 4k+1; \\ -1, & \text{если } \ p = 4k+3. \end{cases}$$

Утверждение.

$$\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \pmod{p}.$$

Доказательство. Прямо следует из критерия Эйлера. □

Следствие.

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

Следствие. *Если*

$$a = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

то

$$\left(\frac{a}{p}\right) = \left(\frac{p_1}{p}\right)^{\alpha_1 \pmod 2} \cdot \left(\frac{p_2}{p}\right)^{\alpha_2 \pmod 2} \cdot \ldots \cdot \left(\frac{p_k}{p}\right)^{\alpha_k \pmod 2}.$$

Вычислять символы Лежандра помогает следующее

Утверждение (Квадратичный закон взаимности)**.** *Пусть p и q — различные нечетные простые числа. Тогда*

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

Доказательство. Предположим сначала, что $p\equiv q\pmod 4$. Без ограничения общности можно считать, что $p>q$. Тогда положим $p-q=4a$. В таком случае $p=4a+q$ и мы имеем

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

Аналогично

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

Но $\left(\frac{a}{p}\right)=\left(\frac{a}{q}\right)$ одинаковы, поскольку $p-q=4a$. Значит,

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

Предположим, что $p\not\equiv q\pmod 4$. Тогда $p\equiv -q\pmod 4$. Положим $p+q=4a$. Тогда мы получаем

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

Аналогично

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

Но $\left(\frac{a}{p}\right)=\left(\frac{a}{q}\right)$ одинаковы, поскольку $p+q=4a$. Тем самым,

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

если p и q имеют различные остатки при делении на 4. □

В законе взаимности фигурируют только нечетные простые числа. С его помощью, однако, можно вычислить и $\left(\frac{2}{p}\right)$, который уже встретился нам в вычислениях.

Утверждение.

$$\left(\frac{2}{p}\right) = (-1)^{\frac{p^2-1}{8}}.$$

С помощью доказанных утверждений можно быстро вычислять, является ли a квадратичным вычетом по модулю m.

$$\left(\frac{983}{1103}\right) = -\left(\frac{1103}{983}\right) = -\left(\frac{120}{983}\right) =$$
$$-\left(\frac{2}{983}\right)^3 \cdot \left(\frac{3}{983}\right) \cdot \left(\frac{5}{983}\right) =$$
$$\left(\frac{983}{3}\right) \cdot \left(\frac{983}{5}\right) = \left(\frac{2}{3}\right) \cdot \left(\frac{3}{5}\right) = \left(\frac{2}{3}\right)^2 = 1.$$

Отметим, что в случае составного модуля, ситуация становится много сложнее и определение того, является ли данное число квадратичным вычетом, требует значительных технических усилий.

Глава IX

Доказательство основной теоремы арифметики

1. Случай натуральных чисел

В третьей главе вы познакомились с формулировкой *основной теоремой арифметики* и узнали, что этот, вроде бы, очевидный факт является на самом деле глубоким свойством натуральных чисел. Теперь мы готовы привести доказательство этой теоремы.

Теорема (Основная теорема арифметики)**.** *Каждое натуральное число n, большее 1, может быть разложено в произведение простых чисел:*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

причем это разложение единственно с точностью до порядка простых сомножителей.

Доказательство. Докажем *существование* такого разложения.
Если n — простое, то $n = n$ и разложение существует.
Если n — составное, то найдутся a, b такие что $n = a \cdot b$, $1 < a, b < n$.
Если a, b — простые, то разложение получено, иначе — продолжаем процесс: $a = k \cdot l$ и т.д. Т.к. $\mathbb{N}$ *ограничено снизу*, то процесс остановится. Существование доказано.

Теперь докажем *единственность* разложения.
Допустим противное. Выберем *наименьшее* число N, имеющее хотя бы два разложения:

$$N = p_1 p_2 \ldots p_n = q_1 q_2 \ldots q_m$$

Заметим, что $p_i \neq q_j$.
Почему мы можем это утверждать?

Действительно, если найдется такая пара, что $p_i = q_j$, то, сокращая, мы получим два разложения для натурального числа, которое меньше N, что противоречит минимальности N.

Предположим, что $p_1 < q_1$, и рассмотрим число

$$\begin{aligned} N' &= N - p_1 q_2 \dots q_m = p_1 p_2 \dots p_n - p_1 q_2 \dots q_m = \\ &= p_1 (p_2 \dots p_n - q_2 \dots q_m) = q_1 q_2 \dots q_m - p_1 q_2 \dots q_m = \\ &= (q_1 - p_1)\, q_2 \dots q_m \end{aligned}$$

Откуда следует, что $(q_1 - p_1)\, q_2 \dots q_m \vdots p_1$. Следовательно $(q_1 - p_1) \vdots p_1$. *Почему мы можем это утверждать?*
Это есть следствие основной теоремы арифметики, которая выполнена для числа $(q_1 - p_1)\, q_2 \dots q_m$, которое меньше N (ведь N обладает свойством минимальности!). Таким образом, $q_1 \vdots p_1$. А значит, $q_1 = p_1$. Противоречие. □

2. Евклидовы кольца и основная теорема арифметики

В течение нашего курса мы успели познакомиться с примерами множеств, которые похожи на целые числа и для которых также выполнена основная теорема арифметики. Но приведенное нами доказательство прямо не обобщается на эти случаи. Нам известны примеры колец, обладающих свойством факториальности, — кольцо многочленов от одной переменной $\mathbb{R}[x]$ (более общо $\Bbbk[x]$, где $\Bbbk$ — произвольное поле) и кольцо гауссовых чисел $\mathbb{Z}[\sqrt{-1}]$. Несмотря на очевидные отличия, эти кольца обладают глубоко скрытой общностью, которая фиксируется определением.

Определение. Кольцо A без делителей нуля, не являющееся полем, называется *евклидовым*, если существует функция

$$N : A \backslash \{0\} \to \mathbb{N} \cup \{0\},$$

называемая *нормой*, удовлетворяющая следующим условиям:

1) $N(ab) \geqslant N(a)$, причем равенство имеет место только тогда, когда элемент b обратим;

2) для любых $a, b \in A$, где $b \neq 0$, существуют такие $q, r \in A$, что $a = qb + r$ и либо $r = 0$, либо $N(r) < N(b)$.

Замечание. Условие 2) в определении нормы означает возможность деления с остатком и существование алгоритма Евклида.

Основные примеры евклидовых колец.

- $\mathbb{Z}$ с нормой $N(a) = |a|$;
- $\Bbbk[x]$ с нормой $N(a) = \deg a$
  (известные нам примеры — $\mathbb{Q}[x], \mathbb{R}[x]$ и $\mathbb{Z}_p[x]$);
- $\mathbb{Z}[\sqrt{-1}]$ с нормой $N(a + b\sqrt{-1}) = a^2 + b^2$.

Замечание. Строгое определение кольца $\mathbb{Z}[\sqrt{-k}]$ будет нами дано в разделе XI.6.

Доказательство того, что представленные функции воистину являются нормами, прямо следует из доказанных нами теорем о делении с остатком, за исключением выполнения условия 2) для нормы гауссова числа.

Утверждение (О делении с остатком в кольце гауссовых чисел)**.** *Для любых* $a, b \in \mathbb{Z}[\sqrt{-1}]$, *где* $b \neq 0$, *существуют такие* $q, r \in \mathbb{Z}[\sqrt{-1}]$, *что* $a = qb + r$ *и либо* $r = 0$, *либо* $N(r) < N(b)$.

Доказательство. Ключевыми для нас окажутся геометрические соображения. Будем изображать гауссовы числа точками на координатной плоскости с целыми координатами.

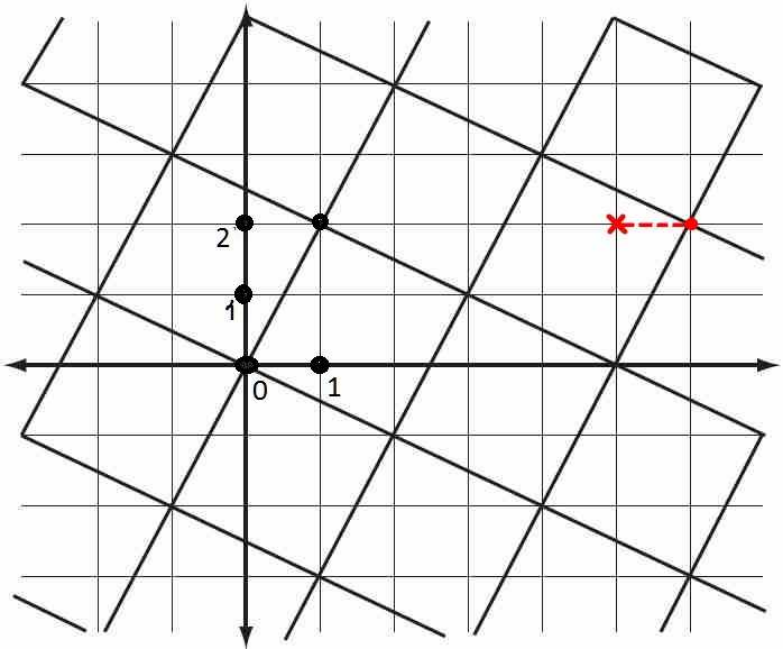
Покажем, как разделить с остатком a на b. Множество гауссовых чисел $k \cdot b$, кратных b, образует квадратную решетку со стороной $\sqrt{N(b)}$, где $N(b) = N(s + t\sqrt{-1}) = s^2 + t^2$.

Замечание. Вы можете убедиться в этом на конкретных примерах. Строгое доказательство утверждения о геометрической интерпретации умножения гауссовых чисел вы узнаете, когда познакомитесь с комплексными числами.

На рисунке ниже приведен пример квадратной решетки для $b = 1 + 2\sqrt{-1}$.

Число a попадает в один из образовавшихся квадратов (на рисунке ниже число a изображено крестиком). В качестве неполного частного q необходимо выбрать число, изображенное ближайшей к a вершиной квадратной решетки. Положим $r = a - qb$. Расстояние от точки внутри квадрата до ближайшей вершины строго меньше длины стороны квадрата, откуда следует

$$\sqrt{N(r)} < \sqrt{N(b)} \quad \Rightarrow \quad N(r) < N(b).$$



Деление с остатком в кольце $\mathbb{Z}[\sqrt{-1}]$.

□

Для доказательства основной теоремы нам необходимо вспомогательное утверждение.

Лемма (Евклид)**.** *Если простой элемент p евклидова кольца делит произведение $a_1 a_2 \dots a_n$, то он делит хотя бы один из сомножителей $a_1, a_2, \dots, a_n$.*

Замечание. В главе III мы рассматривали данную лемму как следствие основной теоремы арифметики. Однако, на самом деле, она может быть доказана без нее. И более того, необходима для доказательства самой основной теоремы в общем случае!

Доказательство. Достаточно доказать лемму для двух элементов. При $n = 2$ предположим, что $a_1 \not\vdots p$. В таком случае НОД$(p, a_1) = 1$, что следует из алгоритма Евклида.

Используя лемму о представлении НОДа (см. главу II), получаем равенство $pu + a_1v = 1$. Умножая обе части на a_2, имеем

$$pua_2 + a_1a_2v = a_2.$$

При этом $pua_2 \vdots p$ и $a_1a_2v \vdots p$, следовательно $pua_2 + a_1a_2v = a_2 \vdots p$.

При $n > 2$ произведение $a_1a_2 \dots a_n$ можно представить в виде $a_1 \cdot (a_2 \dots a_n)$. По только что доказанному или $a_1 \vdots p$, или $a_2 \dots a_n \vdots p$. Во втором случае, продолжая аналогичные рассуждения, получаем, что $a_i \vdots p$, где i — один из индексов $2, \dots, n$. □

Теорема (Основная теорема арифметики для евклидовых колец)**.** *В евклидовом кольце всякий необратимый ненулевой элемент n может быть разложен в произведение простых множителей:*

$$n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdot \ldots \cdot p_k^{\alpha_k},$$

причем это разложение единственно с точностью до перестановки простых множителей и умножения их на обратимые элементы.

Доказательство. 1. Докажем *существование* такого разложения.

Если n — простой, то $n = n$ и разложение существует. Если n — составной, то найдутся *необратимые* a, b такие, что $n = a \cdot b$. А значит, $N(a), N(b) < N(a \cdot b) = N(n)$. Если a, b — простые, то разложение получено, иначе — продолжаем процесс: $a = k \cdot l$ и т.д. Т.к. норма принимает только неотрицательные значения, т.е. *ограничена снизу*, то процесс остановится. Существование доказано.

2. Теперь докажем *единственность* разложения.

Допустим противное. Выберем элемент M с *наименьшей нормой*, имеющий хотя бы два разложения:

$$M = p_1p_2 \dots p_n = q_1q_2 \dots q_m.$$

Из условия минимальности нормы следует, что $p_i \neq q_j$. Из равенства следует, что $q_1q_2 \dots q_m \vdots p_i$. В силу леммы Евклида отсюда следует

существование такого j, что $q_j \vdots p_i$. Это означает, что $q_j = c \cdot p_i$, где c — обратимый элемент. Сокращая равенство $p_1 p_2 \dots p_n = q_1 q_2 \dots q_m$ на p_i приходим к противоречию с минимальностью нормы M.

□

Среди рассмотренных нами множеств встречались и такие, которые не обладают свойством факториальности. Рассмотрим теперь соответствующие примеры и выясним, почему данное доказательство нельзя перенести на них.

Пример Яглома

Напомним, что рассматривается множество четных чисел:

$$2, \quad 4, \quad 6, \quad 8, \quad 10 \quad \dots$$

В третьей главе доказано следующее

Утверждение. *Все четные числа, кратные 4, являются составными, а все некратные 4 — простыми.*

Рассмотрим *наименьшее* число, имеющее *хотя бы два* разложения на простые. Имеем

$$M = 36 = p_1 \cdot p_2 = 2 \cdot 18 = \\ = q_1 \cdot q_2 = 6 \cdot 6.$$

$6 \cdot 6 \vdots 2$, но отсюда не следует, что 6 делится на 2! Лемма Евклида не имеет места для кольца четных чисел! Дело в том, что в этом кольце отсутствует 1, поэтому не выполнено условие 2) для нормы $N(n) = |n|$. (Например, не существует таких четных q и r, что $8 = 6q + r$).

Пример Гильберта

Напомним, что рассматривается множество чисел вида $4k + 1$:

$$1, \quad 5, \quad 9, \quad 13, \quad 17, \quad 21 \quad \dots$$

Рассмотрим *наименьшее* число, имеющее *хотя бы два* разложения на простые. Имеем

$$M = 441 = p_1 \cdot p_2 = 9 \cdot 49 = \\ = q_1 \cdot q_2 = 21 \cdot 21.$$

Аналогично примеру Яглома из того, что $21 \cdot 21 \vdots 9$, не следует, что 21 делится на 9. Более того, множество чисел вида $4k+1$ не является кольцом, поскольку их нельзя складывать. В таком случае не приходится говорить о выполнении леммы Евклида.

Кольцо $\mathbb{Z}[\sqrt{-3}]$

Рассматриваются числа вида:

$$a + b\sqrt{-3}, \quad a, b \in \mathbb{Z}.$$

Рассмотрим теперь число с *наименьшей нормой*, имеющее *хотя бы два* разложения на простые. Имеем

$$M = 4 = p_1 \cdot p_2 = 2 \cdot 2 = \\ = q_1 \cdot q_2 = (1 + \sqrt{-3}) \cdot (1 - \sqrt{-3}).$$

Легко видеть, что $(1 \pm \sqrt{-3})$ не делится на 2. Оказывается, для нормы $N(a+b\sqrt{-3}) = a^2+3b^2$ не выполнено условие 2) и отсутствует алгоритм Евклида.

Необходимо отметить, что существование алгоритма Евколида является достаточным, но не необходимым условием для выполнения основной теоремы арифметики.

Существуют факториальные кольца (т.е. такие, где выполнена основная теорема арифметики), которые не являются евклидовыми. Таким, например, является кольцо многочленов от многих переменных $\Bbbk[x_1, x_2, \ldots, x_n]$.

Глава X

Аксиомы Пеано

Дорогой читатель, вы проделали немалый путь в изучении чисел и их свойств, познакомились с необычными арифметиками, научились решать непростые задачи. Пришло время ответить на следующий вопрос:

Что есть натуральное число?

Натуральные числа — видимо, первый математический объект, который был принят человеком на самой заре нашей цивилизации многие тысячи лет тому назад (хотя по-настоящему математическим он стал совсем недавно — чуть более ста лет назад!).

Как мы сказали в предисловии, во второй половине XIX века родился *формальный язык.* Частью этого формального языка являются *аксиомы.*

В своей книге «Основания геометрии» Гильберт объяснял, что прямая, точка и плоскость появляются только в связи с теми аксиомами, которые для них выбираются. Другими словами, назвать ли их точками, прямыми, плоскостями или же столами, стульями, пивными кружками, — это будут те объекты, для которых справедливы соотношения, выражаемые аксиомами.

В некотором смысле это похоже на то, как значение неизвестного слова проясняется по мере использования его в различных контекстах. Каждое дополнительное предложение, в котором оно участвует, исключает некоторые значения, которое могло бы иметь это слово в предыдущих предложениях. Иными словами, ***аксиомы играют роль и определений!***

Точно так же, как и в геометрии, существую аксиомы, которые отвечают на вопрос, поставленный в начале главы — дают определение

наутральных чисел. Ключевым в этих аксиомах является отношение «следовать за».

I. 1 является натуральным числом.
II. За каждым натуральным числом *следует* одно и только одно число, являющееся натуральным.
III. 1 *не следует* ни за каким натуральным числом.
IV. Каждое натуральное число, отличное от 1, *следует* за одним и только одним натуральным числом.
V. Подмножество натуральных чисел, содержащее число 1, а вместе с каждым натуральным числом и *следующее* за ним число, содержит все натуральные числа.

В математической нотации эти аксиомы записываются следующим образом, где отношение «следовать за» формализуется с помощью введения *функции следования* $S : \mathbb{N} \to \mathbb{N}$:

I. $1 \in \mathbb{N}$.
II. $\forall n \in \mathbb{N} \quad \exists!\, m = S(n) \in \mathbb{N}$.
III. $\nexists n \in \mathbb{N} : S(n) = 1$.
IV. $\forall m \in \mathbb{N}, m \neq 1 \quad \exists!\, n \in \mathbb{N} : m = S(n)$.
V. Если $1 \in \mathbb{M} \subset \mathbb{N}$ и $\forall n \in \mathbb{N}$ из $n \in \mathbb{M}$ следует $S(n) \in \mathbb{M}$, то $\mathbb{N} \subset \mathbb{M}$.

Замечание. Последняя аксиома называется *аксиомой индукции.* Из нее следует справедливость метода математической индукции: если какое-либо предложение доказано для 1 (база индукции) и если из допущения, что оно верно для натурального числа n, вытекает, что оно верно для следующего за n натурального числа (шаг индукции), то это предположение верно для всех натуральных чисел. Чтобы увидеть индукцию в пятой аксиоме, необходимо в качестве $\mathbb{M}$ рассмотреть множество таких натуральных чисел, для которых предложение индукции верно.

Возникает вопрос, а как складывать натуральные числа?

Определение. Операция сложения на множестве $\mathbb{N}$ задается двумя соотношениями, которые корректны в силу аксиом:

1. $n+1=S(n)$;
2. $n+S(m)=S(n+m)$.

В качестве примера сложим 5 и 3:

$$5{+}3 = 5{+}S(2) = S(5{+}2) = S(5{+}S(1)) = S(S(5{+}1)) = S(S(S(5))) = 8.$$

Еще с начальной школы всем известны правила обращения с натуральными числами. В частности, известное выражение «от перестановки слагаемых сумма не меняется» призвано зафиксировать известный всем факт, что $5{+}3 = 3{+}5$. Однако, из формального определения суммы натуральных чисел это не следует! Это утверждение является теоремой, требующей доказательства!

Теорема. *Для любых натуральных m и n выполнено $m{+}n = n{+}m$, т.е. сложение натуральных чисел коммутативно.*

Для доказательства теоремы понадобится следующая

Лемма. *Для любых $m, n \in \mathbb{N}$ выполнено*

$$m + S(n) = S(m) + n.$$

Доказательство. Для доказательства воспользуемся аксиомой V и проведем индукцию по n. *Зафиксируем произвольное* натуральное число m.

1. База индукции. В случае $n = 1$ имеем

$$\begin{aligned} m + S(1) &= S(m+1) && \text{по определению сложения} \\ &= S(S(m)) && \text{по определению сложения с 1} \\ &= S(m) + 1 && \text{по определению сложения с 1.} \end{aligned}$$

2. Шаг индукции. Докажем, что из равенства $m{+}S(n) = S(m){+}n$ следует равенство $m + S(S(n)) = S(m) + S(n)$. Имеем

$$\begin{aligned} &m + S(n) = S(m) + n \;\Leftrightarrow\; \text{в силу аксиом II, IV} \\ &S(m + S(n)) = S(S(m) + n) \;\Leftrightarrow\; \text{по определению сложения} \\ &m + S(S(n)) = S(m) + S(n). \end{aligned}$$

Значит, утверждение верно для любого $n \in \mathbb{N}$. Поскольку мы фиксировали *произольное* $m \in \mathbb{N}$, формула верна для любых натуральных m и n. □

Теперь мы готовы дать доказательство теоремы.

Доказательство. Как и для доказательства леммы, воспользуемся аксиомой V и проведем индукцию по n.

1. База индукции. Докажем, что для произвольного m имеет место равенство

$$m + 1 = 1 + m.$$

Доказательство проведем индукцией по m. Очевидно, $1 + 1 = 1 + 1$. Предположим, что выполнено равенство $k + 1 = 1 + k$. Докажем, что $S(k) + 1 = 1 + S(k)$. Имеем

$$\begin{aligned}
&k + 1 = 1 + k \ \Leftrightarrow \text{ в силу аксиом II, IV} \\
&S(k + 1) = S(1 + k) \ \Leftrightarrow \text{ по определению сложения} \\
&k + S(1) = 1 + S(k) \ \Leftrightarrow \text{ по лемме} \\
&S(k) + 1 = 1 + S(k).
\end{aligned}$$

Таким образом, следуя аксиоме V, заключаем, что равенство $m+1 = 1 + m$ справедливо для любого $m \in \mathbb{N}$.

2. Шаг индукции. Зафиксируем *произвольное* натуральное m и докажем, что из равенства $m + n = n + m$ следует равенство

$$m + S(n) = S(n) + m.$$

Имеем

$$\begin{aligned}
&m + n = n + m \ \Leftrightarrow \text{ в силу аксиом II, IV} \\
&S(m + n) = S(n + m) \ \Leftrightarrow \text{ по определению сложения} \\
&m + S(n) = n + S(m) \ \Leftrightarrow \text{ по лемме} \\
&m + S(n) = S(n) + m.
\end{aligned}$$

Значит, утверждение верно для любого $n \in \mathbb{N}$. Поскольку мы фиксировали *произольное* $m \in \mathbb{N}$, формула верна для любых натуральных m и n. □

Подобным образом доказываются другие известные свойства сложения натуральных чисел.

Аналогично сложению дается определение умножения натуральных чисел.

Определение. Операция умножения на множестве $\mathbb{N}$ задается двумя соотношениями, которые корректны в силу аксиом:

1. $n \cdot 1 = n$;
2. $n \cdot S(m) = n \cdot m + n$.

Можно доказать, что так определенная операция умножения обладает всеми привычными свойствами.

В заключении отметим, что, разумеется, никто не использует упоминаемые выше теоремы при работе с натуральными числами. Главное, что они демонстрируют, — какой высокой оказалась цена за отказ от соблазнов наглядной интерпретации математических объектов. Чего мы достигли, отказавшись от наглядной интуиции? Построение математики на строгой формальной основе открыло невиданные ранее возможности для изучения и описания реального мира! Специальная и общая теории относительности, квантовая механика, статистическая физика, квантовая теория поля, теория струн — все эти поражающие воображение теории не могли бы быть созданы без новой математики.

Глава XI

Что есть число?

В предыдущей главе мы смогли дать определение натурального числа. Теперь необходимо двигаться дальше — построить кольцо целых чисел $\mathbb{Z}$, сделав -1 научным термином, дать определение поля $\mathbb{Q}$ рациональных чисел и поля $\mathbb{R}$ вещественных чисел. Но прежде необходимо познакомиться с одной из ключевых математических идей, которой неявно мы уже пользовались при построении колец остатков.

1. $\mathbb{Z}_m$ как кольцо классов вычетов

Как мы помним, преимущество рассмотрения остатков заключается в том, что если мы фиксируем делитель m, то существует лишь *конечное* множество остатков от деления чисел на m. При этом каждый остаток получается при делении на m *бесконечного* набора чисел.

Итак, рассмотрим произвольное натуральное числа a. Заметим, что условие «a дает при делении на m остаток r» эквивалентно следующей формуле: $a = mt + r$, где $t \in \mathbb{Z}$. Пусть в этой формуле t пробегает все множество целых чисел, в то время как m и r фиксированы. Тогда наша формула дает *все* целые числа, для которых остаток от деления на m равен r.

⋮	⋮	⋮	⋮	⋮	⋮	⋮
-21	-20	-19	-18	-17	-16	-15
-14	-13	-12	-11	-10	-9	-8
-7	-6	-5	-4	-3	-2	-1
0	**1**	**2**	**3**	**4**	**5**	**6**
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
⋮	⋮	⋮	⋮	⋮	⋮	⋮

В таблице приведен пример для $m = 7$. Числа, которые дают одинаковый остаток при делении на 7, расположены в одном столбце.

Таким образом, если задано натуральное число m, то все множество целых чисел можно *разбить* на m *классов*: к одному классу отнести все числа, дающие при делении на m остаток 0, к другому — остаток 1, и так далее. Например, если $m = 7$, то всего классов 7: $7t$, $7t+1$, $\ldots$, $7t+6$. Заметим, что *каждое число обязательно попадает в один и только в один класс.*

Замечание. Обратите внимание на то, что слово «разбиение» из нашего бытового языка только что стало полноправным математическим термином! Разбить множество означает указать такой набор его подмножеств, что

1) Каждый элемент попадает в какое-нибудь подмножество (т.е. подмножества покрывают все множество).

2) Каждый элемент попадает только в одно подмножество (т.е. подмножества не пересекаются).

Определение. Рассмотрим множество целых чисел. Договоримся не различать два целых числа a и b (считать их эквивалентными), если они принадлежат к одному классу. Иными словами, a и b эквивалентны, если $a \equiv b \pmod m$.

Итак, множество целых чисел разбивается на классы, которые будем называть *множеством классов вычетов по модулю* m. Его принято обозначать $\mathbb{Z}/m\mathbb{Z}$. Числа, лежащие в этих классах, называются вычетами.

На самом деле мы только что определили не что иное, как $\mathbb{Z}_m$, т.к. каждому остатку соответствует свой класс вычетов, и наоборот. Как тяжело дались и как дорого стоят слова в этом определении: «договоримся не различать»! Так родился прием, который называется *факторизацией* и который в XX веке очень широко использовался. Именно этот прием позволит нам дать строгое определение целых и рациональных чисел.

Замечание. Обозначения $\mathbb{Z}_m$ и $\mathbb{Z}/m\mathbb{Z}$ взаимозаменяемы. Последнее обычно используют, чтобы подчеркнуть, что кольцо остатков определяется с помощью факторизации кольца целых чисел по отношению эквивалентности, которое мы ввели в определении выше. В таком случае его еще называют *факторкольцом*.

Напомним, что, впервые рассматривая множество остатков $\mathbb{Z}_m$ в главе IV, мы несколько неуклюже ввели на нем стандартные арифметические операции — сложение и умножение. Неуклюже, поскольку для их определения нам потребовалось «выйти за пределы» $\mathbb{Z}_m$. Факторизация доставляет стандартную процедуру для определения операций на полученном фактормножестве. Для этого вновь обратимся к таблице.

⋮	⋮	⋮	⋮	⋮	⋮	⋮
-21	-20	-19	-18	-17	-16	-15
-14	-13	-12	-11	-10	-9	-8
-7	-6	-5	-4	-3	-2	-1
0	**1**	**2**	**3**	**4**	**5**	**6**
7	8	9	10	11	12	13
14	15	16	17	18	19	20
21	22	23	24	25	26	27
⋮	⋮	⋮	⋮	⋮	⋮	⋮

Классы вычетов соответствуют столбцам данной таблицы. Предположим, мы хотим сложить предпоследний и последний ее столбцы, т.е. классы $7t + 5$ и $7t + 6$. Выберем для этого по одному *представителю* из каждого класса. Например, -16 и 27. Сложим их: $-16 + 27 = 11$. Теперь определим, какой класс представляет число 11. Как несложно видеть, это класс $7t + 4$. Тогда суммой классов $7t + 5$ и $7t + 6$ назовем класс $7t + 4$.

В общем случае конструкция совершенно аналогична. Введем для удобства следующее обозначение: класс вычетов, который представлен вычетом c, будем обозначать $[c]_m$.

Замечание. Мы уже не раз отмечали, что сравнение $\equiv$ похоже на равенство $=$. Введенные нами понятия позволяют вскрыть причину этой похожести. Действительно,

$$a \equiv b \pmod{m} \quad \Leftrightarrow \quad [a]_m = [b]_m.$$

Рассмотрим два класса $A, B \in \mathbb{Z}/m\mathbb{Z}$. Выберем в каждом по представителю $a \in A, b \in B$. Таким образом, $[a]_m = A, [b]_m = B$.

Определение. *Суммой* классов $[a]_m$ и $[b]_m$ называется класс $[a+b]_m$.
Произведением классов $[a]_m$ и $[b]_m$ называется класс $[a \cdot b]_m$.

В рассмотренном нами примере получаем

$$[-16]_7 + [27]_7 = [-16 + 27]_7 = [11]_7 = [4]_7,$$

$$[-16]_7 \cdot [27]_7 = [(-16) \cdot 27]_7 = [-432]_7 = [5]_7.$$

Возникает естественный вопрос: а не зависит ли результат сложения (произведения) классов от выбора представителя?

Утверждение (Проверка корректности)**.** *Определения суммы и произведения классов вычетов корректно, т.е. не зависит от выбора представителя.*

Доказательство. Пусть

$$a \equiv a' \pmod{m}, \quad b \equiv b' \pmod{m}.$$

Тогда

$$a + b \equiv a' + b' \pmod{m}, \quad \text{т.е. } [a + b]_m = [a' + b']_m$$

и, аналогично,

$$a \cdot b \equiv a' \cdot b' \pmod{m}, \quad \text{т.е. } [a \cdot b]_m = [a' \cdot b']_m.$$

□

Теперь мы можем утверждать, что построенное множество $\mathbb{Z}/m\mathbb{Z}$ воистину является кольцом, естественно изоморфным кольцу $\mathbb{Z}_m$.

2. Что есть целое число?

Математики примирились с отрицательными числами, к которым вы так привыкли, совсем недавно. Минус единица стала научным термином только во второй половине XIX (!) века. Ни Декарт, Ферма, Лейбниц, Эйлер, Лагранж, ни Гаусс или Коши не могли дать определения отрицательных чисел. Еще в XVIII веке среди ведущих математиков имела место оживленная дискуссия о природе отрицательных чисел. Рассмтрим, например, следующее равенство:

$$\frac{-1}{1} = \frac{1}{-1}$$

Его справедливость не вызывает никаких вопросов. Однако, слева в этом равенстве стоит отношение меньшего числа к большему, а справа, наоборот — большего к меньшему. Как же можно ставить равентсво между этими отношениями?! Даже столь привычный для нас объект, как отрицательное число, *не имеет конкретной наглядной интерпретации.*

Дать аккуратное определение множества $\mathbb{Z}$ целых чисел нам поможет идея факторизации. Рассмотрим множество пар натуральных чисел $\{\ (m,n) \mid m \in \mathbb{N}, n \in \mathbb{N}\}$.

Определение. Пары (m,n) и (k,l) будем считать *эквивалентными*, если $m+l = n+k$.

Приведем примеры эквивалентных пар (знаком $\sim$ мы обозначаем эквивалентность, это аналог $\equiv$ из теориии сравнений):

$$(1,1) \sim (2,2) \sim (3,3) \sim \dots$$
$$(2,1) \sim (3,2) \sim (4,3) \sim \dots$$
$$(1,2) \sim (2,3) \sim (3,4) \sim \dots$$

Множество эквивалентных пар образует класс (по аналогии с множеством чисел, которые дают одинаковый остаток при делении на заданное число). Правда, в отличие от теории сравнений, мы получаем бесконечное множество классов.

Определение. Целым числом называется класс эквивалентных пар.

Множество $\mathbb{N}$ естественным образом вкладывается в $\mathbb{Z}$. Это вложение задается формулой

$$n \mapsto [n+1, 1],$$

где через $[n+1, 1]$ обозначен класс эквивалентности, содержащий пару $(n+1, 1)$.

Замечание. Класс эквивалентности, содержащий пару (m, n) с $m > n$, есть не что иное, как натуральное число $m - n$.

Определение. *Нулем* называется класс $[m, m]$ (обратите внимание на то, что ноль определяется)!

Как складывать и умножать целые числа?

Определение. *Суммой* целых чисел $[m, n]$ и $[k, l]$ называется целое число $[m + k, n + l]$.

Произведением целых чисел $[m, n]$ и $[k, l]$ называется целое число $[mk + nl, ml + nk]$.

Совсем несложно убедиться, что данные определения корректны, т.е. не зависят от выбора представителей.

Теперь мы можем решить уравнение $1 + x = 0$ и тем самым определить минус единицу! В самом деле, $1 = [2, 1]$. Положим $x = [1, 2]$ и получим, что

$$[2, 1] + [1, 2] = [3, 3], \text{ т.е. } 1 + x = 0.$$

Таким образом, $-1 = [1, 2]$, а класс эквивалентности, содержащий пару (m, n) с $m < n$, есть *отрицательное число* $m - n$.

Замечание. Описанная конструкция работает не только для целых чисел, но и в более общих ситуациях, когда требуется «научиться приписывать знак минус», иными словами, вычитать.

3. Что есть рациональное число?

Понятие дроби или рационального числа возникло несколько тысяч лет назад (намного раньше, чем понятие об отрицательных числах), когда, сталкиваясь с необходимостью измерять некоторые вещи (длину, вес, площадь и т. п.), люди поняли, что не удается обойтись целыми числами и необходимо ввести понятие доли: половины, трети и т. п. Иными словами, необходимо было построить числовое множество, в котором возможно делить элементы. Строгое же определение также было дано совсем недавно — чуть более ста лет назад.

Рассмотрим множество пар, состоящих из целых чисел
$\{\ (m,n) \mid m \in \mathbb{Z}, n \in \mathbb{Z}, n \neq 0\}$.

Определение. Пары (m,n) и (k,l) будем считать *эквивалентными*, если $m \cdot l = n \cdot k$.

Приведем примеры эквивалентных пар.

$$(1,1) \sim (2,2) \sim (3,3) \sim \dots$$
$$(2,1) \sim (4,2) \sim (6,3) \sim \dots$$
$$(1,2) \sim (2,4) \sim (3,6) \sim \dots$$

Определение. Рациональным числом называется класс эквивалентных пар.

Множество $\mathbb{Z}$ естественным образом вкладывается в $\mathbb{Q}$. Это вложение задается формулой

$$n \mapsto [n,1],$$

где через $[n,1]$ обозначен класс эквивалентности, содержащий пару $(n,1)$.

Замечание. Класс эквивалентности, содержащий пару (m,n), есть не что иное, как дробь $\frac{m}{n}$.

Как складывать и умножать рациональные числа?

Определение. *Суммой* рациональных чисел $[m, n]$ и $[k, l]$ называется рациональное число $[ml + nk, nl]$.

Произведением рациональных чисел $[m, n]$ и $[k, l]$ называется рациональное число $[mk, nl]$.

Совсем несложно убедиться, что данные определения корректны, т.е. не зависят от выбора представителей.

Класс $[k, k]$ является числом (рациональным) 1. Действительно, для любого рационального числа $[m, n]$ имеем

$$[m, n] \cdot [k, k] = [mk, nk] = [m, n].$$

Как мы знаем, множество $\mathbb{Q}$ является *полем*, т.е. в нем не просто существуют операции сложения и умножения, имеющие известные свойства, но также каждый ненулевой элемент имеет обратный. Действительно, рассмотрим ненулевое рациональное число $[m, n]$. Имеем

$$[m, n] \cdot [n, m] = [mn, mn] = 1.$$

В таком случае можно утверждать, что $[m, n]^{-1} = [n, m]$.

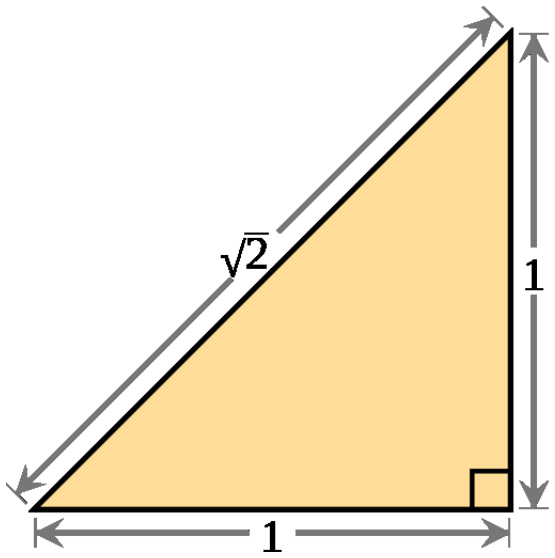
Замечание. Абсолютно аналогичные построения можно провести не только для кольца $\mathbb{Z}$, но и для любого кольца без делителей нуля. В таком случае построенное поле называется *полем частных* исходного кольца. В частности, если в качестве исходного выбрать колько многочленов от одной переменной $\Bbbk[x]$, то получится так называемое поле рациональных дробей, обозначаемое $\Bbbk(x)$.

Мы построили следующую цепочку вложенных числовых множеств

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q}.$$

Каждый шаг был мотивирован совершенно ясной потребностью: научиться вычитать и научиться делить. Может ли возникнуть причина продложать эту цепочку? Древнегреческие математики верили, что при измерении достаточно рациональных чисел. Действительно, с помощью рациональных чисел возможно построение сколь угодно

малых величин, выражаемых рациональными числами вида $1/n$. Этот факт убеждает, что рациональными числами можно измерить вообще любые геометрические расстояния. Совершенно потрясающим открытием (школа Пифагора) явилось существование величин, которые не могут быть измерены рациональными числами! Греки были настолько поражены, что засекретили это открытие, поскольку оно разрушало гармонично построенную ими теорию чисел, а значит, и всего сущего, ведь, как они считали, все сущее есть число! Пример появился с неожиданной стороны — из теоремы Пифагора. Если рассмотреть квадрат со стороной 1, то длина его диагонали будет равна $\sqrt{2}$.



Теорема (О $\sqrt{2}$)**.** *Число* $\sqrt{2}$ *не является рациональным.*

Доказательство. Предположим противное. Пусть существует такое рациональное число, т.е. *нескратимая* дробь $\frac{m}{n}$, что $\frac{m}{n} = \sqrt{2}$. Тогда

$$\frac{m}{n} = \sqrt{2} \;\Leftrightarrow\; \frac{m^2}{n^2} = 2 \;\Leftrightarrow\; m^2 = 2n^2.$$

Откуда следует, что число m четное, т.е. $m = 2k$. В таком случае

$$(2k)^2 = 2n^2 \;\Leftrightarrow\; 4k^2 = 2n^2 \;\Leftrightarrow\; 2k^2 = n^2.$$

Откуда следует, что число n четное, т.е. дробь $\frac{m}{n}$ *сократима*. Противоречие. □

4. Что есть вещественное число?

Итак, появилась необходимость продолжить цепочку

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \ ?$$

Известно, что Евдоксом Книдским была построена теория чисел, включающая несоизмеримые, т.е. иррациональные величины. И только спустя две тысячи лет люди вернулись к исследованию этих вопросов. Только во второй половине XIX века, когда развитие математики потребовало перестройки самих основ на новом уровне строгости, и стал зарождаться формальный язык. В работах Карла Вейерштрасса, Рихарда Дедекинда, Георга Кантора и других была построена строгая теория вещественных чисел. Далее мы опишем конструкцию, принадлежащую Вейерштрассу, которая, кстати, появилась до строгого определения натуральных и целых чисел.

Определение. Положительным вещественным числом α называется бесконечная десятичная дробь

$$\alpha = n_0, n_1 n_2 \dots n_k \dots .$$

На первый взгляд, это определение кажется более простым, чем рассмотренные нами ранее. Однако, за ним стоит история длинною в две тысячи лет. Ключевым словом в этом определении является «бесконечность».

5. Бесконечность в математике

Пифагорейцам было известно, что длина диагонали квадрата не выражается никаким рациональным числом. Как они открыли этот удивительный факт? Вряд ли им было доступно приведенное выше доказательство, поскольку алгебраическая техника тогда не была достаточно развита. Приведем геометрическое доказательство, которое вполне могло быть известно последователям Пифагора.

Предположим противное. А именно, что существует такое количество частей n, на которое мы разделим сторону единичного квадрата,

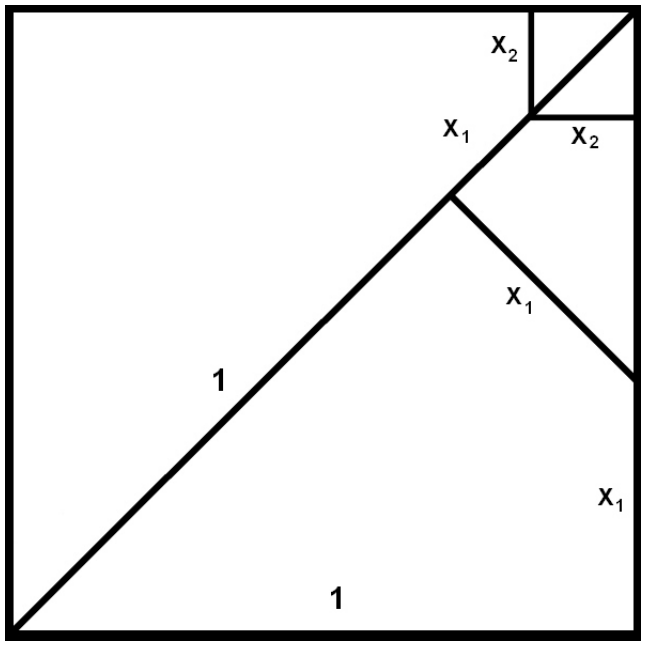
и получим отрезок длины $\frac{1}{n}$, который целое число раз укладывается на диагонали, т.е. длина диагонали будет равна $\frac{m}{n}$. Иными словами, мы предполагаем, что сторона квадрата и его диагональ *соизмеримы*.

Отложим сторону квадарата на диагонали, разделив ее на два отрезка длины 1 и x_1. Построим новый квадрат со стороной x_1 на части диагонали исходного. Легко видеть, что диагональ нового квадрата равна $1 - x_1$. Тогда имеем следующее неравенство

$$x_1 < 1 - x_1 \quad \Leftrightarrow \quad x_1 < \frac{1}{2}.$$

С новым квадратом проведем аналогичную процедуру и отложим его сторону на его диагонали, разделив ее на два отрезка длины x_1 и x_2. Построим следующий квадрат со стороной x_2, диагональ которого равна $x_1 - x_2$. Имеем

$$x_2 < x_1 - x_2 \quad \Leftrightarrow \quad x_2 < \frac{1}{2} \cdot x_1 < \frac{1}{2^2}.$$



Геометрическое доказательство иррациональности $\sqrt{2}$.

Продолжая далее, на k-ом шаге получим отрезок длины

$$x_k < \frac{1}{2^k}.$$

Из предположения и построения следует, что в каждый из отрезков $x_1, x_2, \ldots, x_k, \ldots$ отрезок $1/n$ умещается целое число раз. Нами получено противоречие, поскольку, начиная с некоторого k_0, величина $1/2^k$ не превосходит $1/n$.

Замечание. Описанная выше процедура представляет собой не что иное, как геометрический аналог алгоритма Евклида. В отличие от того, который мы рассматривали в главе II, этот алгоритм является бесконечным, что и приводит нас к противоречию.

Нельзя исключать того, что во времена Пифагора делались попытки дать определение вещественного числа, наподобие того, которое было дано Вейерштрассем спустя более чем две тысячи лет, — измерять отрезок с любой наперед заданной точностью. Утверждать этого мы не можем, поскольку не сохранилось никаких письменных упоминаний об этом, но мы вправе задать себе следующий вопрос. В силу каких причин такое определение появилось только лишь спустя две тысячи лет?

В V веке до н.э. Зенон Элейский формулирует свои знаменитые апории, из который нам наиболее известны благодаря упоминанию у Аристотеля следующие четыре: «Ахиллес и черепаха», «Дихотомия», «Летящая стрела» и «Стадион». Они объединены общим лейтмотивом: «нельзя впускать бесконечность в математику». Приведем примеры таких рассуждений. Допустим, Ахиллес бежит в десять раз быстрее, чем черепаха, и находится позади нее на расстоянии в тысячу шагов. За то время, за которое Ахиллес пробежит это расстояние, черепаха в ту же сторону проползет сто шагов. Когда Ахиллес пробежит сто шагов, черепаха проползет еще десять шагов, и так далее. Процесс будет продолжаться до бесконечности, Ахиллес так никогда и не догонит черепаху. В «Дихотомии» Зенон утверждает, чтобы преодолеть путь, нужно сначала преодолеть половину пути, а чтобы преодолеть половину пути, нужно сначала преодолеть половину половины, и так до бесконечности. Поэтому движение никогда не начнется. А летящая стрела не может вовсе лететь, так как в

каждый момент времени она покоится, а поскольку она покоится в каждый момент времени, то она покоится всегда. Иными словами, нельзя пройти бесконечное число интервалов за конечное время! Не исключено, что подобные рассуждения Зенон приводил в качестве аргументов против в том числе попыток давать определения в духе Вейерштрасса.

Работы Зенона нам известны благодаря Аристотелю. Ему же принадлежит следующий анализ понятия бесконечности. Бесконечность может рассматриваться как неограниченность некоторого процесса, например, когда утверждается возможность продолжить бесконечно и непрерывно любую прямую, то имеется в виду, что процесс можно непрерывно продолжать, но существование такого самостоятельного объекта, как бесконечная прямая, из него не следует. Такого рода процессы и совокупности объектов, их описывающие, называют *потенциальной бесконечностью*. Альтернативой является понятие *актуальной бесконечности*, которая означает рассмотрение конечно неизмеримых объектов как данность, как реально существующих, но при этом как единых и целостных, с которыми возможно оперировать.

Например, когда мы говорим, что, каковым бы ни было натуральное число, можно прибавить к нему 1 и получить большее число, мы обращаемся к понятию потенциальной бесконечности, доказывая тем самым, что множество натуральных чисел неограничено. Но если мы произносим «рассмотрим множество натуральных чисел», то обращаемся к понятию актуальной бесконечности, имея в виду бесконечное множество, как целое.

Существование потенциальной бесконечности Аристотель допускает. Но на актуальную бесконечность был наложен запрет, поскольку, следуя Зенону, введение ее в математику приводило, как тогда казалось, к неразрешимым противоречиям.

В III веке до н.э. Евклид формализовал и узаконил запрет Аристотеля в виде следующей аксиомы: «часть меньше целого».

Еще в 1638 году (почти две тысячи лет спустя!) Галилей в труде «Беседы и математические доказательства двух новых наук» вкладывает в уста Сальвиати такие рассуждения. Действительно, множество квадратов натуральных чисел является частью всего множества натуральных чисел. Однако, между ними можно установить взаимно-однозначное соответствие:

$$n \longleftrightarrow n^2,$$

т.е. часть может быть равна целому? Другой собеседник «Бесед» Сагредо, который отражает взгляды самого Галилея, возражает:«Свойства равенства, а также большой и меньшей величины не имеют места там, где дело идет о бесконечности».

По всей видимости, первым, кто стал вводить бесконечность в маетматику, был Ньютон. В 1665-1666 годах в Лондоне бушевала эпидемия чумы. Занятия в Тринити-колледже, где работал Ньютон, были прекращены и персонал распущен до окончания эпидемии. Ньютон уехал домой в Вулсторп и там посвятил все свое время исследованиям. Именно тогда он развил методы дифференциального и интегрального исчисления и открыл закон всемирного тяготения. В его работах по дифференциальному исчислению появились бесконечные ряды, которые впоследствии были названы *рядами Тейлора* (Тейлор был учеником Ньютона).

$$a_0 + a_1 x + a_2 x^2 + a_3 x^3 + \dots$$

С помощью этих рядов Ньютон научился решать дифференциальные уравнения. Он сам не публиковал этих работ, возможно по причине того, что не мог привести строгих обоснований методов, которые использовал. Такие обоснования появились спустя почти 200 лет.

Окончательно запрет на использование актруальной бесконечности был снят Георгом Кантором в 70-ых годах XIX века. Сразу после этого появилось несколько определений вещественных чисел, одно из которых мы и рассматриваем.

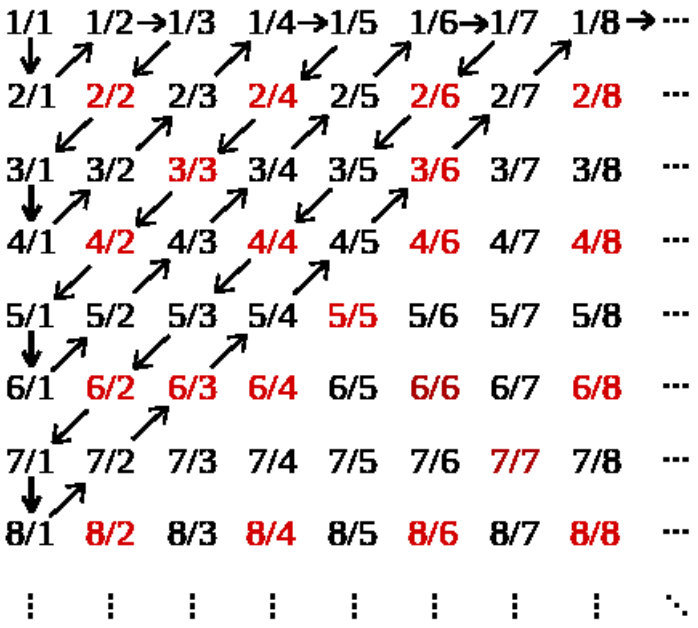
Кантор преодолел оковы нашей наглядной интуиции и ввел в математику многие парадоксальные конструкции (парадокс — *правда*, похожая на ложь!). Эти парадоксы есть *цена* снятия запрета Аристотеля. Часть действительно может быть равна целому: четных чисел столько же, сколько и всех натуральных, поскольку между ними можно установить взаимно однозначное соответствие

$$n \longleftrightarrow 2 \cdot n.$$

В этом случае принято говорить, что они имеют одинаковую мощность. Оказалось, что множества, на первый взгляд разные, имеют одинаковую мощность. Кажется совершенно ясным, что рациональных чисел «больше», чем натуральных. Однако, это не так.

Теорема. *Множества* $\mathbb{Q}$ *и* $\mathbb{N}$ *имеют одинаковую мощность.*

Доказательство. Чтобы это установить, достаточно пронумеровать рациональные числа, т. е. установить биекцию между множествами рациональных и натуральных чисел. Примером такого построения может служить следующий простой алгоритм. Составляется бесконечная таблица обыкновенных дробей, на каждой i-ой строке в каждом j-ом столбце которой располагается дробь $\frac{i}{j}$.



В процессе обхода, который изображен на рисунке, каждому новому рациональному числу ставится в соответствие очередное натуральное число. Т. е. дроби 1/1 ставится в соответствие число 1,

дроби $2/1$ — число 2, дроби $1/2$ — число 3 и т.д. Чтобы каждое рациональное число встречалось по одному разу, нумеруются только несократимые дроби. Совсем несложно сообразить, что, пронумеровав положительные рациональные числа, можно пронумеровать все элементы множества $\mathbb{Q}$. □

Множества, мощность которых равна мощности множества натуральных чисел, называются *счетными*. Иными словами, их элементы можно пересчитать и установить тем самым биекцию между исходным множеством и множеством $\mathbb{N}$. Бывают ли *несчетные* множества? Кантор дает ответ и на этот вопрос.

Теорема. *Множество вещественных чисел на отрезке* $[0,1]$ *не является счетным.*

Доказательство. Для удобства будем записывать вещественные числа в двоичной системе счисления. Предположим противное, а именно что нам удалось каким-то образом занумеровать все вещественные числа на отрезке $[0,1]$. Теперь мы предъявим число, которое не имеет номера и придем к противоречию.

$$
\begin{array}{ll}
s_1 & = 0\,0\,0\,0\,0\,0\,0\,0\,0\,0\,0\ldots \\
s_2 & = 1\,1\,1\,1\,1\,1\,1\,1\,1\,1\,1\ldots \\
s_3 & = 0\,1\,0\,1\,0\,1\,0\,1\,0\,1\,0\ldots \\
s_4 & = 1\,0\,1\,0\,1\,0\,1\,0\,1\,0\,1\ldots \\
s_5 & = 1\,1\,0\,1\,0\,1\,1\,0\,1\,0\,1\ldots \\
s_6 & = 0\,0\,1\,1\,0\,1\,1\,0\,1\,1\,0\ldots \\
s_7 & = 1\,0\,0\,0\,1\,0\,0\,0\,1\,0\,0\ldots \\
s_8 & = 0\,0\,1\,1\,0\,0\,1\,1\,0\,0\,1\ldots \\
s_9 & = 1\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\ldots \\
s_{10} & = 1\,1\,0\,1\,1\,1\,0\,0\,1\,0\,1\ldots \\
s_{11} & = 1\,1\,0\,1\,0\,1\,0\,0\,1\,0\,0\ldots \\
\vdots & \quad \vdots\;\vdots\;\vdots\;\vdots\;\vdots\;\vdots\;\vdots\;\vdots\;\vdots\;\vdots\;\ddots
\end{array}
$$

$$
s = 1\,0\,1\,1\,1\,0\,1\,0\,0\,1\,1\ldots
$$

Рассмотрим *первый* знак после запятой у *первого* числа. Если это ноль, то первым знаком искомого числа будет единица (иначе наоборот). Рассмотрим *второй* знак после запятой у *второго* числа и проделаем аналогичную операцию. Продолжая эту операцию до бесконечности, мы получим вещественное число, которое не имеет своего номера, ведь оно отличается от каждого занумерованного числа хотя бы в одном разряде! □

Замечание. Какое простое в техническом смысле, но какое сложное в смысле идейном доказательство!

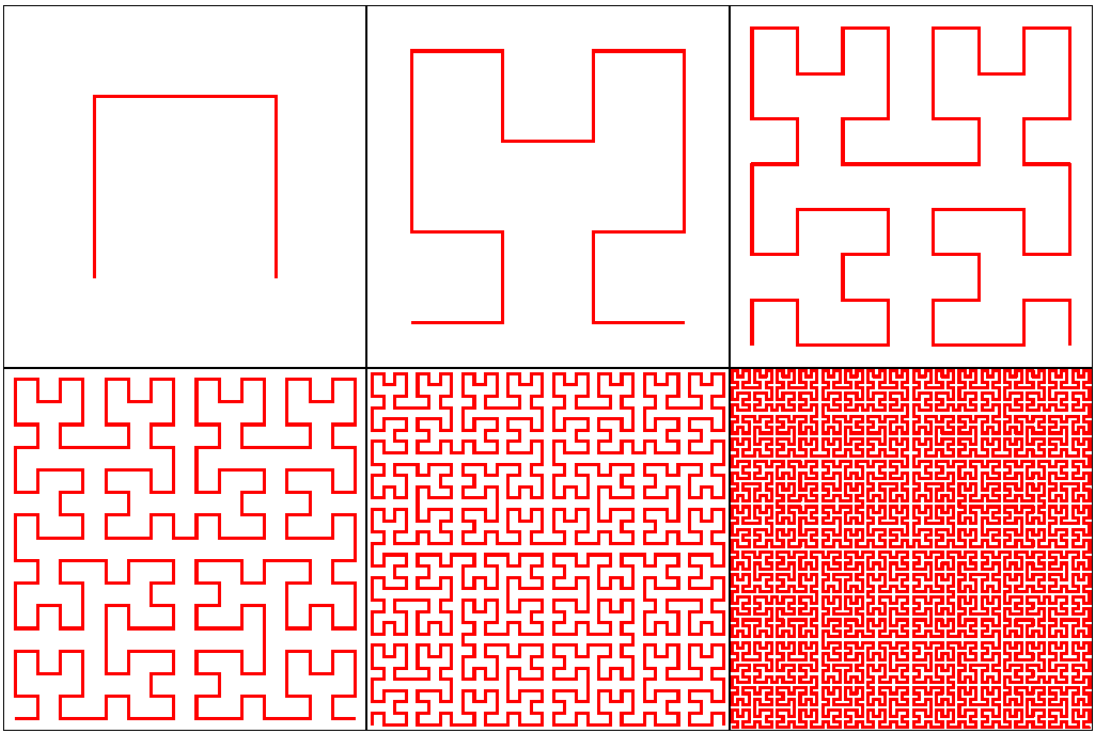
Таким образом, вещественных чисел оказалось больше, чем натуральных. Одна бесконечность оказалась больше другой! А найдется ли бесконечность, которая больше, чем бесконечность множества вещественных чисел? Логично было бы предполагать, что на плоскости точек больше, чем на прямой... Оказывается, это не так! На плоскости точек столько же, сколько и на прямой! Сам Кантор писал об этом открытии в письме Дедекинду так: «Я вижу это, но не верю этому».

Замечание. Мощность множества всех вещественных чисел называется континуумом. В списке уже упоминавшихся нами проблем Гильберта первое место занимает континуум-гипотеза — выдвинутая в 1877 году Георгом Кантором гипотеза о том, что не существует «промежуточных» мощностей между счетным множеством и континуумом.

Решение обозначенной проблемы совершенно поразительно. Невозможно доказать или опровергнуть континуум-гипотезу в рамках принятой аксиоматики теории множеств (аксиоматике Цермело-Френкеля). Таким образом, гипотеза Кантора является независимой от системы аксиом теории множеств.

Замечание. Теория множеств доставляет много парадоксальных примеров. Например, существует непрерывное отображение отрезка на квадрат!

Рассмотрим единичный отрезок и единичный квадрат. На 1-м шаге построения разделим квадрат средними линиями на 4 равных квадрата, а отрезок — на 4 равные части. Получим квадраты и отрезки 1-го уровня. На каждом последующем шаге делим квадраты и отрезки предыдущего уровня на 4 части — получаем квадратики и отрезочки следующего уровня. Зададим порядок обхода квадратиков каждого уровня. Для 1-го, 2-го, ..., 6-го уровня порядок обхода показан на рисунке. Порядок обхода определяет взаимно однозначное соответствие между множеством квадратиков n-го уровня и множеством отрезочков n-го уровня.



Впоследствии Кантору удалось доказать, что какую бы мощность ни имело бесконечное множество, всегда найдется множество, имеющее большую мощность. Таким образом, он открыл существование бесконечного количества бесконечностей. Воистину потрясающим образом устроен мир.

Судьба Кантора была очень тяжелой. Его работы подвергались жесточайшей критике со стороны коллег. Даже самые сильные и яркие умы не всегда способны преодолеть инерцию мышления.

Возвращаясь к произвольным вещественным числам, возникает естественный вопрос, как складывать и умножать эти числа? Для этого нам потребуется вспомогательное понятие.

Определение. Число $\alpha_k = n_0, n_1 n_2 \dots n_k$ называется *приближением числа α по недостатку.*

Число $\alpha'_k = n_0, n_1 n_2 \dots n_k + \frac{1}{10^k}$ называется *приближением числа α по избытку.*

Чтобы сложить два вещественных числа

$$\alpha = n_0, n_1 n_2 \dots n_k \dots \quad \text{и} \quad \beta = m_0, m_1 m_2 \dots m_k \dots,$$

образуем последовательность их десятичных приближений по избытку: $\alpha'_0, \alpha'_1, \dots, \alpha'_k, \dots$ и $\beta'_0, \beta'_1, \dots, \beta'_k, \dots$. Сложим их покомпонентно и получим такую последовательность:

$$\alpha'_0 + \beta'_0,\ \alpha'_1 + \beta'_1,\ \dots,\ \alpha'_k + \beta'_k,\ \dots.$$

Заметим, что $\alpha'_k \geqslant \alpha'_{k+1}$ и $\beta'_k \geqslant \beta'_{k+1}$. Рассмотривая целые части сумм, получаем

$$[\alpha'_0 + \beta'_0] \geqslant [\alpha'_1 + \beta'_1] \geqslant \dots \geqslant [\alpha'_k + \beta'_k] \dots.$$

Эти целые части являются натуральными числами, и поэтому среди них есть наименьшее число. Тогда все числа, идущие за ним одинаковы, т.е. последовательность стабилизируется. Полученное число и есть $[\alpha + \beta]$. Далее следим за цифрой десятых в последовательности сумм. По тем же соображениям она начинает повторяться, и мы переходим к цифре сотых и т.д. так одна за одной будут определяться цифры числа $\alpha + \beta$.

Аналогичным образом будет определяться умножение вещественных чисел. Нужно только рассмотреть последовательность произведений приближений по избытку.

Теперь мы готовы дать ответ на вопрос, который поставил в тупик древнегреческих математиков, и определить, что же такое $\sqrt{2}$. Пусть

$$\sqrt{2} = x_0, x_1 x_2 x_3 \dots.$$

Имеем

$$\begin{aligned}
&x_0^2 < 2, \;\; (x_0+1)^2 > 2 \;\; \Rightarrow \;\; x_0 = 1\\
&(1,x_1)^2 < 2, \;\; (1,x_1+0,1)^2 > 2 \;\; \Rightarrow \;\; x_1 = 4\\
&(1,4x_2)^2 < 2, \;\; (1,4x_2+0,01)^2 > 2 \;\; \Rightarrow \;\; x_2 = 1\\
&(1,41x_3)^2 < 2, \;\; (1,41x_3+0,001)^2 > 2 \;\; \Rightarrow \;\; x_3 = 4\\
&\ldots\ldots\ldots\ldots
\end{aligned}$$

Иными словами, $\sqrt{2}$ — это *алгоритм*, который позволяет отыскать любой знак после запятой у решения уравнения $x^2 = 2$.

Теперь у нас появился корень уравнения $x^2 = 2$. Аналогичным образом можно построить решение уравнения $x^2 = 3$ и так далее. Предположим, нам удалось рассмотреть все корни всевозможных уравнений вида

$$x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0 = 0,$$

где a_i — целые числа. Такие числа (корни многочленов) называются *алгебраическими*. Исчерпывается ли множество вещественных чисел алгебраическими? Иными словами, существуют ли числа, которые не являются корнем никакого многочлена с целыми коэффициентами? Мы готовы дать ответ и на этот вопрос.

Теорема. *Множество алгебраических чисел* $\mathbb{A}$ *является счетным.*

Доказательство. Рассмотрим многочлен

$$p = x^n + a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$$

и его *высоту*

$$h = n + |a_{n-1}| + |a_{n-2}| + \ldots + |a_0|.$$

Высота многочлена с целыми коэффициентами является натуральным числом. Очевидно, что существует лишь конечное число многочленов с фиксированной высотой. Следовательно, мы легко сможем их занумеровать. Как нам известно, количество корней любого многочлена ограничено его степенью. Таким образом, имея нумерацию

многочленов, можно пронумеровать их корни, то есть все алгебраические числа. □

Следствие. *Существуют трансцендентные, то есть не алгебраические числа.*

Доказательство. Множество $\mathbb{A}$ счетно, а множество всех вещественных чисел $\mathbb{R}$ имеет мощность континуума. □

6. Факторкольца $\Bbbk[x]/(f)$

Ранее в конце главы IV мы, проводя аналогию между кольцом целых чисел $\mathbb{Z}$ и кольцом многочленов $\mathbb{R}[x]$, задали вопрос: что получится, если вместо кольца остатков $\mathbb{Z}_m$ мы рассмотрим кольцо остатков $\mathbb{R}[x]/(f)$, где $f \in \mathbb{R}[x]$? Работа, проделанная нами в разделе XI.1, позволяет ответить на этот вопрос и развить соответствующую теорию, что мы сейчас и сделаем.

Ранее в главе VII мы рассматривали многочлены с коэффициентами не только в поле вещественных чисел $\mathbb{R}$, но и в конечных полях $\mathbb{Z}_p$. Можно рассматривать многочлены с коэффициентами и в других полях (например, в $\mathbb{Q}$). Чтобы построить общую теорию для всех таких полей, договоримся через $\Bbbk$ обозначать произвольное поле (конечное или бесконечное). В дальнейшем полезно проследить, как будут выглядеть получаемые нами результаты в двух наиболее важных случаях: $\Bbbk = \mathbb{R}$ и $\Bbbk = \mathbb{Z}_p$.

По аналогии с рассмотрением кольца целых чисел договоримся считать два многочлена эквивалентными, если они дают одинаковый остаток при делении на фиксированный многочлен f (о делении многочленов с остатком см. главу II). Обозначается это так же, как и в случае колец вычетов: $p \equiv q \pmod f$.

Определение. Кольцо многочленов $\Bbbk[x]$ разбивается на классы, состоящие из эквивалентных многочленов. Множество этих классов называется *факторкольцом кольца многочленов* $\Bbbk[x]$ и обозначается через $\Bbbk[x]/(f)$.

Замечание. Отметим, что если поле $\Bbbk$ конечно (например, $\Bbbk = \mathbb{Z}_p$), то и факторкольцо $\Bbbk[x]/(f)$ будет конечным. Напротив, если $\Bbbk$ бесконечно (например, $\Bbbk = \mathbb{R}$ или $\mathbb{Q}$), то и факторкольцо будет бесконечным, поскольку оно содержит в себе основное поле $\Bbbk$ (многочлены нулевой степени попадают в различные классы, потому что, очевидно, имеют различные остатки при делении на f).

Операции на множестве $\Bbbk[x]/(f)$ определяются стандартным образом. Класс многочленов, который представлен многочленом g, будем обозначать $[g]_f$.

Определение. *Суммой* классов $[p]_f$ и $[q]_f$ называется класс $[p+q]_f$.
Произведением классов $[p]_f$ и $[q]_f$ называется класс $[p \cdot q]_f$.

Таким образом, множество $\Bbbk[x]/(f)$ является кольцом.

Прежде чем двигать дальше, рассмотрим несколько примеров. В дальнейшем для удобства будем работать не с классами, а с их представителями.

Пример. 1. Пусть $\Bbbk$ — произвольное поле $f(x) = x - a$, где $a \in \Bbbk$. В этом случае $\Bbbk[x]/(f) = \Bbbk$. Можно показать, что многочлены первой степени — это единственные многочлены, факторкольца которых совпадают с основным полем $\Bbbk$. Этот важнейший факт имеет большое значение в алгебраической геометрии и называется теоремой Гильберта о нулях.

2. Пусть $\Bbbk = \mathbb{R}$ и $f(x) = x^2 + 1$. Тогда остатки многочленов от деления на f имеют степень не выше 1, т.е. равны $a + bx$, где a, $b \in \mathbb{R}$.

3. Пусть $\Bbbk$ — произвольное поле и f — произвольный многочлен степени n. Тогда остатки многочленов от деления на f имеют степень не выше $n-1$, т.е. равны $a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}$, где a_0, $\ldots$, $a_{n-1} \in \Bbbk$.

Посмотрим, как выглядит операция умножения в наших факторкольцах (ведь именно с операцией умножения были связаны основные наши результаты в случае $\mathbb{Z}_m$).

Рассмотрим факторкольцо $\mathbb{R}[x]/(x^2+1)$. Попробуем умножить в нем два элемента $a+bx$ и $c+dx$. Для этого нам нужно перемножить их как обычные многочлены (в результате получится многочлен степени 2), а затем взять его остаток от деления на x^2+1. Используя язык сравнений, получаем:

$$(a+bx)\cdot(c+dx) = (ac)+(ad+bc)x+(bd)x^2 \equiv$$
$$\equiv (ac)+(ad+bc)x+(bd)(-1) = (ac-bd)+(ad+bc)x.$$

В частности, $x^2 = -1$! И вновь мы получаем это удивительное равенство!

Замечание. Данный пример является очень важным в математике. Отметим, что факторкольцо $\mathbb{R}[x]/(x^2+1)$ обозначается через $\mathbb{C}$ и называется *полем комплексных чисел*.

Комплексные числа вошли в математику очень интересным образом. В 1545 году была напечатана книга итальянского математика Джероламо Кардано «Великое искусство». Одним из главных ее результатов была формула для решения уравнений третьей степени. Еще в Древнем мире умели решать квадратные уравнения, а уравнения третьей степени научились решать только в XVI веке.

Произвольное уравнение третьей степени с помощью замены переменной можно привести к виду

$$x^3+px+q=0.$$

Формула Кардано дает решение этого уравнения

$$x = \sqrt[3]{-\frac{q}{2}+\sqrt{\frac{p^3}{27}+\frac{q^2}{4}}}+\sqrt[3]{-\frac{q}{2}-\sqrt{\frac{p^3}{27}+\frac{q^2}{4}}}.$$

Давайте ее проверим. Рассмотрим, например, уравнение

$$(x-1)(x-2)(x+3) = x^3-7x+6=0.$$

Следуя Кардано, получаем

$$x = \sqrt[3]{-3+\sqrt{-\frac{343}{27}+9}} + \sqrt[3]{-3-\sqrt{-\frac{343}{27}+9}} =$$
$$= \sqrt[3]{-3+\sqrt{-\frac{100}{27}}} + \sqrt[3]{-3-\sqrt{-\frac{100}{27}}}.$$

Что это за число?! Это 1, 2 или -3? И что такое $\sqrt{-\frac{100}{27}}$?

Так в математику вошли комплексные числа (еще раньше, чем отрицательные!), среди которых есть решение уравнения $x^2+1=0$. Как мы выяснили выше, решением этого уравнения является не какой-то непонятный символ, квадрат которого равен -1, а класс $[x]_{x^2+1}$ многочлена x по модулю x^2+1.

Отметим также, что зачастую удобно использовать другие (эквивалентные нашему) формы записи комплексных чисел. Например, можно мыслить комплексное число z как пару вещественных чисел (a,b), или же как точку на координатной плоскости. Отмечая вектор с началом в начале координат и концом в данной точке (a,b) и раскладывая его по ортонормированному базису, получаем стандартную форму записи комплексного числа $z=a+bi$, где i — так называемая мнимая единица (единичный вектор на оси ординат).

Теперь попробуем сформулировать и доказать аналоги теорем из главы IV. Начнем с теоремы о делителях нуля.

Теорема. *Ненулевой элемент $g \in \Bbbk[x]/(f)$ является делителем нуля тогда и только тогда, когда* $\deg$ *НОД*$(g,f) > 0$.

Ее доказательство абсолютно аналогично доказательству теоремы о делителях единицы из раздела IV.2.

Также имеет место и теорема о делителях единицы.

Теорема. *Элемент $g \in \Bbbk[x]/(f)$ является делителем единицы тогда и только тогда, когда НОД*$(g,f)=1$.

Доказательство. Заметим, что перенести доказательство этой теоремы из раздела IV.3 напрямую не удастся, поскольку в случае многочленов мы не имеем аналога леммы о колоде карт (точнее говоря, лемма о колоде карт отсутствует в случае бесконечного поля $\Bbbk$). Поэтому мы воспользуемся другим соображением (кстати говоря, это же соображение работает и в случае колец остатков $\mathbb{Z}_m$).

Рассмотрим элемент $g \in \Bbbk[x]/(f)$, взаимно простый с f. По лемме о представимости НОДа существуют такие многочлены $u(x)$ и $v(x) \in \Bbbk[x]$, что $fu + gv = 1$. Переходя к остаткам по модулю f, получаем $g \cdot v \equiv 1 \pmod f$, откуда $g^{-1} = v$. Тем самым мы явно предъявили обратный эелмент g^{-1}, что и требовалось. □

Из приведенных теорем следует важное утверждение.

Следствие. *Факторкольцо $\Bbbk[x]/(f)$ является полем тогда и только тогда, когда многочлен f неприводим над $\Bbbk$.*

Замечание. Факторкольцо $\mathbb{Z}_m$ является полем тогда и только тогда, когда число m является простым.

Пример. 1. В факторкольце $\mathbb{R}[x]/(x^2-1)$ делителями нуля являются многочлены вида $a(x\pm1)$, где $a \neq 0$, а все остальные многочлены являются делителями единицы.

2. Факторкольцо $\mathbb{C} = \mathbb{R}[x]/(x^2+1)$ действительно является полем, поскольку многочлен x^2+1 неприводим над $\mathbb{R}$.

3. Факторкольцо $\mathbb{R}[x]/(f)$, где $\deg f > 1$ и нечетна, никогда не является полем, поскольку любой многочлен нечетной степени приводим.

4. Факторкольцо $\mathbb{R}[x]/(x^4+4)$ не является полем, поскольку $x^4+4 = (x^2-2x+2)(x^2+2x+2)$.

Следующий вопрос, который мы рассмотрим, связан с вычислением обратных элементов в факторкольце $\mathbb{R}[x]/(f)$ (напомним, что в случае кольца вычетов $\mathbb{Z}_m$ этот вопрос тесно связан с решением линейных диофантовых уравнений). Доказательство теоремы о

делителях единицы доставляет нам общий алгоритм вычисления обратных элементов, однако в данном случае можно поступить проще.

Рассмотрим наш пример факторкольца $\mathbb{C} = \mathbb{R}[x]/(x^2+1)$, выберем в нем произвольный ненулевой элемент $a+bx$ и попробуем вычислить для него обратный элемент $(a+bx)^{-1}$. Согласно следствию, такой элемент всегда существует, поскольку наше факторкольцо является полем.

По определению обратный элемент $(a+bx)^{-1} = c+dx$ таков, что $(a+bx)\cdot(c+dx) \equiv 1 \pmod{x^2+1}$. Раскрывая скобки и приравнивая коэффициенты, получаем систему уравнений

$$\begin{cases} ac - bd &= 1 \\ ad + bc &= 0. \end{cases}$$

Решая эту систему относительно неизвестных c и d, получаем ответ:

$$(a+bx)^{-1} = \frac{1}{a^2+b^2}(a-bx).$$

Замечание. Напомним, что для элемента $z = a+bx$ элемент $a-bx$ называется сопряженным и обозначается через $\bar{z}$, а произведение $N(z) = z\cdot\bar{z}$ называется нормой элемента z и равно a^2+b^2. Поэтому последнюю формулу можно записать так: $z^{-1} = \bar{z}/N(z)$.

Теперь посмотрим на еще одну интересную конструкцию. До сих пор мы с вами рассматривали факторкольца вида $\Bbbk[x]/(f)$, где $\Bbbk$ является полем. Однако все наши рассуждения можно провести для случая, когда $\Bbbk$ является кольцом (обычно произвольное кольцо обозначается через R от английского слова «ring»), а многочлен f является приведенным (т.е. его старший коэффициент обратим). В этом случае все предыдущие наши результаты без изменений переносятся на этот случай (естественно, слово «поле» в них необходимо заменить на слово «кольцо»).

Рассмотрим факторкольцо $\mathbb{Z}[x]/(x^2+k)$. Его элементами являются выражения вида $a+bx$, где $a, b \in \mathbb{Z}$ и $x^2 = -k$. Если написать формально $x = \sqrt{-k}$, то мы получаем ни что иное как кольцо $\mathbb{Z}[\sqrt{-k}]$,

которое мы уже рассматривали ранее! Но теперь мы понимаем «символ» $\sqrt{-k}$: это класс многочлена x по модулю x^2+k! А что если в качестве исходного кольца взять $\mathbb{Z}_m$? С этими конструкциями и их интересными свойствами вы можете ознакомиться в Добавлении «О функции Эйлера алгебраических расширений колец вычетов».

Далее, поясним, какую роль могут играть факторкольца $\mathbb{Z}_p[x]/(f)$ над конечными полями $\mathbb{Z}_p$. Ранее в главе VII мы обсуждали вопрос о том, какие бывают конечные поля. Можно доказать, что количество элементов в конечном поле всегда является степенью простого числа. Оказывается, что верно и обратное: для любой степени простого p^n существует единственное поле с таким количеством элементов. Возникает вопрос, как оно устроено?

В случае $n=1$ все просто — это наши поля вычетов $\mathbb{Z}_p$. Однако уже для $n=2$ ситуация становится совсем не простой. Например, поле из четырех элементов — это вовсе не $\mathbb{Z}_4$ (так как 4 — не простое число, это множество вообще не является полем).

Оказывается, построить поля из p^n элементов помогает наша конструкция факторколец. А именно, рассмотрим кольцо многочленов $\mathbb{Z}_p[x]$ и неприводимый мнонгочлен f степени n. Можно показать (и этот момент является самым трудным!), что такой многочлен обязательно существует для всех p и n. Тогда факторкольцо

$$\mathbb{F}_{p^n} = \mathbb{Z}_p[x]/(f)$$

является полем согласно нашему следствию, а его элементы имеют вид $a_0 + a_1 x + \ldots + a_{n-1}x^{n-1}$, где a_0, a_1, $\ldots$, $a_{n-1} \in \mathbb{Z}_p$. Несложно видеть, что общее количество таких элементов в точности равно p^n.

В частности, для $p=n=2$ можно взять $f(x)=x^2+x+1$, а поле $\mathbb{F}_4$ из четырех элементов 0, 1, x, $1+x$ со следующей таблицей умножения:

$\times$	1	x	$1+x$
1	1	x	$1+x$
x	x	$1+x$	1
$1+x$	$1+x$	1	x

В заключение этого раздела мы дадим ответ на вопрос, который наверняка задают себе все восьмиклассники, когда проходят квадратные корни. Во всех учебниках алгебры можно в изобилии встретить задачи вида *избавьтесь от иррациональности в знаменателе дроби.*

Зачем убирать иррациональность в знаменателе?! Ведь иногда окончательный ответ получается гораздо более громоздким, нежели исходное выражение! Например,

$$\frac{1}{\sqrt{5}+\sqrt{2}+1} = \frac{1}{2}(3+2\sqrt{2}-\sqrt{5}-\sqrt{10}).$$

Дело вот в чем. Рассмотрим теперь в качестве основного поля $\Bbbk$ поле $\mathbb{Q}$ рациональных чисел. Давайте попробуем посмотреть, что представляет собой факторкольцо $\mathbb{Q}[\sqrt{2}] = \mathbb{Q}[x]/(x^2-2)$. Согласно нашим результатам, это факторкольцо состоит из элементов вида $a+bx$, где $a, b \in \mathbb{Q}$ и $x^2-2=0$. Иначе говоря,

$$\mathbb{Q}[x]/(x^2-2) = \{a+b\sqrt{2} : a, b \in \mathbb{Q}\}.$$

Тем самым наша конструкция факторкольца дает нам возможность работать с иррациональными числами, не используя само определение иррациональных чисел как бесконечных десятичных дробей (см. XI.4)! Отметим также, что поле $\mathbb{Q}[\sqrt{2}]$ называется *расширением поля* $\mathbb{Q}$.

Заметим, что многочлен x^2-2 *неприводим над полем* $\mathbb{Q}$, поэтому факторкольцо $\mathbb{Q}[x]/(x^2-2)$ является полем. Значит, для каждого ненулевого элемента существует обратный. А теперь давайте вычислим обратный элемент $(\sqrt{2}+1)^{-1}$. Если рассматривать этот элемент как вещественное число, то это просто обычная дробь $\frac{1}{\sqrt{2}+1}$ с иррациональностью в знаменателе. Однако если мы хотим рассмотреть этот

элемент как элемент факторкольца $\mathbb{Q}[x]/(x^2-2)$, иррациональность в знаменателе содержаться не должна!

Избавление от иррациональности в знаменателе есть ни что иное как нахождение обратных элементов в факторкольцах вида $\mathbb{Q}[x]/(f)$!

Доказательство теоремы о делителях единицы дает нам алгоритм избавления от иррациональности в знаменателе. Напомним его.

Рассмотрим произвольный неприводимый над $\mathbb{Q}$ многочлен f степени n и какой-то его корень α (отметим еще раз, что $\alpha \notin \mathbb{Q}$, иначе $f(x) \vdots (x-\alpha)$). Пусть

$$g = a_0 + a_1\alpha + a_2\alpha^2 + \ldots + a_{n-1}\alpha^{n-1}$$

— произвольный ненулевой элемент факторкольца $\mathbb{Q}[x]/(f)$. Возьмем многочлен $\widetilde{g}(x) = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}$ (мы просто заменили в элементе g элемент α на переменную x). Поскольку f неприводим, $\text{НОД}(f, \widetilde{g}) = 1$. По лемме о представимости НОДа существуют такие многочлены $u(x)$ и $v(x) \in \mathbb{Q}[x]$, что $fu + \widetilde{g}v = 1$. Подставляя теперь в это равенство $x = \alpha$, получаем $g \cdot v(\alpha) = 1$, откуда $g^{-1} = v(\alpha)$.

Тем самым мы имеем *каноническую процедуру* избавления от иррациональностей в знаменателе! Причем не только от квадратичных, но и даже от таких, которые и записать-то не получится. Например, можно вычислить α^{-1}, где α — корень многочлена $x^5 - x + 1 = 0$ (согласно теореме Абеля, корень такого многочлена невозможно записать, используя четыре арифметические операции и операции извлечения корней). Для этого нужно вычислить представление НОДа многочленов x и x^5-x+1 с помощью алгоритма Евклида, взять коэффициент, стоящий перед x и подставить в него $x = \alpha$. Получившийся элемент и будет обратным к α. Имеем:

$$(x^5 - x + 1) \cdot (1) + x \cdot (-x^4 + 1) = 1,$$

откуда $\alpha^{-1} = 1 - \alpha^4$.

Замечание. Обратите внимание, что у многочлена может быть несколько корней, однако формула для нахождения обратных эле-

ментов не зависит от выбора корня! Она определяется только самим многочленом.

Замечание. Выше мы рассмотрели расширения поля $\mathbb{Q}$ с помощью только одного иррационального корня α. Однако подобным образом можно рассматривать и поля вида $\mathbb{Q}[\alpha_1, \ldots, \alpha_k]$, где α_i — корни неприводимых над $\mathbb{Q}$ многочленов f_i.

7. Что дальше?

Итак, мы продолжили цепочку вложенных числовых множеств

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C}.$$

Поле комплексных чисел, которое мы определили в предыдущем разделе, обладает одним ключевым свойством — оно *алгебраически замкнуто*, т.е. *любой* многочлен в $\mathbb{C}[x]$, отличный от константы, имеет корень. А это означает, что любой многочлен над полем комплексных чисел может быть разложен на *линейные* множители. Это свойство оказывается чрезвычайно важным как в алгебре (теорема об алгебраической замкнутости поля $\mathbb{C}$ называется «основной теоремой алгебры»), так и в геометрии.

Оказывается, что и это не конец нашей цепочки! Можно продолжать ее построение

$$\mathbb{N} \subset \mathbb{Z} \subset \mathbb{Q} \subset \mathbb{R} \subset \mathbb{C} \subset ???$$

Но это уже совсем другая история...

Литература

[1] Арнольд В.И. *Группы Эйлера и арифметика геометрических прогрессий*. М.: МЦНМО, 2003.

[2] Арнольд В.И. *Экспериментальная математика*. М.: Фазис, 2005.

[3] Аршинов М.Н., Садовский Л.Е. *Грани алгебры*. М.: Факториал, 2008.

[4] Винберг Э.Б. *Курс алгебры*. М.: МЦНМО, 2011.

[5] Виленкин Н.Я. *Рассказы о множествах*. М.: МЦНМО, 2013.

[6] Виленкин Н.Я. и др *Алгебра 8*. М.: Просвещение, 2006.

[7] Дэвенпорт Г. *Высшая арифметика. Введение в теорию чисел*. М.: Наука, 1965 г.

[8] Прасолов В.В. *Многочлены*. М.: МЦНМО, 2014.

[9] Радемахер Г., Теплиц О. *Числа и фигуры*. М.: Физматгиз, 1962.

[10] Шафаревич И.Р. *Основные понятия алгебры*. Ижевск: РХД, 2001.

Добавления

О геометрии диаграмм Юнга перестановок Арнольда

Д. А. Байгушев[(1)]

1. Введение

В работе [3] В. И. Арнольдом был предложен новый способ исследования перестановок, основанный на рассмотрении диаграмм Юнга.

Разобьем перестановку на циклы. Упорядочив длины циклов по убыванию, получим множество $\{a_1, a_2, ..., a_h\}$, где h — количество циклов перестановки. Построим фигуру, у которой в первом ряду будет a_1 единичных квадратиков, во втором — a_2 единичных квадратиков, и т. д. Данную фигуру мы будем называть *диаграммой Юнга перестановки.*

С диаграммой Юнга естественно связать некоторые геометрические характеристики:

- *высота* h диаграммы Юнга,
- *длина* $l := a_1$ диаграммы Юнга,
- *площадь* $n := a_1 + \ldots + a_h$ диаграммы Юнга,
- *вертикальная и горизонтальная асимметрии* $\mu := h/l$ *и* $\eta := l/h$ диаграммы Юнга,
- *плотность* $\lambda := n/(h \cdot l)$ диаграммы Юнга.

В [3] Арнольдом были приведены вычисления *средних параметров* $\widehat{h}$, $\widehat{l}$, $\widehat{\mu}$ и $\widehat{\lambda}$ диаграмм Юнга перестановок длины не больше 7 и

[(1)] *Лицей «Вторая школа», Москва, Россия*
e-mail: IDanila24@gmail.com

высказаны гипотезы об асимптотиках этих средних (при $n \to \infty$):

$$\widehat{h}(n) \sim c_1 \ln n, \quad \widehat{l}(n) \sim c_2 n, \quad \widehat{\lambda}(n) \sim c_3 / \ln n, \quad \widehat{\mu}(n) \sim c_4 \ln n / n.$$

Также В.И. Арнольд предложил исследовать специальный класс перестановок, которые строятся следующим образом. Рассмотрим множество $\{1, 2, \ldots, n\}$. Разобьем его на три непустых блока $\{A, B, C\}$ размеров a, b и c соответственно и переставим их в порядке $\{C, B, A\}$. Получившуюся перестановку мы будем называть (C, B, A)-*перестановкой* (или *перестановкой Арнольда*) и будем обозначать ее через $\sigma(a, b, c)$.

Задача об исследовании (C, B, A)-перестановок является дискретным аналогом известной *задачи о перекладывании отрезков*, поставленной Арнольдом в 1958 г. (см. [1]).

Целью данной работы является исследование геометрии диаграмм Юнга (C, B, A)-перестановок Арнольда. В частности, мы явно вычислим средние параметры диаграмм Юнга (C, B, A)-перестановок. Оказывается, что некоторые из них совпадают с асимптотиками Арнольда, а некоторые нет.

2. Об асимптотике эргодических перестановок Арнольда

Перестановки можно рассматривать как динамические системы на конечном пространстве. Одним из важных свойств динамической системы является всюду плотность ее траекторий. В случае перестановок это условие означает, что перестановка состоит из одного цикла. Такие перестановки мы будем называть *эргодическими*.

Основной целью данного пункта является решение следующей задачи, сформулированной В.И. Арнольдом в 1958 г.: *какова (асимптотически) доля эргодических* (C, B, A)-*перестановок*?

Замечание. Отметим, что доля обычных эргодических перестановок длины n равна $1/n$ и стремится к 0 при $n \to \infty$.

Для решения этой задачи нам понадобится следующее определение.

Определение. Назовем *шагами перестановки* σ величины $\sigma(i) - i$, где $i = 1, \ldots, n$.

Отметим, что в (C, B, A)-перестановках возможны всего три шага, которые мы обозначим через S_C, S_B и S_A соответственно. А именно, S_C — шаг, на который увеличивается число, переходящее в блок C, S_B — в блок B и S_A — в блок A.

Легко видеть, что

$$S_C = a + b, \quad S_B = a - c, \quad S_A = -b - c.$$

Теорема 1 (Критерий эргодичности)**.** *Перестановка Арнольда* $\sigma(a, b, c)$ *эргодична тогда и только тогда, когда НОД*$(a+b, b+c) = 1$.

Замечание. Второе условие теоремы в свою очередь равносильно условию НОД$(S_A, S_B, S_C) = 1$.

Доказательство. «$\Rightarrow$» Если НОД$(S_C, S_B, S_A) = d \neq 1$, то, передвигаясь по циклу с шагами S_C, S_B и S_A, мы не сможем попасть из 1 в 2, т. к. мы будем попадать только в числа, сравнимые с 1 по модулю d.

«$\Leftarrow$» Рассмотрим первый цикл перестановки. Пройдя по нему один раз, мы получим: $xS_A + yS_B + zS_C = 0$ (здесь x — количество шагов S_A в цикле, y — количество шагов S_B и z — количество шагов S_C).

Подставив в это равенство значения шагов, получаем:

$$(x + y)(b + c) = (y + z)(a + b).$$

Т. к. НОД $(a + b, b + c) = 1$, то $\begin{cases} x + y \geqslant a + b \\ y + z \geqslant b + c \end{cases}$.

Сложим два получившихся неравенства: $x + 2y + z \geqslant a + 2b + c$.

Т. к. $y \leqslant b$, то $x + y + z \geqslant a + b + c = n$, т.е. длина первого цикла не меньше n. Но это означает, что она в точности равна n, и перестановка $\sigma(a, b, c)$ эргодична. $\square$

С помощью критерия эргодичности мы решим задачу Арнольда.

Теорема 2. *Доля эргодических перестановок Арнольда асимптотически равна* $6/\pi^2$.

Доказательство. Рассмотрим перестановку Арнольда $\sigma(a,b,c)$. Положим $x := b+c = n-a$ и $y := a+b = n-c$. Тогда множество перестановок Арнольда соответствует множеству точек

$$\triangle = \{(x,y) \in \mathbb{Z}^2 : x < n,\ y < n,\ x+y > n\}.$$

В то же время согласно критерию эргодичности множество эргодических перестановок Арнольда соответствует множеству точек в $\triangle$ со взаимно простыми координатами.

Согласно теореме Арнольда [2] о равномерной распределенности, доля точек в $\triangle$ со взаимно простыми координатами асимптотически (при $n \to \infty$) равна $1/\zeta(2) = 6/\pi^2$, где ζ — дзета-функция римана, что и требовалось доказать (смотри III.6). □

В заключение отметим, что эта теорема хорошо подтверждается численными экспериментами (см. таблицу).

размер пер-вки	10	10^2	10^3	10^4
всего пер-вок	36	4851	498501	49985001
эргодических пер-вок	24	2964	303392	30389486
доля эргодических пер-вок	**0,6**6667	**0,61**1008	**0,608**609	**0,6079**72
константа $6/\pi^2$	**0,6**07927	**0,6**07927	**0,607**927	**0,6079**27

3. Высота диаграмм Юнга перестановок Арнольда

С точки зрения геометрии диаграмм Юнга эргодичность перестановки означает, что ее диаграмма Юнга имеет высоту 1.

Следующая теорема обобщает этот результат.

Теорема 3 (Обобщение критерия эргодичности)**.** *Высота диаграммы Юнга перестановки Арнольда* $\sigma(a,b,c)$ *равна*

$$h = НОД(a+b, b+c).$$

Доказательство. 1. Если НОД$(S_C, S_B, S_A) = h$, то количество циклов в перестановке $\sigma(a, b, c)$ не меньше h, т.к. передвигаясь по перестановке с шагами S_C, S_B и S_A, мы будем попадать только в числа, сравнимые друг с другом по модулю h. Поэтому числа 1, 2, ..., h заведомо лежат в разных циклах.

2. Теперь докажем, что количество циклов не больше h. Для этого достаточно доказать, что каждый цикл нашей перестановки содержит *все* числа, сравнимые друг с другом по модулю h.

Рассмотрим произвольный цикл перестановки. Пройдя по нему один раз, мы получим: $xS_A + yS_B + zS_C = 0$ (здесь x — количество шагов S_A в цикле, y — количество шагов S_B и z — количество шагов S_C).

Подставив в это равенство значения шагов, получаем:

$$(x + y)(b + c) = (y + z)(a + b).$$

Т. к. НОД $(a + b, b + c) = h$, то

$$x + y \geqslant \frac{a + b}{h} \qquad \text{и} \qquad y + z \geqslant \frac{b + c}{h}.$$

Сложив эти неравенства, получаем:

$$x + 2y + z \geqslant \frac{a + 2b + c}{h}.$$

Пусть все числа в нашем цикле сравнимы с числом k по модулю h, где $1 \leqslant k \leqslant h$. Легко видеть, что $y \geqslant [\frac{a+b-k}{h}] - [\frac{a-k}{h}]$. Следовательно, $x + y + z \geqslant [\frac{a+b+c-k}{h}] + 1 = [\frac{n-k}{h}] + 1$.

Т. к. количество чисел, сравнимых с k по модулю h, равно $[\frac{n-k}{h}] + 1$, то в выбранном нами цикле будут *все* числа, сравнимые с k по модулю h.

Следовательно, в нашей перестановке ровно h циклов, что и требовалось доказать. □

Замечание. Из доказательства теоремы следует, что циклы (C, B, A)-перестановки состоят из всех чисел, сравнимых друг с другом по модулю h. В частности, длина диаграммы Юнга равна $[\frac{n-1}{h}] + 1$.

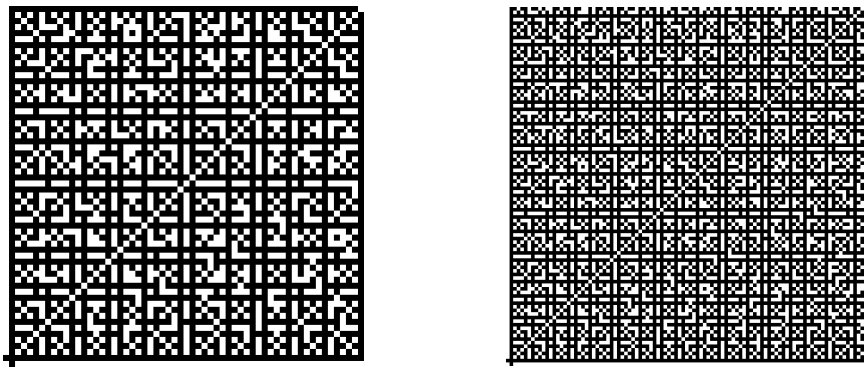
Следствие 1. *Доля $\delta_n(h)$ перестановок Арнольда, имеющих диаграмму Юнга высоты h площади n, асимптотически равна $\frac{1}{\zeta(2)} \cdot \frac{1}{h^2}$ при $n \to \infty$ и фиксированном h.*

Доказательство. Рассмотрим перестановку Арнольда $\sigma(a, b, c)$. Положим $x := \frac{b+c}{h} = \frac{n-a}{h}$ и $y := \frac{a+b}{h} = \frac{n-c}{h}$. Тогда согласно теореме 3 множество перестановок Арнольда с h циклами соответствует множеству точек со взаимно простыми координатами в

$$\triangle_n = \{(x, y) \in \mathbb{Z}^2 : x < n/h,\ y < n/h,\ x + y > n/h\}.$$

Таким образом, нам необходимо вычислить долю точек в $\triangle_n$ со взаимно простыми координатами при $n \to \infty$. Для этого мы воспользуемся следующей теоремой.

Теорема (Арнольда о равномерной распределенности; см. [2], см. также III.6)**.** *Множество целочисленных точек со взаимно простыми координатами равномерно распределено на плоскости, т.е. число точек этого множества в гомотетично растянутой в N раз области плоскости становится асимптотически пропорциональным произведению площади этой области на число N^2 при $N \to \infty$. Коэффициент этой пропорциональности (плотность) оказывается равным $1/\zeta(2) = 6/\pi^2$.*



Равномерное распределение: черным цветом показаны точки со взаимно простыми координатами, а белым — остальные.

Применим теорему Арнольда о равномерной распределенности к выпуклым оболочкам множеств $\triangle_n$. Их площади асимптотически равны $|\triangle_n|$, а количество перестановок Арнольда асимптотически равно $|\triangle_n|h^2$, поэтому доля перестановок Арнольда с h циклами

асимптотически (при $n \to \infty$) равна $\frac{1}{\zeta(2)}\frac{1}{h^2}$, что и требовалось доказать. □

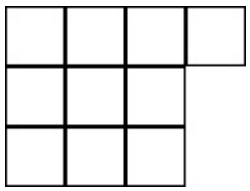
4. Диаграммы Юнга и их средние параметры

Используя следствие 1, мы вычислим средние параметры диаграмм Юнга (C, B, A)-перестановок. Для этого нам понадобится следующее определение.

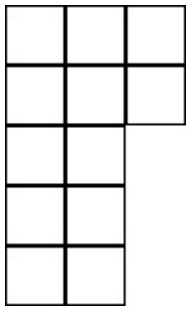
Определение. Назовем диаграмму Юнга *календарной*, если длины ее строк различаются не более чем на 1.

Из доказательства теоремы 3 следует следующее красивое свойство диаграмм Юнга (C, B, A)-перестановок.

Теорема 4. *Все диаграммы Юнга (C, B, A)-перестановок календарны.*



Перестановка $\sigma(1, 2, 7)$



Перестановка $\sigma(2, 3, 7)$

Теорема 5. *Календарная диаграмма Юнга однозначно задается своей высотой.*

Доказательство. В самом деле, длины строк календарной диаграммы Юнга однозначно выражаются через ее площадь и высоту. □

Теперь мы готовы вычислить средние размеры диаграмм Юнга перестановок Арнольда.

Теорема 6. *1. Средняя высота диаграмм Юнга перестановок Арнольда асимптотически равна $\frac{1}{\zeta(2)} \ln n$.*

2. Средняя длина диаграмм Юнга перестановок Арнольда асимптотически равна $\frac{\zeta(3)}{\zeta(2)} n$.

3. Средняя плотность диаграмм Юнга перестановок Арнольда асимптотически равна 1.

4. Средняя вертикальная асимметрия диаграмм Юнга перестановок Арнольда асимптотически равна $\frac{1}{\zeta(2)}$.

5. Средняя горизонтальная асимметрия диаграмм Юнга перестановок Арнольда асимптотически равна $\frac{\zeta(4)}{\zeta(2)}n$.

Доказательство. Если f — некоторая функция на диаграммах Юнга перестановок Арнольда, то согласно теорема 5 ее можно рассматривать как функцию от высоты h диаграммы Юнга. Поэтому ее среднее $\widehat{f}$ вычисляется по формуле

$$\widehat{f}(n) = \sum_{h=1}^{n} \delta_n(h) f(h) \sim \frac{1}{\zeta(2)} \cdot \sum_{h=1}^{n} \frac{f(h)}{h^2}.$$

Покажем, как вычислить, например, среднюю длину $\widehat{l}$ диаграммы Юнга. Полагая в предыдущей формуле $f(h) = l(h) = \left[\frac{n-1}{h}\right] + 1$, получаем:

$$\widehat{l}(n) \sim \frac{1}{\zeta(2)} \sum_{h=1}^{n} \frac{1}{h^2} \left(\left[\frac{n-1}{h}\right] + 1 \right) \sim \frac{1}{\zeta(2)} \sum_{h=1}^{n} \frac{n}{h^3} \sim \frac{\zeta(3)}{\zeta(2)} n.$$

Аналогично доказываются и другие формулы. □

Ниже представлено сравнение средних характеристик (C, B, A)-перестановок и перестановок общего положения, где $c \approx 0.6243$ — постоянная Голомба.

Средние	(C, B, A)-перестановки	Пер-ки общего положения
Высота $\widehat{h}$	$\frac{1}{\zeta(2)} \ln n$	$\ln n$
Длина $\widehat{l}$	$\frac{\zeta(3)}{\zeta(2)} n$	cn
Плотность $\widehat{\lambda}$	1	$\frac{1}{\ln n}$
Асимметрия $\widehat{\mu}$	$\frac{1}{\zeta(2)}$	$\frac{c_4 \ln n}{n}$
Асимметрия $\widehat{\eta}$	$\frac{\zeta(4)}{\zeta(2)} n$	?

В заключение рассмотрим вопрос о *предельной форме* календарных диаграмм Юнга.

Определение. Рассмотрим набор диаграмм Юнга, первые строки которых имеют общий квадрат, повернутых на 90° вокруг этого квадрата и сжатых в $\sqrt{n}$ раз. Тогда при $n \to \infty$ множество диаграмм Юнга будут заполнять некоторую область. Эту область мы и будем называть *предельной формой набора диаграмм Юнга.*

Теорема 7. *Предельная форма множества календарных диаграмм Юнга представляет собой криволинейный треугольник, ограниченный осями координат и гиперболой* $y = 1/x$.

Доказательство. Рассмотрим произвольную диаграмму Юнга. Самая длинная строка будет иметь длину $[\frac{n-1}{h}] + 1$, следующая за ней — $[\frac{n-2}{h}] + 1$, и т. д. до последней, которая будет иметь длину $[\frac{n-h}{h}] + 1$.

Располагая диаграмму Юнга так, как было описано выше, находим, что самая длинная строка ее имеет длину $([\frac{n-h}{h}] + 1)/n \sim 1/h$. Значит, предельная форма имеет указанный вид. □

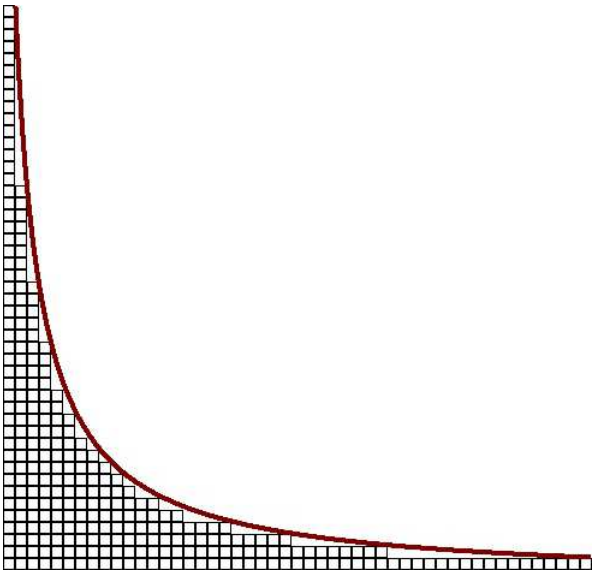


Рис. **.** $n = 100$

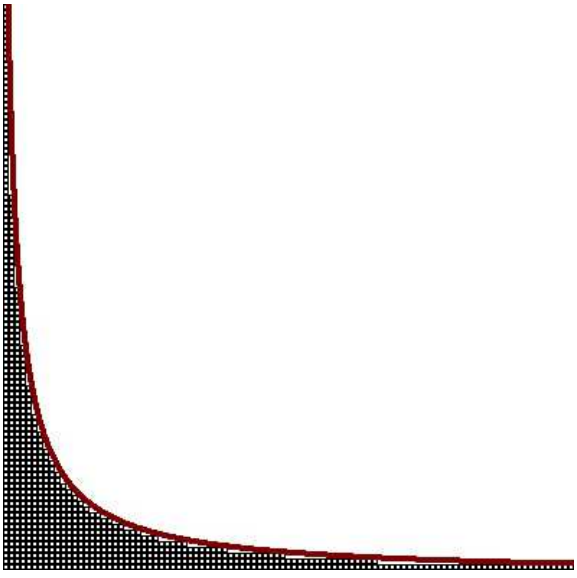


Рис. **.** $n = 200$

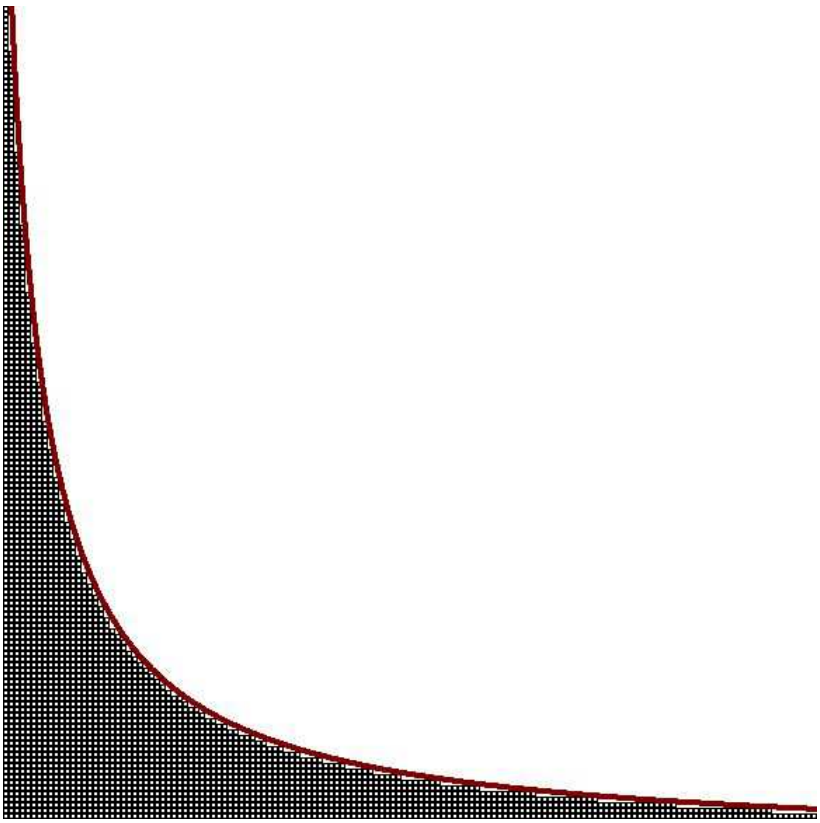


Рис. . $n = 1000$

Литература

[1] Арнольд В.И. *Задачи Арнольда*. М.: ФАЗИС, 2000 г.

[2] Арнольд В.И. *Равномерное распределение неделимых векторов в целочисленном пространстве*. Изв. РАН. Сер. матем. **79**:1 (2009), с. 21–29.

[3] Арнольд В.И. *Экспериментальное наблюдение математических фактов*. М.: МЦНМО, 2006 г.

О матричных аналогах функции Эйлера

Д. А. Байгушев[(1)]

1. Введение и обзор результатов

В теории чисел хорошо известна функция Эйлера. Она обладает большим количеством интересных свойств и встречается в задачах из самых разных областей математики (см., например, [6, 7]). Поэтому целесообразным представляется обобщить эту функцию на случай матриц (с определением матриц и их простыми свойствами можно ознакомиться в [5]).

Напомним, что функция Эйлера ставит в соответствие числу m количество обратимых элементов из кольца $\mathbb{Z}_m$, т. е. $\varphi(m) := |\mathbb{Z}_m^*|$. Рассмотрим вместо $\mathbb{Z}_m$ множество $\mathrm{Mat}(2, \mathbb{Z}_m)$ матриц 2×2 с элементами из кольца $\mathbb{Z}_m$. Вместо $\mathbb{Z}_m^*$ рассмотрим множества $\mathrm{GL}(2, \mathbb{Z}_m)$ обратимых матриц и $\mathrm{SL}(2, \mathbb{Z}_m)$ матриц с определителем 1. Тогда определим матричные функции Эйлера следующим образом.

Определение (см. также [3])**.** *Первой матричной функцией Эйлера* назовем функцию $\Phi'(m) := |\mathrm{SL}(2, \mathbb{Z}_m)|$.

Второй матричной функцией Эйлера назовем функцию $\Phi(m) := |\mathrm{GL}(2, \mathbb{Z}_m)|$.

Целью данной работы является изучение матричных функций Эйлера. Результаты представлены в Таблице 1.

[(1)] *Лицей «Вторая школа», Москва, Россия*
e-mail: IDanila24@gmail.com

Связь трех функций Эйлера		
$\Phi(m) = \Phi'(m) \cdot \varphi(m)$		
φ	Φ'	Φ
Мультипликативность: если НОД $(a, b) = 1$, то		
$\varphi(ab) = \varphi(a) \cdot \varphi(b)$	$\Phi'(ab) = \Phi'(a) \cdot \Phi'(b)$	$\Phi'(ab) = \Phi(a) \cdot \Phi(b)$
Значения		
$m \cdot \prod\limits_{p\mid m} (1 - \frac{1}{p})$	$m^3 \cdot \prod\limits_{p\mid m} (1 - \frac{1}{p^2})$	$m^4 \cdot \prod\limits_{p\mid m} (1 - \frac{1}{p}) \cdot (1 - \frac{1}{p^2})$
Теорема Эйлера: для любого A из $\mathbb{Z}_m^*$, $\mathrm{SL}(2, \mathbb{Z}_m)$ или $\mathrm{GL}(2, \mathbb{Z}_m)$		
$A^{\varphi(m)} \equiv 1 \bmod m$	$A^{\Phi'(m)} \equiv E \bmod m$	$A^{\Phi(m)} \equiv E \bmod m$
Рост в среднем		
$\frac{m}{\zeta(2)}$	$\frac{m^3}{\zeta(3)}$	$m^4 \cdot \prod\limits_{p} \left(1 - \frac{1}{p^2} - \frac{1}{p^3} + \frac{1}{p^4}\right)$

Таблица 1.

Также естественно определить *n-мерную матричную функцию Эйлера* $\Phi'_n(m)$ как количество матриц размера $n \times n$ с элементами из $\mathbb{Z}_m$ с определителем 1, т.е. $\Phi'_n(m) := |\mathrm{SL}(n, \mathbb{Z}_m)|$. В разделе 4 найдена формула для значений $\Phi'_n(m)$, а также изучен ее рост в среднем. А именно,

$$\Phi'_n(m) \widehat{\sim} m^{n^2-1} \cdot \prod_p \left(\frac{1}{p} \cdot \left(1 - \frac{1}{p^n}\right)^{n-1} + 1 - \frac{1}{p}\right).$$

2. Значения матричных функций Эйлера

В этом разделе мы изучим значения матричных функций Эйлера (см. также [4]).

В Таблице 2 представлены первые 30 значений всех трех функций Эйлера.

m	$\Phi'(m)$	$\Phi(m)$	m	$\Phi'(m)$	$\Phi(m)$	m	$\Phi'(m)$	$\Phi(m)$
1	0	0	**11**	1320	13200	**21**	8064	96768
2	6	6	**12**	1152	4608	**22**	7920	79200
3	24	48	**13**	2184	26208	**23**	12144	267168
4	48	96	**14**	2016	12096	**24**	9216	73728
5	120	480	**15**	2880	23040	**25**	15000	300000
6	144	288	**16**	3072	24576	**26**	13104	157248
7	336	2016	**17**	4896	78336	**27**	17496	314928
8	384	1536	**18**	3888	23328	**28**	16128	193536
9	648	3888	**19**	6840	123120	**29**	24360	682080
10	720	2880	**20**	5760	46080	**30**	17280	138240

Таблица 2.

Теорема 1. *Имеет место равенство* $\Phi(m) = \Phi'(m) \cdot \varphi(m)$.

Доказательство. Поставим в соответствие каждой матрице $M \in \mathrm{SL}(2, \mathbb{Z}_m)$ $\varphi(m)$ матриц из $\mathrm{GL}(2, \mathbb{Z}_m)$.

Пусть

$$\mathbb{Z}_m^* = \{k_1, k_2, ..., k_{\varphi(m)}\} \quad \text{и} \quad K_i := \begin{pmatrix} k_i & 0 \\ 0 & 1 \end{pmatrix}.$$

Поставим в соответствие матрице M матрицы $M \cdot K_1, \ldots, M \cdot K_{\varphi(m)}$.

Докажем, что разным матрицам из $\mathrm{SL}(2, \mathbb{Z}_m)$ мы поставили в соответствие разные матрицы. Допустим, что $M_1 \cdot K_a \equiv M_2 \cdot K_b \pmod m$. Тогда $M_2^{-1} \cdot M_1 \equiv K_b \cdot K_a^{-1}$ и

$$1 = \det(M_2^{-1} \cdot M_1) \equiv \det(K_b \cdot K_a^{-1}) = \frac{\det K_b}{\det K_a} = \frac{b}{a}.$$

Значит, $a \equiv b \pmod m$.

Кроме этого, каждая матрица из $\mathrm{GL}(2, \mathbb{Z}_m)$ может быть получена таким способом.

Итак, $|\mathrm{GL}(2, \mathbb{Z}_m)| = |\mathrm{SL}(2, \mathbb{Z}_m)| \cdot |\mathbb{Z}_m^*|$, т. е. $\Phi(m) = \Phi'(m) \cdot \varphi(m)$. □

Теорема 2. *Функция* $\Phi(m)$ *мультипликативна, т.е.* $\Phi(ab) = \Phi(a) \cdot \Phi(b)$ *при* $НОД(a, b) = 1$.

Доказательство. Поставим в соответствие каждому целому числу $A \in [0, m^4)$ матрицу из $\mathrm{Mat}(2, \mathbb{Z}_m)$ следующим образом. Пусть $A = \overline{a_1 a_2 a_3 a_4}_m$ — представление числа A в m-ичной системе счисления. Сопоставим числу A матрицу $M_A := \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}$.

Теперь построим следующую таблицу:

$a^4 \cdot 0 + 0$	$a^4 \cdot 0 + 1$	$\cdots$	$a^4 \cdot 0 + (a^4 - 1)$
$a^4 \cdot 1 + 0$	$a^4 \cdot 1 + 1$	$\cdots$	$a^4 \cdot 1 + (a^4 - 1)$
$\vdots$	$\vdots$	$\cdots$	$\vdots$
$a^4 \cdot (b^4 - 1) + 0$	$a^4 \cdot (b^4 - 1) + 1$	$\cdots$	$a^4 \cdot (b^4 - 1) + (a^4 - 1)$

В каждом столбце стоят разные числа по модулю b^4, а в строке — по модулю a^4, при этом во всей таблице стоят разные остатки по модулю $(ab)^4$. Значит, если рассматривать числа в таблице как матрицы из $\mathrm{Mat}(2, \mathbb{Z}_{ab})$, в строках стоят разные матрицы по модулю a, а в столбцах — по модулю b.

Заметим, что если определитель матрицы взаимно прост с a и b, то он взаимно прост с ab, и наоборот (т.к. $\mathrm{НОД}(a, b) = 1$). Поэтому количество обратимых матриц равно $\Phi(ab) = \Phi(a) \cdot \Phi(b)$. □

Теперь найдем формулу для вычисления значений матричных функций Эйлера Φ' и Φ. Из теорем 1 и 2 следует, что достаточно вычислить значение первой матричной функции Эйлера Φ' при $m = p^k$.

Лемма. *Имеет место равенство*

$$\Phi'(p^k) = p^{3k} \cdot \left(1 - \frac{1}{p^2}\right).$$

Доказательство. Нам надо найти количество решений сравнения $ad - bc \equiv 1 \bmod p^k$. Рассмотрим два случая.

1) $a = 0$. Тогда d — любое и $bc \equiv -1 \bmod p^k$. Значит, b — делитель единицы в $\mathbb{Z}_{p^k}$ и однозначно задает c. Количество таких решений равно $p^k \cdot \varphi(p^k) = p^{2k} - p^{2k-1}$.

2) $a = a' \cdot p^l$ и НОД$(a', p) = 1$. Количество таких a равно $\varphi(p^{k-l}) = p^{k-l} - p^{k-l-1}$. Тогда

$$bc + 1 \equiv 0 \mod p^l \quad \text{и} \quad d \equiv \left(\frac{bc+1}{p^l}\right) \cdot (a')^{-1} \mod p^{k-l}.$$

Значит, в $\mathbb{Z}_{p^l}$ элемент b — делитель единицы и однозначно задает c. Поэтому количество таких пар (b, c) в $\mathbb{Z}_{p^k}$ равно

$$\varphi(p^l) \cdot p^{k-l} \cdot p^{k-l} = p^{2k-l} - p^{2k-l-1}.$$

Далее, значение d в $\mathbb{Z}_{p^{k-l}}$ задается однозначно. Поэтому количество различных d в $\mathbb{Z}_{p^k}$ равно p^l.

Итак, количество решений сравнения $ad - bc \equiv 1 \mod p^k$ равно

$$(p^{2k} - p^{2k-1}) + \sum_{l=0}^{k-1} \left((p^{k-l} - p^{k-l-1}) \cdot (p^{2k-l} - p^{2k-l-1}) \cdot p^l\right) = p^{3k} - p^{3k-2}.$$

□

Из этой леммы с помощью теорем 1 и 2 мы немедленно получаем следующее следствие.

Следствие 1. *Если $m = p_1^{\alpha_1} \cdot \ldots \cdot p_s^{\alpha_s}$ —разложение m на простые множители, то*

$$\Phi'(m) = m^3 \cdot \left(1 - \frac{1}{p_1^2}\right) \cdot \ldots \cdot \left(1 - \frac{1}{p_s^2}\right)$$
$$\Phi(m) = m^4 \cdot \left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_1^2}\right) \cdot \ldots \cdot \left(1 - \frac{1}{p_s}\right)\left(1 - \frac{1}{p_s^2}\right).$$

Теорема 3. *1) Если $A \in \mathrm{SL}(2, \mathbb{Z}_m)$, то $A^{\Phi'(m)} \equiv E \ mod\ m$.*

2) Если $A \in \mathrm{GL}(2, \mathbb{Z}_m)$, то $A^{\Phi(m)} \equiv E \ mod\ m$.

Доказательство. Мы докажем только п.1), т.к. п.2) доказывается аналогично.

Пусть $\mathrm{SL}(2, \mathbb{Z}_m) = \{X_1, ..., X_{\Phi'(m)}\}$. Тогда множество матриц

$$\{X_1 \cdot A, X_2 \cdot A, \ldots, X_{\Phi'(m)} \cdot A\}$$

совпадает с $\mathrm{SL}(2, \mathbb{Z}_m)$.

Имеем:

$$(X_1 \cdot A) \cdot (X_2 \cdot A) \cdot \ldots \cdot (X_{\Phi'(m)} \cdot A) \equiv X_1 \cdot X_2 \cdot \ldots \cdot X_{\Phi'(m)} \mod m.$$

Отсюда $A^{\Phi'(m)} \equiv E \bmod m$. □

3. Рост в среднем матричных функций Эйлера

Из Таблицы 2 видно, что матричные функции Эйлера быстро растут (уже при $m = 11$ значения функций превосходят 10^3, а при $m = 40$ они превосходят 10^4). Также видно, что их значения «скачут», т. е. при малых изменениях аргумента наблюдается большой разброс значений. Особенно хорошо это видно на графике первой матричной функции Эйлера, построенном в логарифмических координатах $(\lg m, \lg \Phi'(m))$.

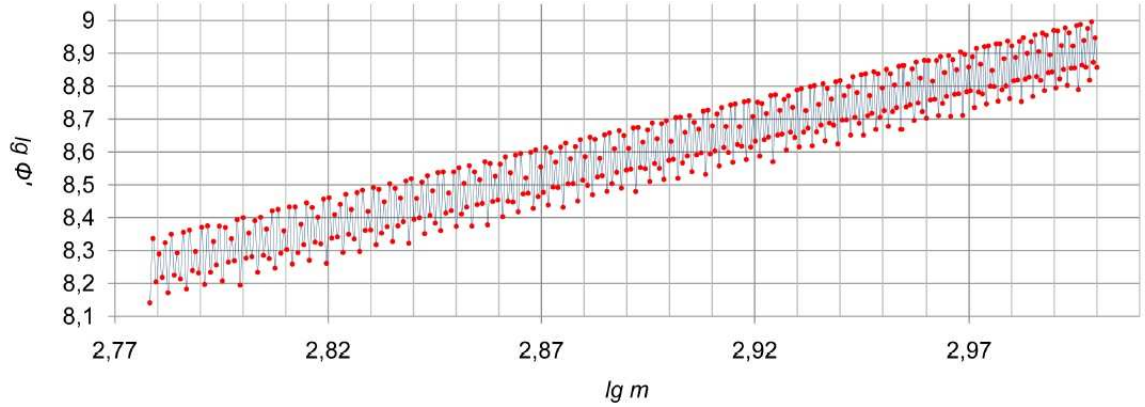


График функции Φ' в логарифмических координатах

В этом разделе мы исследуем рост матричных функций Эйлера. Для этого нам понадобится следующее определение (см. [2] и также VI.4).

Определение. Будем говорить, что функция f *растет в среднем* так же, как функция g (обозначим $f \mathrel{\widehat{\sim}} g$), если

$$\frac{f(1) + \ldots + f(m)}{g(1) + \ldots + g(m)} \to 1 \quad \text{при } m \to \infty.$$

Иначе говоря, $f \mathrel{\widehat{\sim}} g$, если средние арифметические $\frac{f(1)+\ldots+f(m)}{m}$ и $\frac{g(1)+\ldots+g(m)}{m}$ асимптотически равны, т.е. их отношение стремится к 1 при $m \to \infty$.

Рост в среднем используется для оценки «скачущих» функций, таких как матричная функция Эйлера. Идея заключается в том, что «скачки» вверх и вниз компенсирую друг друга при подсчете среднего арифметического, в результате чего рост усредненных функций будет исследовать легче.

Этот подход оказывается полезен уже при изучении обычной функции Эйлера. Так, хорошо известно, что $\varphi(m) \mathrel{\widehat{\sim}} \frac{m}{\zeta(2)}$ (см. [2] и VI.4), где ζ — дзета-функция Римана.

Наша цель — изучить рост в среднем матричных функций Эйлера (см. также замечания Арнольда по этому вопросу в [3, с. 37]).

3.1. Первая матричная функция Эйлера

Мы начнем с первой матричной функции Φ'.

Теорема 4. *Имеет место следующая асимптотика в среднем:*

$$\Phi'(m) \mathrel{\widehat{\sim}} \frac{m^3}{\zeta(3)}.$$

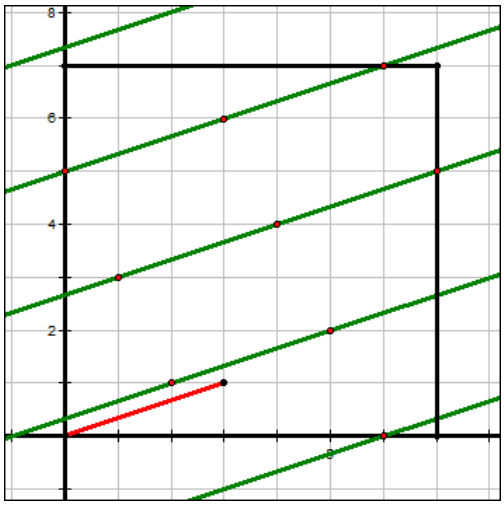
Замечание. Константа $1/\zeta(3)$ примерно равна 0,8319074.

Доказательство. Для каждого фиксированного $M \in \mathbb{N}$ рассмотрим в трехмерном пространстве $\mathbb{Z}^3$ с координатами (x, y, m) пирамиду P', заданную неравенствами $0 \leqslant x, y \leqslant m \leqslant M$.

Покрасим точки в пирамиде P' в черный цвет, если $\text{НОД}(x, y, m) = 1$, и в белый в противном случае.

1. Поставим в соответствие каждой черной точке m матриц с определителем 1, причем разным точкам мы поставим в соответствие разные матрицы. Пусть $A(x, y, m)$ — фиксированная черная точка. Рассмотрим в плоскости $\mathbb{Z}_m^2$ вектор v с координатами $(x', y') :=$

$(x, y)/\text{НОД}(x, y)$. Проведем «целочисленную» прямую l (т.е. множество решений уравнения вида $b \cdot x' - a \cdot y' \equiv 1$), параллельную v и проходящую на расстоянии $\frac{1}{|v|}$ от вектора v, где $|v|$ — его длина.



Пример для точки $A(1, 3, 7)$

Заметим, что эта прямая проходит ровно через m точек из $\mathbb{Z}_m^2$ (см. рисунок выше): не умаляя общности можно считать, что $x' \neq 0$, тогда a может быть произвольным и $b \equiv \frac{ay'+1}{x'} \bmod m$.

Обозначим векторы с концами в этих целочисленных точках через $u_1, \ldots, u_m$. Площадь параллелограмма, натянутого на векторы v и u_i всегда сравнима с 1 по модулю m для всех $i = 1, \ldots, m$.

Построим m матриц M_i' с определителем 1, записав в их столбцы координаты векторов v и u_i, т.е. $M_i' := (v|u_i)$. Преобразуем их в m матриц $M_i = (v \cdot \text{НОД}(x, y) \mid u_i/\text{НОД}(x, y))$ с определителем 1.

Наконец, поставим в соответствие точке A матрицы $\{M_1, \ldots, M_m\}$. Легко видеть, что разным точкам мы поставили в соответствие разные матрицы (т. к. первые столбцы у них разные), и каждой матрице соответствует единственная черная точка.

2. Рассмотрим четырехмерное пространство $\mathbb{Z}^4 = \mathbb{Z}^3 \times \mathbb{Z}^1$ с координатами (x, y, m, h). Достроим пирамиду $P' \subset \mathbb{Z}_3$ до четырехмерной пирамиды P (заданной неравенствами $0 \leqslant x, y, h \leqslant m \leqslant M$), надстроив над каждой точкой столбец высотой m и цветом, совпадающим с цветом точки. Тогда из п.1 следует, что количество черных точек в P равно $\Phi'(1) + \ldots + \Phi'(M)$.

Вероятность выбрать черную точку в P равна

$$\rho = \frac{\Phi'(1) + \ldots + \Phi'(M)}{1^3 + \ldots + M^3}.$$

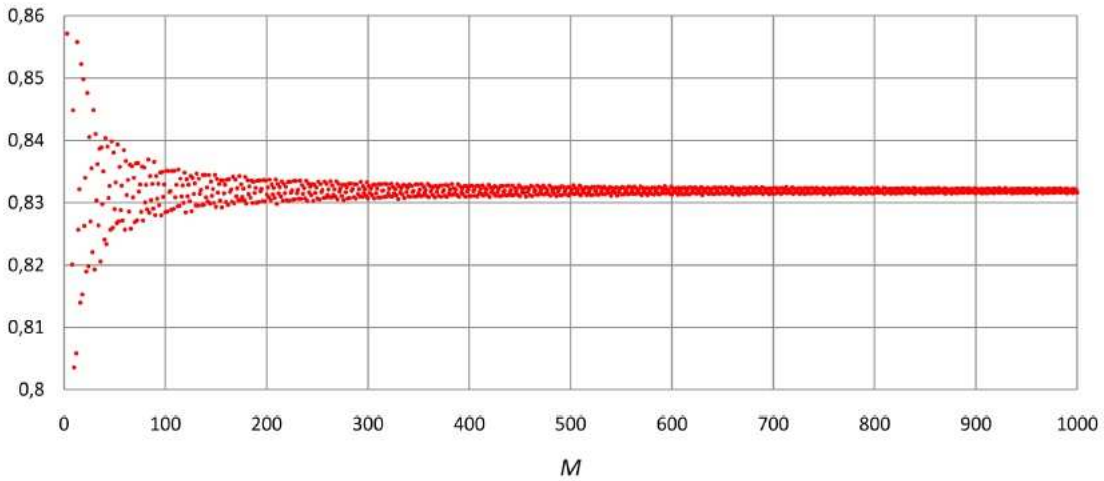
Теперь найдем предел этой вероятности при $M \to \infty$. Для этого достроим пирамиду P до куба C со стороной M белыми точками. Новая вероятность ρ' выбрать черную точку в кубе C равна $\rho \cdot \frac{\mathrm{Vol}(P)}{\mathrm{Vol}(C)}$, где Vol — объем.

Посчитаем вероятность выбрать черную точку в C. Вероятность того, что три числа (x, y, m) взаимно просты, стремится к $\frac{1}{\zeta(3)}$ при $M \to \infty$. Вероятность того, что точка принадлежит пирамиде P, равна отношению объемов $\frac{\mathrm{Vol}(P)}{\mathrm{Vol}(C)}$.

Значит, вероятность выбрать черную точку в C стремится к $\frac{\mathrm{Vol}(P)}{\mathrm{Vol}(C)} \cdot \frac{1}{\zeta(3)}$ при $M \to \infty$.

Таким образом, $\rho \to \frac{1}{\zeta(3)}$, т.е. $\Phi(m) \mathrel{\widehat{\sim}} \frac{m^3}{\zeta(3)}$, что и требовалось доказать. □

Отметим, что теорема 4 хорошо подтверждается численными экспериментами (см. рисунок и Таблицу 3).



Отношения $\sum\limits_{m \leqslant M} \Phi'(m) / \sum\limits_{m \leqslant M} m^3$ при $M \leqslant 1000$

M	100	250	500	1000	2000
$\sum\limits_{m\leqslant M}\Phi'(m)/\sum\limits_{m\leqslant M}m^3$	0,**8**28496	0,**83**0673	0,**831**615	0,**831**645	0,**831**859
константа $1/\zeta(3)$	0,831907	0,**831**907	0,**831**907	0,**831**907	0,**831**907
погрешность	$\sim 10^{-2}$	$\sim 10^{-3}$	$\sim 10^{-4}$	$\sim 10^{-4}$	$\sim 10^{-5}$

Таблица 3.

3.2. Вторая матричная функция Эйлера

Теперь изучим рост в среднем второй матричной функции Эйлера Φ.

Теорема. *Имеет место следующая асимптотика:*

$$\Phi(m) \mathrel{\widehat{\sim}} m^4 \cdot \prod_{p\ -\ \text{простое}} \Big(1-\frac{1}{p^2}-\frac{1}{p^3}+\frac{1}{p^4}\Big).$$

Замечание. Константа $\prod\limits_p (1-\frac{1}{p^2}-\frac{1}{p^3}+\frac{1}{p^4})$ примерно равна 0,535896, т.е. обратимых матриц среди произвольных в среднем чуть больше половины.

Доказательство. Для каждого фиксированного $M \in \mathbb{N}$ рассмотрим в четырехмерном пространстве $\mathbb{Z}^4$ с координатами (x, y, d, m) пирамиду P', заданную неравенствами $0 \leqslant x, y, d \leqslant m \leqslant M$.

Покрасим точки в пирамиде P' в черный цвет, если НОД$(x, y, m) = 1$ и НОД$(d, m) = 1$, и в белый в противном случае.

Аналогично п.1 доказательства теоремы 4, поставим в соответствие каждой черной точке m обратимых матриц, причем разным точкам поставим в соответствие разные матрицы.

Далее, рассмотрим пятимерное пространство $\mathbb{Z}^5 = \mathbb{Z}^4 \times \mathbb{Z}^1$ с координатами (x, y, d, m, h). Достроим пирамиду $P' \subset \mathbb{Z}^4$ до пятимерной пирамиды P (заданной неравенствами $0 \leqslant x, y, d, h \leqslant m \leqslant M$), надстроив над каждой точкой столбец высотой m и цветом, совпадающим с цветом точки. Заметим, что количество черных точек в P равно $\Phi(1) + \ldots + \Phi(M)$.

Вероятность выбрать черную точку в P равна

$$\rho = \frac{\Phi(1) + \ldots + \Phi(M)}{1^4 + \ldots + M^4}.$$

Посчитаем вероятность выбрать черную точку в P другим способом. Зафиксируем простое число p и рассмотрим два случая.

1) $m \vdots p$ (вероятность этого равна $\frac{1}{p}$). Тогда НОД$(x, y) \not\vdots p$ с вероятностью $1 - \frac{1}{p^2}$ и $d \not\vdots p$ с вероятностью $1 - \frac{1}{p}$. Поэтому вероятность всех этих событий равна

$$\frac{1}{p} \cdot \Big(1 - \frac{1}{p^2}\Big) \cdot \Big(1 - \frac{1}{p}\Big).$$

2) $m \not\vdots p$ (вероятность этого равна $(1 - \frac{1}{p})$).

Таким образом, вероятность выбрать черную точку в P равна

$$\rho = \prod_{p \leqslant M} \Big(\frac{1}{p} \cdot \big(1 - \frac{1}{p^2}\big) \cdot \big(1 - \frac{1}{p}\big) + \big(1 - \frac{1}{p}\big)\Big).$$

Наконец, при $M \to \infty$ получаем

$$\Phi(m) \,\widehat{\sim}\, m^4 \cdot \prod_p \Big(1 - \frac{1}{p^2} - \frac{1}{p^3} + \frac{1}{p^4}\Big),$$

что и требовалось доказать. □

Замечание. Заметим, что

$$\prod_p \Big(1 - \frac{1}{p^2} - \frac{1}{p^3} + \frac{1}{p^4}\Big) = \frac{c}{\zeta(2) \cdot \zeta(3)},$$

где

$$c = \prod_p \Big(1 + \frac{1}{(p^3 - 1) \cdot (p + 1)}\Big) \approx 1{,}05963.$$

В [1] В.И Арнольд ставит вопрос о связи роста в среднем произведения двух функций и произведения ростов в среднем самих этих функций. В рассмотренных им и А.А. Карацубой примерах произведение средних оказывалось во много раз больше среднего

произведения, причем это их отношение стремилось к бесконечности с ростом m.

В случае трех функций Эйлера имеем

$$\varphi(m) \widehat{\sim} \frac{m}{\zeta(2)}, \quad \Phi'(m) \widehat{\sim} \frac{m^3}{\zeta(3)} \quad \text{и} \quad \Phi(m) = \Phi'(m)\varphi(m) \widehat{\sim} \frac{m^4}{\zeta(2)\zeta(3)} \cdot c,$$

т.е. среднее произведение отличается от произведения средних на константу c, очень близкую к 1. По всей видимости, это первый «нетривиальный» пример подобной ситуации.

Судя по всему, это говорит о некоторой (еще не известной) связи между ростом трех матричных функций Эйлера: несмотря на то, что сам их рост очень нерегулярен (есть большие «скачки» значений), эта нерегулярность схожа для всех трех функций Эйлера (т.е. «скачки» приходятся на примерно одни и те же значения аргументов).

4. Обобщения: n-мерные матричные функции Эйлера

По аналогии с матричными функциями Эйлера для матриц размера 2×2, в этом разделе мы введем и исследуем матричные функции Эйлера для матриц произвольного размера.

Определение. *Первой n-мерной матричной функцией Эйлера* назовем функцию $\Phi'_n(m) := |\mathrm{SL}(n, \mathbb{Z}_m)|$.

Второй n-мерной матричной функцией Эйлера назовем функцию $\Phi_n(m) := |\mathrm{GL}(2, \mathbb{Z}_m)|$.

Ниже мы исследуем значения этих функций и асимптотику роста в среднем.

4.1. Значения

Теорема 5. *Матричные функции Эйлера обладают следующими свойствами:*

1) $\Phi_n(m) = \Phi'_n(m) \cdot \varphi(m)$;

2) $\Phi_n(ab) = \Phi_n(a) \cdot \Phi_n(b)$ *при* $НОД(a, b) = 1$;

3) Если $A \in \mathrm{SL}(n, \mathbb{Z}_m)$, *то* $A^{\Phi'_n(m)} \equiv E \ mod\ m$.
4) если $A \in \mathrm{GL}(n, \mathbb{Z}_m)$, *то* $A^{\Phi_n(m)} \equiv E \ mod\ m$.

Доказательства этих свойств получаются несложной модификацией аналогичных свойств для двумерных матричных функций Эйлера.

Гораздо более сложным оказывается вопрос о явных формулах для значений n-мерных матричных функций Эйлера. Для их нахождения нам потребуется следующая лемма.

Лемма. *Любая* d*-мерная плоскость* $H \subset \mathbb{Z}_m^n$ *состоит из* m^d *точек.*

Доказательство. Плоскость H есть множество векторов (или, что тоже самое, точек) вида $t_1a_1 + \ldots + t_da_d + b$, где $\{a_1, \ldots, a_d, b\}$ — независимые векторы и $t_1, \ldots, t_d \in \mathbb{Z}_m$.

Рассмотрим пространство $\mathbb{Z}_p^d = \langle a_1, a_2, \ldots, a_d \rangle$. Тогда каждый набор $\{t_1, \ldots, t_d\}$ есть просто точка в этом пространстве. Т.к. в пространстве $\mathbb{Z}_m^d$ ровно m^d точек, то и плоскость H также содержит ровно m^d точек. □

Теорема 6. *Имеет место равенство*

$$\Phi_n(p^k) = (p^k)^{n^2} \cdot \left(1 - \frac{1}{p}\right) \cdot \ldots \cdot \left(1 - \frac{1}{p^n}\right).$$

Доказательство. Заметим, что достаточно доказать утверждение теоремы для случая $k = 1$, т.к. случай $k > 1$ следует из него по лемме Гензеля (см. [6]).

Т.к. p — простое, то $\mathrm{GL}(n, \mathbb{Z}_p)$ есть множество матриц с определителем, отличным от 0. Как известно (см. [5]), определитель матрицы — это объем параллелепипеда, натянутого на векторы-столбцы матрицы.

Значит, $\mathrm{GL}(n, \mathbb{Z}_p)$ состоит из матриц, векторы-столбцы которых независимы. Посчитаем количество таких наборов векторов столбцов.

По лемме, для первого вектора-столбца есть $p^n - 1$ вариантов (годится любой ненулевой вектор). Для второго вектора есть $p^n - p$

вариантов (т.к. на прямой, натянутой на первый вектор, лежит p векторов, и он не должен совпадать ни с одним из них), для третьего — $p^n - p^2$ вариантов, и т.д.

Таким образом, количество всех обратимых матриц равно

$$\Phi_n(p) = (p^n-1)\cdot(p^n-p)\cdot\ldots\cdot(p^n-p^{n-1}) = p^{n^2}\cdot\left(1-\frac{1}{p}\right)\cdot\ldots\cdot\left(1-\frac{1}{p^n}\right).$$

□

Следствие 2. *Если $m = p_1^{\alpha_1}\cdot\ldots\cdot p_s^{\alpha_s}$ —разложение m на простые множители, то*

$$\Phi_n(m) = m^{n^2}\cdot\prod_{i=1}^{s}\left(1-\frac{1}{p_i}\right)\cdot\ldots\cdot\left(1-\frac{1}{p_i^n}\right)$$

$$\Phi'_n(m) = m^{n^2-1}\cdot\prod_{i=1}^{s}\left(1-\frac{1}{p_i^2}\right)\cdot\ldots\cdot\left(1-\frac{1}{p_i^n}\right).$$

4.2. Рост в среднем

В этом разделе мы исследуем рост в среднем первой n-мерной матричной функции Эйлера.

Теорема 7. *Имеет место следующая асимптотика:*

$$\Phi'_n(m) \mathrel{\widehat{\sim}} m^{n^2-1}\cdot\prod_{p\ -\ \text{простое}}\left(\frac{1}{p}\cdot\left(1-\frac{1}{p^n}\right)^{n-1}+1-\frac{1}{p}\right).$$

Доказательство. Для каждой матрицы $B \in \mathrm{GL}(n,\mathbb{Z}_m)$ обозначим через $\overline{b}_1$, …, $\overline{b}_n$ векторы–столбцы, из которых она состоит. Координаты самих векторов $\overline{b}_i$ обозначим через b_i^1, …, b_i^n. Также для каждого i введем обозначение НОД$(\overline{b}_i)$:=НОД$(b_i^1,\ldots,b_i^n)$.

Зафиксируем натуральное число M и рассмотрим в пространстве $\mathbb{Z}^{n^2-n+1}$ с координатами $(b_1^1,\ldots,b_{n-1}^n;m)$ пирамиду

$$P' = \{(b_i^j;m) : 0 \leqslant b_1^1,\ldots,b_{n-1}^n \leqslant m \leqslant M\}.$$

Далее, для каждой точки $A(b_i^j; m) \in P'$ рассмотрим $n-1$ вектор $\bar{b}_1$, $\bar{b}_2$, …, $\bar{b}_{n-1}$, где $\bar{b}_i := (b_i^1, b_i^2, \ldots, b_i^n)$. Тогда покрасим точку в черный цвет, если

$$\text{НОД}(\bar{b}_1, m) = \text{НОД}(\bar{b}_2, m) = \ldots = \text{НОД}(\bar{b}_{n-1}, m) = 1 \qquad (*)$$

и векторы $\bar{b}_1$, $\bar{b}_2$, …, $\bar{b}_{n-1}$ независимы, и в белый в противном случае.

1. Поставим в соответствие каждой черной точке m^{n-1} обратимых матриц, причем разным точкам — разные матрицы.

Пусть $A(b_i^j, m)$ — фиксированная черная точка, $\bar{b}_1$, $\bar{b}_2$, …, $\bar{b}_{n-1}$ — соответствующие ей векторы. Рассмотрим в пространстве $\mathbb{Z}_m^{n-1}$ набор векторов

$$\{\bar{b}'_1, \bar{b}'_2, \ldots, \bar{b}'_{n-1}\}, \quad \text{где } \bar{b}'_i := \bar{b}_i / \text{НОД}(\bar{b}_i).$$

Построим гиперплоскость L, параллельную $\langle \bar{b}_1, \ldots, \bar{b}_{n-1} \rangle$ и проходящую на расстоянии $\frac{1}{\text{Vol}(W)}$ от нее (здесь W — параллелепипед, натянутый на вектора $\bar{b}_1$, $\bar{b}_2$, …, $\bar{b}_{n-1}$).

Аналогично случаю $n = 2$, гиперплоскость L проходит ровно через m^{n-1} точек с целочисленными координатами: $\bar{b}_n^1$, $\bar{b}_n^2$, …, $\bar{b}_n^{m^{n-1}}$. Площадь параллелепипеда, натянутого на векторы $\bar{b}_1$, $\bar{b}_2$, …, $\bar{b}_{n-1}$, $\bar{b}_n^i$ всегда равна 1.

Наконец, построим m^{n-1} матриц $M'_i := (\bar{b}_1 | \ldots | \bar{b}_{n-1} | \bar{b}_n^i)$ с определителем 1. Преобразуем их в m^{n-1} производных матриц

$$M_i := (\bar{b}_1 \cdot \text{НОД}(\bar{b}_1) | \ldots | \bar{b}_{n-1} \cdot \text{НОД}(\bar{b}_{n-1}) | \bar{b}_n^i \cdot t),$$

где

$$t \equiv \prod_{i=1}^{n-1} \text{НОД}(\bar{b}_i)^{-1} \bmod m.$$

Поставим в соответствие точке A матрицы $\{M_1, \ldots, M_m^{n-1}\}$. Легко видеть, что разным точкам мы поставили в соответствие разные матрицы, и каждой матрице соответствует единственная черная точка.

2. Рассмотрим пространство $\mathbb{Z}^{n^2-n+2} = \mathbb{Z}^{n^2-n+1} \times \mathbb{Z}^1$ и достроим пирамиду P' до пирамиды P, надстроив над каждой точкой столбец

высотой m^{n-1} и цветом, совпадающим с цветом точки. Тогда из п.1 следует, что количество черных точек в P равно $\Phi'_n(1)+\ldots+\Phi'_n(M)$.

3. Вероятность выбрать черную точку в P равна

$$\rho = \frac{\Phi'_n(1)+\ldots+\Phi'_n(M)}{1^{n^2-1}+\ldots+M^{n^2-1}}.$$

Теперь вычислим предел вероятности ρ при $M \to \infty$.

Прежде всего докажем, что вероятность P_{ind} того, что векторы $\overline{b}_1$, $\overline{b}_2$, $\ldots$, $\overline{b}_{n-1}$, удовлетворяющие условию $(*)$, независимы, стремится к 1 при $M \to \infty$.

Пусть $N(M)$ — количество наборов из $n-1$ зависимых векторов, удовлетворяющих условию $(*)$, и $F(M)$ — количество точек в пирамиде P, удовлетворяющих условию $(*)$. Как известно, $F(M) \sim \zeta(n+1)\cdot M^n$ при $M \to \infty$. Кроме того, из леммы о количество точек в d-мерной плоскости следует, что

$$N(M) \leqslant (m^n)^{n-2}\cdot m^{n-2} = m^{n^2-n-2}.$$

Таким образом,

$$P_{\text{ind}} = 1 - \frac{N(M)}{F(M)^{n-1}} \geqslant 1 - \frac{M^{n^2-n-2}}{M^{n^2-n}\cdot\zeta(n+1)^{n-1}} = \\ = 1 - \frac{1}{M^2\cdot\zeta(n+1)^{n-1}} \to 1$$

при $M \to \infty$.

Таким образом, предел вероятности ρ при $M \to \infty$ равен пределу вероятности выбрать точку в P, удовлетворяющую условию $(*)$.

Посчитаем эту вторую вероятность. Для этого зафиксируем простое число p и рассмотрим два случая.

1) $m \vdots p$ (вероятность этого равна $\frac{1}{p}$). Тогда НОД$(\overline{b}_1)$, $\ldots$, НОД$(\overline{b}_{n-1})$ не делятся на p с вероятностью $\left(1-\frac{1}{p^n}\right)^{n-1}$. Поэтому вероятность всех этих событий равна $\frac{1}{p}\cdot\left(1-\frac{1}{p^n}\right)^{n-1}$.

2) $m \not\vdots p$ (вероятность этого равна $(1-\frac{1}{p})$).

Таким образом, вероятность выбрать точку в P, удовлетворяющую условию $(*)$, равна

$$\prod_{p\leqslant M}\left(\frac{1}{p}\cdot\left(1-\frac{1}{p^n}\right)^{n-1}+1-\frac{1}{p}\right).$$

Наконец, при $M\to\infty$ получаем

$$\Phi'_n(m)\mathrel{\widehat{\sim}} m^{n^2-1}\cdot\prod_{p}\left(\frac{1}{p}\cdot\left(1-\frac{1}{p^n}\right)^{n-1}+1-\frac{1}{p}\right),$$

что и требовалось доказать. □

Литература

[1] В. И. Арнольд. *Задачи семинара 2003–2004*. М., МЦНМО, 2005 г.

[2] В. И. Арнольд. *Группы Эйлера и арифметика геометрических прогрессий*. М.: МЦНМО, 2003.

[3] В. И. Арнольд. *Перестановки* // УМН, 2009, **64**:4(388), с. 3–44.

[4] Д. А. Байгушев. *О матричной функции Эйлера* // Труды математического центра им. Н. И. Лобачевского, т. 45, 2012, с. 12–14.

[5] Э. Б. Винберг. *Курс алгебры*. М.: Факториал, 2002 г.

[6] И. М. Виноградов. *Основы теории чисел*. М.: ГИТТЛ, 1952 г.

[7] Г. Дэвенпорт. *Высшая арифметика. Введение в теорию чисел*. М.: Наука, 1965 г.

О функции Эйлера алгебраических расширений колец вычетов

Г. А. Юргин[1]

1. Введение

В теории чисел огромную роль играет *функция Эйлера* φ, ставящая в соответствие натуральному числу $m > 1$ количество обратимых элементов кольца вычетов $\mathbb{Z}_m$, т.е. $\varphi(m) := |\mathbb{Z}_m^*|$. Представляется естественным изучить обобщения функции Эйлера на другие кольца. В работе [1] были изучены *матричные функции Эйлера*, ставящие в соответствие натуральному числу $m > 1$ количество обратимых матриц в кольцах $\mathrm{GL}(2, \mathbb{Z}_m)$ и $\mathrm{SL}(2, \mathbb{Z}_m)$.

Целью данной работы является изучение функции Эйлера алгебраических расширений колец вычетов $\mathbb{Z}_m$.

Пусть α является корнем многочлена $Q \in \mathbb{Z}[x]$ степени n и не является корнем никакого многочлена меньшей степени. Обозначим через $\mathbb{Z}[\alpha]$ множество чисел вида

$$a_{n-1}\alpha^{n-1} + a_{n-2}\alpha^{n-2} + \ldots + a_1\alpha + a_0,$$

где $a_i \in \mathbb{Z}$. На эти числа легко распространяются операции сложения и умножения. Частным случаем таких множеств являются числа вида $a + b\sqrt{-1}$, известные как *гауссовы целые числа*.

Аналогично, через $\mathbb{Z}_m[\alpha]$ будем обозначать множество чисел вида

$$a_{n-1}\alpha^{n-1} + a_{n-2}\alpha^{n-2} + \ldots + a_1\alpha + a_0,$$

[1] *Лицей «Вторая школа», Москва, Россия*
e-mail: g.y.98@mail.ru

где все $a_i \in \mathbb{Z}_m$, причем m взаимно просто со старшим коэффициентом Q (мы будем считать, что это условие выполнено далее везде).

Определение. Через $\varphi_\alpha(m)$ обозначим количество обратимых элементов кольца $\mathbb{Z}_m[\alpha]$, т.е. $\varphi_\alpha(m) := |\mathbb{Z}_m[\alpha]^*|$.

В работе доказаны основные свойства функции Эйлера $\varphi_\alpha(m)$, вычислены ее значения и исследован рост в среднем функции Эйлера $\varphi_i(m)$ колец вычетов гауссовых целых чисел $\mathbb{Z}_m[i]^*$. Также в работе исследуются обобщения алгебраических расширений колец вычетов по модулям гауссовых целых чисел, вычисляется соответствующая им функция Эйлера и исследуется ее асимптотика.

2. Примеры

Прежде всего приведем конкретные примеры вычислений значений функции Эйлера некоторых расширений колец вычетов $\mathbb{Z}_m$.

Пусть α — корень многочлена $x^2 - 2$. Тогда

$$\mathbb{Z}_m[\alpha] = \{a + b\sqrt{2} : a, b \in \mathbb{Z}_m\}.$$

Сложение и умножение таких чисел будет выглядеть следующим образом:

$$(a + b\sqrt{2}) + (c + d\sqrt{2}) = (a + c) + (b + d)\sqrt{2},$$
$$(a + b\sqrt{2}) \cdot (c + d\sqrt{2}) = (ac + 2bd) + (ad + bc)\sqrt{2}.$$

Значения $\varphi_{\sqrt{2}}(m)$ для разных m приведены в таблице (во второй строке представлено число элементов в кольце $\mathbb{Z}_m[\sqrt{2}]$. В первую строку включены только те числа, по модулю которых 2 — квадратичный вычет, т. к. в этом случае $x^2 - 2$ неприводим над $\mathbb{Z}_m$).

m	3	5	9	11	13	15	19	25	27	53	65	99
m^2	9	25	81	121	169	225	361	625	729	2809	4225	9801
$\varphi_{\sqrt{2}}(m)$	8	24	72	120	168	192	360	600	648	2808	4032	8640

2) Пусть α — корень многочлена $3x^2 + x - 1$. Тогда сложение элементов, как и в случае 1), происходит покомпонентно. А чтобы вычислить произведение двух элементов, можно их перемножить как многочлены от α, а потом избавиться от слишком больших степеней α, пользуясь тем, что $3\alpha^2 + \alpha - 1 = 0$. Например, по модулю 5

$$(2\alpha+1)(\alpha+3) = 2\alpha^2+7\alpha+3 = -3\alpha^2-3\alpha-2 = -(3\alpha^2+\alpha-1)-2\alpha-3 = 3\alpha+2.$$

Замечание. Условие взаимной простоты m и старшего коэффициента не может быть опущено, т. к. иначе возникают трудности при умножении. Если в примере выше положить, скажем, $m = 9$, то уже не ясно, как избавиться от члена $2\alpha^2$.

3. Свойства функции Эйлера $\varphi_\alpha(m)$

В этом разделе мы докажем основные свойства функции Эйлера $\varphi_\alpha(m)$.

Теорема 1. *Если α является корнем многочлена n-й степени, неприводимого над $\mathbb{Z}_p$, где p — простое, то $\varphi_\alpha(p) = p^n - 1$, т.е. $\mathbb{Z}_p[\alpha]$ является полем.*

Теорема 2 (Свойство мультипликативности)**.** *Если k и m — натуральные взаимно простые числа, причем α является корнем многочлена, неприводимого над $\mathbb{Z}_p$ для всех p, являющихся простыми делителями k или m, то*

$$\varphi_\alpha(k) \cdot \varphi_\alpha(m) = \varphi_\alpha(km).$$

Теорема 3 (О значениях $\varphi_\alpha(m)$)**.** *Если $m = p_1^{\gamma_1} \cdot \ldots \cdot p_t^{\gamma_t}$ — натуральное число и его разложение на простые, причем α является корнем многочлена n-й степени, неприводимого над $\mathbb{Z}_{p_i}$ для всех i, то*

$$\varphi_\alpha(m) = m^n \prod_{i=1}^{t} \left(1 - \frac{1}{p_i^n}\right).$$

Теорема 1 является известным утверждением о факторкольце $\mathbb{Z}_p[x]/(f)$, где $f \in \mathbb{Z}_p[x]$ — неприводимый многочлен (см. раздел XI.6).

Прежде чем переходить к доказательству теорем 2 и 3, докажем важное утверждение.

Утверждение. *Элемент*

$$z = a_{n-1}\alpha^{n-1} + a_{n-2}\alpha^{n-2} + \ldots + a_1\alpha + a_0,$$

где $z \in \mathbb{Z}_m[\alpha]$, *обратим тогда и только тогда, когда*

$$НОД(a_0, a_1, \ldots, a_{n-1}, m) = 1.$$

Доказательство. $\boxed{\Rightarrow}$ Если НОД$(a_0, a_1, \ldots, a_{n-1}, m) > 1$, то существует индекс i такой, что все коэффициенты делятся на p_i. Но тогда у любого элемента, полученного умножением z на произвольный элемент, все коэффициенты также будут делиться на p_i. Значит, элемент z необратим.

$\boxed{\Leftarrow}$ От противного. Допустим, что z необратим. Тогда найдется такой $w \neq 0$, что $zw = 0$. Из условия $w \neq 0$ следует, что НОД коэффициентов элемента w не равен 0 в $\mathbb{Z}_m$. Мы знаем, что НОД$(a_0, a_1, \ldots, a_{n-1}, m) = 1$. По лемме Гаусса НОД коэффициентов zw равен произведению НОДов коэффициентов z и w. Значит, $zw \neq 0$. Противоречие.

□

Перейдем к доказательству теорем 2 и 3.

Доказательство теоремы 2. Из доказанного утверждения получаем, что элемент z обратим по модулю mk тогда и только тогда, когда он обратим по модулю m и по модулю k. По китайской теореме об остатках каждый коэффициент в точности задается остатками по модулю m и k, т. е. элемент в точности задается двумя наборами остатков (по модулю m и по модулю k для каждого коэффициента). Число способов выбрать остатки по модулю m так, чтобы элемент был обратим по модулю m, равно $\varphi_\alpha(m)$, по модулю k — $\varphi_\alpha(k)$, значит, всего имеем $\varphi_\alpha(m)\varphi_\alpha(k)$ обратимых элементов по модулю mk, что и требовалось доказать. □

Доказательство теоремы 3. Случай 1. $m = p^\gamma$, где p — простое. Из из доказанного утверждения получаем, что элемент необратим тогда и только тогда, когда все его коэффициенты делятся на p. Всего есть n коэффициентов и $p^{\gamma-1}$ способов выбрать каждый так, чтобы он делился на p. Значит, необратимых элементов $p^{(\gamma-1)n}$, а обратимых $p^{\gamma n} - p^{(\gamma-1)n} = m^n(1 - \frac{1}{p^n})$.

Случай 2. $m = p_1^{\gamma_1} p_2^{\gamma_2} ... p_t^{\gamma_t}$ Из свойства мультипликативности получаем

$$\varphi_\alpha(m) = m^n \prod_{i=1}^{t} \left(1 - \frac{1}{p_i^n}\right)$$

что и требовалось доказать. □

4. Функция Эйлера колец вычетов гауссовых чисел

Наша цель — получить формулу роста в среднем функции $\varphi_i(m)$. Для этого изучим функцию $\varphi_i(m)$ для всех натуральных m.

Сначала напомним следующее

Определение. *Нормой числа* $z := a + bi$ *по модулю* m будем называть число $N(z) := a^2 + b^2 \pmod m$.

Лемма. *Если* $a+bi$ *имеет норму, взаимно простую с* m, $c+di$ *и* $e+fi$ *— произвольные элементы, причем* $(a + bi)(c + di) = (a + bi)(e + fi)$, *то* $c + di = e + fi$.

Доказательство.

$$(a + bi)(c + di) = (a + bi)(e + fi) \Rightarrow \begin{cases} a(c - e) + b(d - f) = 0 \\ a(d - f) - b(c - e) = 0 \end{cases}$$

Здесь и далее во всем доказательстве все равенства по модулю m. Обозначим $c - e = s$ и $d - f = t$. Тогда

$$\begin{cases} as + bt = 0 \quad (1) \\ at - bs = 0 \quad (2) \end{cases} \Rightarrow \begin{cases} abs + b^2 t = 0 \\ a^2 t - abs = 0 \end{cases}$$

Складывая два последних равенства, получаем $t = 0$ (т. к. $a^2 + b^2$ взаимно просто с m). Подставляя $t = 0$ в (1) и (2), имеем

$$\begin{cases} as = 0 \\ bs = 0 \end{cases} \Rightarrow \begin{cases} a^2 s = 0 \\ b^2 s = 0 \end{cases} \Rightarrow s(a^2 + b^2) = 0 \Rightarrow s = 0.$$

То есть $s = 0$ и $t = 0$. Откуда следует, что $c = e$ и $d = f$. □

Теорема 4. *Элемент $\alpha \in \mathbb{Z}_m[i]$ обратим тогда и только тогда, когда $N(\alpha)$ и m взаимно просты.*

Доказательство. «И только тогда». Если у элемента α норма не взаимно проста с m, то из мультипликативности нормы для любого элемента β норма $N(\alpha\beta)$ тоже не взаимно проста с m. Но $N(1) = 1$ взаимно проста с m, значит, $\alpha\beta \neq 1$.

«Тогда». Пусть элемент α таков, что $N(\alpha)$ и m взаимно просты. Умножим каждый элемент $\mathbb{Z}_m[\alpha]$ на α . Из леммы следует, что в полученном наборе элементов любые два различны. Значит, в нем каждый элемент встречается ровно по одному разу, т. е. в нем есть и 1. Значит, α обратим. □

Оказывается, что формулы для значений функции Эйлера $\varphi_i(m)$ различаются в зависимости от того, какие именно простые делители входят в m. Для удобства обозначим через $\mathbb{P}_1$ (соответственно $\mathbb{P}_{-1}$) множество всех простых чисел, дающих остаток 1 (соответственно -1) при делении на 4.

Теорема 5. *1. Если $p \in \mathbb{P}_1$, то*

$$\varphi_i(p^\alpha) = p^{2\alpha}\left(1 - \frac{1}{p}\right)^2.$$

2. Если $p \in \mathbb{P}_{-1}$, то

$$\varphi_i(p^\alpha) = p^{2\alpha}\left(1 - \frac{1}{p^2}\right).$$

3. Имеет место равенство

$$\varphi_i(2^\alpha) = 2^{2\alpha-1}.$$

Доказательство. 1. Рассмотрим элемент $a+bi$. Из теоремы 4 следует, что он обратим тогда и только тогда, когда его норма не делится на p. Посчитаем, сколько существует значений b таких, что a^2+b^2 делится на p, при различных a.

1) Пусть a делится на p. В этом случае a^2+b^2 делится на p тогда и только тогда, когда b делится на p. Среди чисел от 0 до $p^\alpha-1$ есть $p^{\alpha-1}$ чисел, кратных p, значит, в случае 1) получаем $p^{2\alpha-2}$ необратимых элементов.

2) Пусть a не делится на p. Тогда необходимо, чтобы b^2 было сравнимо с $-a^2$ по модулю p. Остаток $-a^2$ является квадратичным вычетом по модулю p, так как p дает остаток 1 при делении на 4 (о квадратичных вычетах см. VIII.2). Отсюда следует, что уравнение $b^2 \equiv -a^2 \pmod{p}$ имеет ровно два решения среди чисел от 0 до $p-1$. Значит, среди чисел от 0 до $p^\alpha - 1$ оно имеет ровно $2p^{\alpha-1}$ решений. В то же время среди чисел от 0 до $p^\alpha - 1$ существует ровно $p^\alpha - p^{\alpha-1}$ значений a, не кратных p, поэтому случай 2) дает $2p^{\alpha-1}(p^\alpha - p^{\alpha-1})$ различных необратимых элементов.

Итого мы имеем

$$2p^{\alpha-1}(p^\alpha - p^{\alpha-1}) + p^{2\alpha-2} = 2p^{2\alpha-1} - p^{2\alpha-2}$$

необратимых элементов. Всего элементов $p^{2\alpha}$, значит, обратимых ровно

$$p^{2\alpha} - 2p^{2\alpha-1} + p^{2\alpha-2} = p^{2\alpha}\left(1 - \frac{1}{p}\right)^2.$$

2. Если p простое и дает остаток -1 при делении на 4, то -1 является квадратичным невычетом по модулю p, поэтому этот пункт следует из теорем 1,2,3.

3. Рассмотрим элемент $a+bi$. Среди чисел от 0 до $2^\alpha - 1$ четных и нечетных поровну. Поэтому одинакова вероятность четырех исходов:

1) a и b четны;

2) a и b нечетны;

3) a четно, b нечетно;

4) a нечетно, b четно.

Из теоремы 4 следует, что элемент $a + bi$ обратим тогда и только тогда, когда имеют место исходы 3) или 4). Значит, обратимых элементов ровно половина, т.е. $2^{2\alpha-1}$. □

Из доказательства теорем 1,2,3 сразу следует, что функция Эйлера φ_i мультипликативна. Таким образом, мы немедленно получаем

Следствие. *Имеет место формула*

$$\varphi_i(m) = \varepsilon(m) \cdot m^2 \cdot \prod_{p_k \in \mathbb{P}_{-1}} \left(1 - \frac{1}{p_k^2}\right) \cdot \prod_{q_l \in \mathbb{P}_1} \left(1 - \frac{1}{q_l}\right)^2.$$

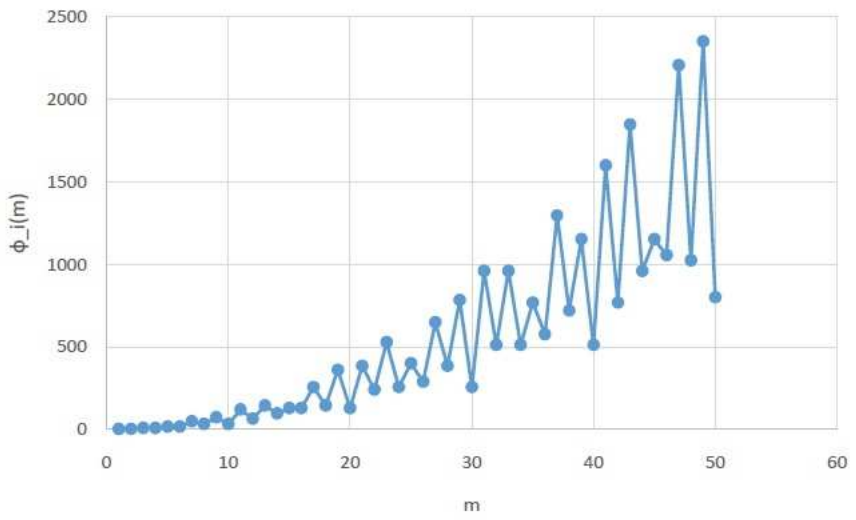
Здесь $\varepsilon(m) = 1$, *если* m *нечетно, и* $\varepsilon(m) = \frac{1}{2}$, *если* m *четно, а произведения ведутся по всем простым делителям числа* m.

В следующей таблице представлено несколько значений функции φ_i.

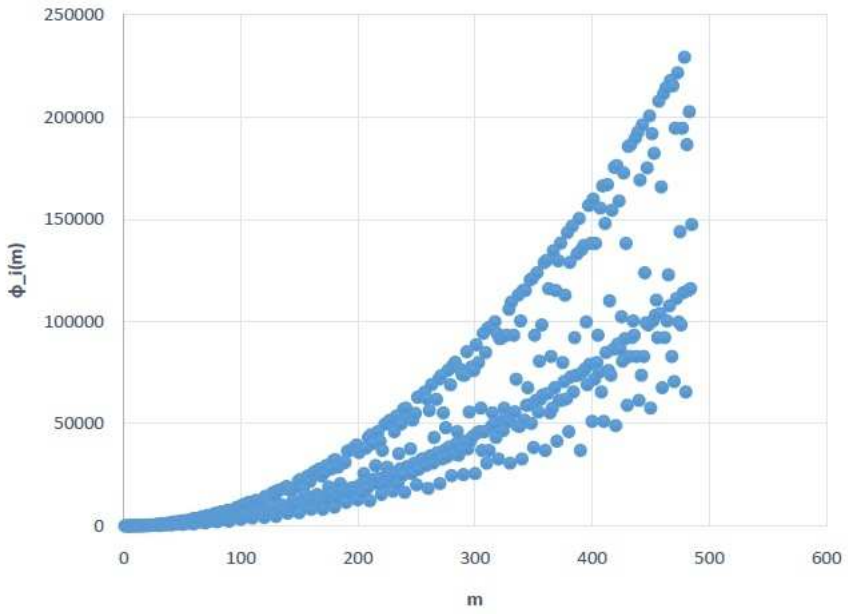
m	$\varphi_i(m)$	m	$\varphi_i(m)$
1	1	**13**	144
2	2	**17**	256
3	8	**19**	360
4	8	**25**	400
5	16	**33**	960
6	16	**45**	1152
7	48	**50**	800
8	32	**63**	3456
9	72	**77**	5760
10	32	**99**	8640

5. Рост в среднем функции Эйлера φ_i

В этом разделе мы изучим рост функции Эйлера φ_i. Ниже представлены графики этой функции



Значения $\varphi_i(m)$ на отрезке от 1 до 50.



Значения $\varphi_i(m)$ на отрезке от 1 до 500.

Отметим, что функция Эйлера растет крайне нерегулярно. Поэтому говорить о ее асимптотике нельзя. Однако можно «сгладить» скачки функции Эйлера, если рассмотреть ее средние арифметические. Это приводит нас к следующему понятию (см. [2] и также VI.4).

Определение. Будем говорить, что две функции f, g *одинаково растут в среднем*, если

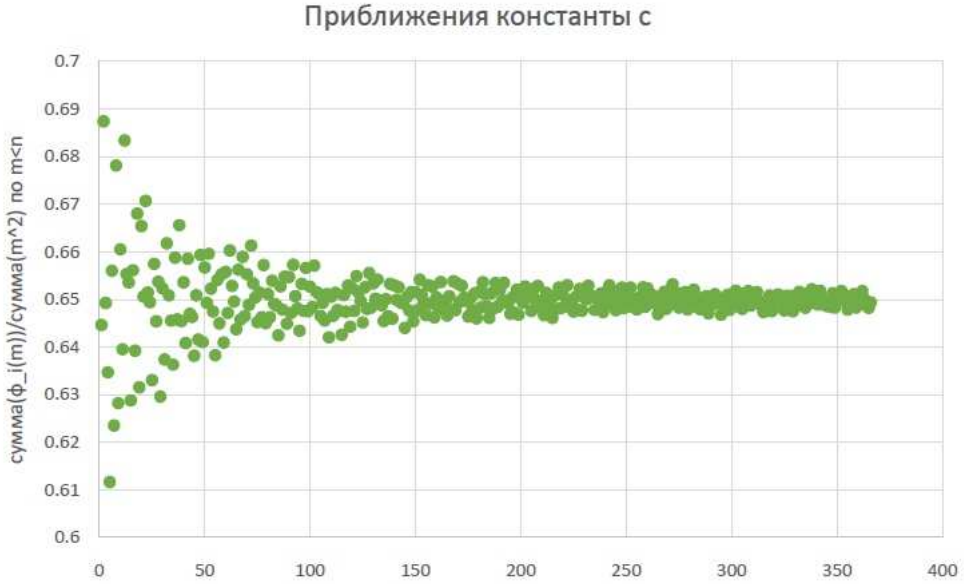
$$\frac{\sum\limits_{m\leqslant M} f(m)}{\sum\limits_{m\leqslant M} g(m)} \to 1 \quad \text{при } M \to \infty.$$

Обозначать это будем так $f \mathrel{\widehat{\sim}} g$.

Теперь сформулируем основную теорему этого раздела.

Теорема 6. *Имеет место следующая асимптотика в среднем:*

$$\varphi_i(m) \mathrel{\widehat{\sim}} cm^2, \quad \text{где } c = \frac{3}{4} \prod_{p\in\mathbb{P}_{-1}} \left(1 - \frac{1}{p^3}\right) \cdot \prod_{q\in\mathbb{P}_1} \left(1 - \frac{2}{q^2} + \frac{1}{q^3}\right) \approx 0,6498.$$



Доказательство. Назовем *правильными* трехмерные вектора вида (a, b, m), где $0 \leqslant a, b < m$. Сопоставим правильному вектору вида (a, b, m) элемент $a + bi$ гауссова кольца вычетов $\mathbb{Z}_m[i]$. Тогда каждому элементу каждого кольца $\mathbb{Z}_m[i]$ будет сопоставлен ровно один

правильный вектор. Будем называть правильный вектор (a, b, m) *необратимым по простому модулю* p, если $a^2 + b^2$ и m делятся на p. Будем называть правильный вектор (a, b, m) *необратимым*, если существует простое p, по модулю которого он необратим. Из теоремы 4 следует, что правильный вектор соответствует обратимому элементу тогда и только тогда, когда сам вектор обратим.

Лемма. *Если m делится на простое p, то число элементов с нормой, не кратной p, равно $p^{2\alpha}\sigma(p)$, где*

$$\sigma(p) = \begin{cases} \left(1 - \frac{1}{p}\right)^2, & \text{если } p \text{ дает остаток } -1 \text{ при делении на } 4; \\ 1 - \frac{1}{p^2}, & \text{если } p \text{ дает остаток } 1 \text{ при делении на } 4; \\ \frac{1}{2}, & \text{если } p = 2. \end{cases}$$

Доказательство. Из теоремы 5 следует, что если $m = p^\alpha$, то доля элементов с нормой, некратной p, равна $\sigma(p)$. Если же $m = p^\alpha t$, где t и p взаимно просты, то доля элементов с нормой, некратной p, будет такой же (это следует из китайской теоремы об остатках). Значит, число элементов с нормой, некратной p, равно $p^{2\alpha}\sigma(p)$, что и требовалось. □

Из леммы получаем, что если m делится на p, то существует ровно $p^2(1 - \sigma(p))$ правильных векторов, необратимых по модулю p и имеющих третью координату m. Если же m не делится на p, то правильных векторов, необратимых по модулю p и имеющих третью координату m не существует. Всего правильных векторов с третьей координатой m ровно m^2. Отсюда вероятность выбрать вектор, необратимый по модулю p, при случайном выборе правильного вектора, равна

$$\lim_{n\to\infty}\frac{p^2(1-\sigma(p))+(2p)^2(1-\sigma(p))+\ldots+(np)^2(1-\sigma(p))}{1^2+2^2+\ldots+(np)^2}=$$
$$=\lim_{n\to\infty}p^2(1-\sigma(p))\frac{1^2+2^2+\ldots+n^2}{1^2+2^2+\ldots+(np)^2}=$$
$$=\lim_{n\to\infty}p^2(1-\sigma(p))\frac{n^3/3}{n^3p^3/3}=\frac{1-\sigma(p)}{p}.$$

Значит, вероятность выбрать таким образом вектор, обратимый по модулю p, составляет

$$1-\frac{1-\sigma(p)}{p}.$$

Ясно, что эти события при разных p независимы, поэтому вероятность P выбрать обратимый вектор равна

$$\prod_{p\in\mathbb{P}}\left(1-\frac{1-\sigma(p)}{p}\right)=\frac{3}{4}\prod_{p\in\mathbb{P}_{-1}}\left(1-\frac{1}{p^3}\right)\cdot\prod_{q\in\mathbb{P}_1}\left(1-\frac{2}{q^2}+\frac{1}{q^3}\right).$$

Но при любом m среди m^2 правильных векторов вида (a,b,m) существует ровно $\varphi_i(m)$ обратимых. Поэтому

$$P=\lim_{n\to\infty}\frac{\varphi_i(1)+\varphi_i(2)+\ldots+\varphi_i(n)}{1^2+2^2+\ldots+n^2}$$

Значит,

$$\lim_{n\to\infty}\frac{\varphi_i(1)+\varphi_i(2)+\ldots+\varphi_i(n)}{1^2+2^2+\ldots+n^2}=\frac{3}{4}\prod_{p\in\mathbb{P}_{-1}}\left(1-\frac{1}{p^3}\right)\cdot\prod_{q\in\mathbb{P}_1}\left(1-\frac{2}{q^2}+\frac{1}{q^3}\right).$$

□

6. Продолженная функция Эйлера и ее свойства

В предыдущих разделах была рассмотрена функция Эйлера расширений колец вычетов от натуральных аргументов. В этой части мы рассмотрим продолжение функции Эйлера гауссовых кольец вычетов по модулям произвольных целых гауссовых чисел.

Пусть z — гауссово целое число. Все точки на комплексной плоскости, соответствующие числам, кратным z, образуют квадратную решетку (см. также доказательство евклидовости кольца гауссовых чисел в IX.2). Квадраты, из которых состоит решетка, будем называть ее элементарными ячейками. Условимся, что точка, лежащая линиях сетки, принадлежит той ячейке, в которую попадает ее образ при прибавлении к ней числа $\varepsilon z(i+1)$ при малых ε.

Определение. *Вычетами по модулю* z будем называть все гауссовы целые числа, лежащие в одной ячейке с числом 0. Через $\mathbb{Z}_z[i]$ обозначим кольцо вычетов по модулю z. *Функцией Эйлера* $\varphi_i(z)$ от произвольного гауссового целого числа z будем называть функцию, сопоставляющую z число обратимых элементов множества $\mathbb{Z}_z[i]$.

Отметим следующее известное

Утверждение. *Неприводимыми в* $\mathbb{Z}[i]$ *являются те и только те числа, которые удовлетворяют одному из следующих свойств:*

1) Норма числа равна 2.

2) Норма числа равна простому числу из $\mathbb{P}_1$.

3) Число является действительным простым числом из $\mathbb{P}_{-1}$.

Утверждение. *Во множестве* $\mathbb{Z}_z[i]$ *ровно* $N(z)$ *элементов.*

Доказательство. Элементарная ячейка решетки, соответствующая z, является квадратом с площадью $N(z)$. По формуле Пика $N(z) = a + b/2 - 1$, где строго внутри ячейки лежит ровно a целочисленных точек, а на ее границе лежит ровно b целочисленных точек. Но ячейке принадлежат все точки строго внутри нее, одна ее вершина и точки на двух ее сторонах, не являющиеся вершинами. Легко видеть, что это в точности $a+b/2-1$ точек. То есть число точек, принадлежащих ячейке, равно $N(z)$, что и требовалось. □

Теорема 7. *Пусть дано неприводимое число* $p \in \mathbb{Z}[i]$, *пусть гауссово целое* $z \in \mathbb{Z}[i]$ *кратно* p. *Тогда доля элементов* $\mathbb{Z}_z[i]$, *кратных* p, *равна* $\frac{1}{N(p)}$.

Доказательство. Элементарная ячейка решетки Z, соответствующей числу z, является квадратом с площадью $N(z)$. Элементарная ячейка решетки P, соответствующей числу p, является квадратом с площадью $N(p)$. Каждый узел Z является также узлом P, значит, при совмещении двух ячеек решетки Z при помощи параллельного переноса совмещаются все находящиеся в них узлы P. То есть внутри разных ячеек Z поровну узлов P.

Пусть в одной ячейке Z содержится t узлов P. Возьмем какую-нибудь ячейку Z и сопоставим каждому узлу P внутри этой ячейки ячейку решетки P, крайним узлом которой этот узел является. Суммарная площадь всех сопоставленных ячеек равна $tN(p)$.

Возьмем теперь множество S ячеек Z, образующих большой прямоугольник $s \times s$ ячеек и сделаем такое же сопоставление для каждой ячейки из S. Ясно, что отношение суммарной площади ячеек множества S к суммарной площади всех сопоставленных ячеек стремится к 1 при $s \to \infty$. Отношение этих площадей равно $\frac{s^2 N(z)}{s^2 t N(p)} = \frac{N(z)}{tN(p)}$, поэтому $t = \frac{N(z)}{N(p)}$.

Доля элементов $\mathbb{Z}_z[i]$, кратных p, равна отношению числа узлов P в ячейке Z к числу всех точек в ячейке Z, то есть $\frac{t}{N(z)} = \frac{1}{N(p)}$, что и требовалось. □

На функции, определенные на $\mathbb{Z}[i]$, естественным образом обобщается понятие роста в среднем. А именно, будем говорить, что функции f и g, определенные на $\mathbb{Z}[i]$, *одинаково растут в среднем*, если

$$\lim_{n\to\infty} \frac{\sum\limits_{N(z)<n} f(z)}{\sum\limits_{N(z)<n} g(z)} = 1.$$

Основная теорема данного раздела формулируется следующим образом.

Теорема 8. *Имеет место асимптотика в среднем*

$$\varphi_i(z) \widehat{\sim} CN(z), \quad \text{где} \quad C := \frac{3}{4} \prod_{p \in \mathbb{P}_{-1}} \left(1 - \frac{1}{p^4}\right) \cdot \prod_{q \in \mathbb{P}_1} \left(1 - \frac{1}{q^2}\right)^2 \approx 0,6637.$$

Доказательство. Будем рассматривать пары вида (z, r), где z является гауссовым целым числом, а r является элементом множества $\mathbb{Z}_z[i]$. Будем говорить, что пара (z, r) обратима, если r обратим в $\mathbb{Z}_z[i]$.

Пусть гауссово целое число p неприводимо. Будем говорить, что пара (z, r) необратима по модулю p, если z и r кратны p. Пара обратима тогда и только тогда, когда она обратима по модулю всех неприводимых p. Через $T(p)$ обозначим вероятность того, что случайно выбранная пара окажется необратима по модулю p.

Всего пар с первым элементом z ровно $N(z)$. Если z не кратно p, то неприводимых по модулю p пар с первым элементом z не существует. Если z кратно p, то по теореме 7 неприводимых по модулю p пар с первым элементом z будет $\frac{N(z)}{N(p)}$.

Доля гауссовых целых z, кратных p, среди всех гауссовых целых чисел равна $\frac{1}{N(p)}$. При этом рост в среднем величины $N(z)$ одинаков для z, кратных p, и для z, не кратных p. Отсюда получаем $T(p) = \frac{1}{N^2(n)}$.

Вероятность того, что случайно выбранная пара окажется обратима по модулю p, равна $1 - T(p) = 1 - \frac{1}{N^2(p)}$. События «выпадение пары, обратимой по модулю» при разных неприводимых p независимы. Значит, вероятность выбрать обратимую пару при случайной выборке равна

$$\prod_{p \in \mathbb{P}[i]} \left(1 - \frac{1}{N^2(p)}\right)$$

Из утверждения о неприводимых элементах в $\mathbb{Z}[i]$ следует, что это произведение равно

$$\frac{3}{4} \prod_{p \in \mathbb{P}_1} \left(1 - \frac{1}{p^2}\right)^2 \cdot \prod_{q \in \mathbb{P}_{-1}} \left(1 - \frac{1}{q^4}\right).$$

Но число обратимых пар с первым элементом z равно в точности $\varphi_i(z)$. Значит, вероятность выбрать обратимую пару при случайной выборке равна

$$\lim_{n\to\infty} \frac{\sum\limits_{N(z)<n} \varphi_i(z)}{\sum\limits_{N(z)<n} N(z)}.$$

Отсюда получаем, что

$$\lim_{n\to\infty} \frac{\sum\limits_{N(z)<n} \varphi_i(z)}{\sum\limits_{N(z)<n} N(z)} = \frac{3}{4} \prod_{p\in\mathbb{P}_1} \left(1-\frac{1}{p^2}\right)^2 \prod_{q\in\mathbb{P}_{-1}} \left(1-\frac{1}{q^4}\right)$$

□

Литература

[1] Арнольд В.И. *Группы Эйлера и арифметика геометрических прогрессий.* – М.: МЦНМО, 2003. – 44с.

[2] Байгушев Д.А. *О росте в среднем матричной функции Эйлера//* Труды Казанского математического общества им. Н.И. Лобачевского, – **28**, – 2013

[3] Винберг Э.Б. *Курс алгебры.* – М.: Изд-во "Факториал Пресс", 2001. – 544с.