С. Коутинхо

ВВЕДЕНИЕ В ТЕОРИЮ ЧИСЕЛ АЛГОРИТМ RSA

Перевод с английского С. А. Кулешова

Под редакцией С. К. Ландо

ПОСТМАРКЕТ

МОСКВА

2001

Криптография! Многие еще с детства заинтригованы этим процессом. Кто не помнит «пляшущих человечков»Конан Дойля? Но реальная схема шифрования и проще, и сложнее, чем об этом написано в знаменитом рассказе классика. С одной системой шифрования и знакомит эта книга.

Увидев в названии математическую теорию, некоторые из вас сочтут книгу скучной и неинтересной. Ошибаетесь! Пособие написано живо, интересно и очень доступно. Для понимания сути достаточно знаний средней школы. Но несмотря на простой стиль изложения, все утверждения снабжены строгими доказательствами или ссылками на литературу.

Круг читателей очень широк: от школьников, интересующихся теорией чисел или шифрованием, до банковских и корпоративных программистов, желающих глубже вникнуть в основы своей деятельности.

Содержание

«... всякая опытность и наблюдение человека, страстно к чему-нибудь привязанного, могут быть полезны для людей, разделяющих его любовь к тому же предмету.»

С. Т. Аксаков
«Записки об ужении рыбы»

Предисловие

Эта книга в первую очередь — книга по теории чисел, по науке, которую древние называли «царицей математики». И действительно, вряд ли в какой-нибудь другой части математики встретится такое количество столь простых и изящных по формулировке, и столь трудных для решения задач, настолько оторванных от практических надобностей человека. Мало какая область математики может похвастаться столь древней и славной историей. Сегодняшняя же теория чисел, называемая зачастую математиками попросту «арифметикой», вдобавок, еще и очень сложна по своим методам, почерпнутым из почти всех прочих математических наук.

Предыдущий абзац относится ко всем без исключения книгам по теории чисел. В чем же особенность книги, предлагаемой сейчас читателю? Пожалуй, в том, что ничто из сказанного выше для нее неверно. Во-первых, во второй половине двадцатого века у теории чисел неожиданно появились приложения к вещам вполне практическим. Сегодня компьютер, телефон, кредитная карточка и многие другие устройства используют довольно тонкие результаты этой науки. Как пели

по радио в моем детстве: «Чтоб водить корабли или летчиком стать, надо прежде всего арифметику знать». Яркий пример подобного приложения — криптосистема с открытым ключем, подробно разбираемая в этой книге. Идея такой криптосистемы проста чрезвычайно: перемножить два больших целых числа очень легко, а вот найти сомножители, зная произведение — довольно трудно.

Вторая особенность этой книги — ее простота. Книга написана автором для студентов первого курса, что в переводе на русский язык означает, что она вполне доступна интересующемуся школьнику 9, 10, 11 класса. Никаких, или почти никаких, предварительных знаний ее чтение не потребует.

Прочтя ее, Вы узнаете многие интересные факты из теории чисел и из ее истории, поймете, на чем основаны ее приложения к криптографии. Если же Вы не сочтете за труд прорешать предлагаемые в ней задачи, то после этого Вы будете готовы читать и более сложные книги. А главное, Вы почувствуете красоту теории чисел, про которую Эсхил говорил: «наука чисел, из наук важнейшая».

М.А.Цфасман

Андрэа и Даниэлю

И желание мое — чтобы читатель заметил, что я сделал себе развлечение из писания; а чтобы чтение не было для него нудным и скучным занятием, я сдобрил повествование — не скабрезностями, но шутками невинными и беспечными; и если ты — человек строгий и нрава мрачного, то судить о них не смей, ибо, как говорят мудрые, бывают оскорбления нанесенные и такие, которые, не будучи нанесенными, принимаются на свой счет.

Айзек Уолтон[1]*, «Совершенный рыболов».*

[1]Izaak Walton (1593-1683), английский писатель. Книга «The Compleat Angler» вышла в Лондоне в 1653 году.

Предисловие автора

Эта книга приглашает Вас в путешествие, конечная цель которого — (RSA), знаменитая система шифрования с открытым ключом Ривеста, Шамира и Адлемана (Rivest, Shamir, Adleman). Путешествие будет неспешным, с большим количеством остановок, на которых можно полюбоваться окружающим пейзажем и познакомиться с историческими достопримечательностями.

На самом деле, книга посвящена скорее математическим вопросам, чем криптографии. Хотя мы и изучим подробно работу системы шифрования RSA, детали ее реализации останутся в стороне. Вместо этого мы сосредоточимся на возникающих в связи с ней математических проблемах — разложение числа на простые множители и определение, является ли число составным или простым. Эти вопросы принадлежат к числу старейших в области, известной под именем *теории чисел*, которая с античных времен служит источником множества интригующих задач. В этой области работали такие математики, как Эвклид (Euclid), Ферма (Fermat), Эйлер (Euler), Лагранж (Lagrange), Лежандр (Legendre), Гаусс (Gauss), Риман (Riemann), а также, во времена не столь отдаленные, А. Вейль (Weil), Делинь (Deligne) и Уайлс (Wiles).

Предлагаемый в этой книге подход к теории чисел отличается от классического подхода старых монографий в некоторых важных аспектах. Мы всюду подчеркиваем алгоритмиче-

скую сторону дела, не забывая строго обосновать все встречающиеся на нашем пути алгоритмы. Разумеется, со времен Эвклида теория чисел была пронизана алгоритмами, однако до самого последнего времени этот подход казался несколько старомодным. Мы относимся к нему очень серьезно. Так, доказательству Эвклида бесконечности множества простых чисел предшествует обсуждение формулы прайморіала, а существование примитивных корней по модулю простого числа доказывается с помощью алгоритма Гаусса, который тот изобрел для вычисления корней.

Таким образом, в действительности эта книга об алгоритмической теории чисел и ее применениях к системе шифрования RSA. Но хотя цель и обозначена очень четко, изложение ни в коем случае не зациклено на ней. На самом деле, мы не всегда следуем кратчайшим путем, предпочитая тот, который способен пролить больше света на интересующий нас вопрос. Этим объясняется обращение к понятию группы, с помощью которого различные методы разложения числа на простые множители получают в главах 10 и 11 единое толкование. Наше отклонение в теорию групп заводит аж до теоремы Лагранжа и содержит обсуждение групп симметрий.

В основу книги легли записки лекций, предназначенных для программистов-первокурсников. Некоторые ее особенности объясняются слабой подготовкой студентов. Так, она предполагает у читателя лишь незначительные предварительные математические познания. Практически не используется ничего, кроме формулы для суммы геометрической прогрессии и биномиального разложения. Кроме того, хотя предметом книги и являются алгоритмы, никакого предварительного знакомства с программированием не требуется. Можно ожидать, однако (благодаря самому выбору предмета), что процент компьютерно грамотных читателей окажется значительным. Поэтому в конце каждой главы приведены (необязательные) упражнения, иллюстрирующие описанные в тексте алгоритмы.

Цель многих из них состоит в получении числовых данных для проверки известных формул или гипотез в теории чисел. Их можно отнести к тому, что некоторые называют *компьютерным экспериментом*.

Скажем несколько слов о стиле. Иногда математические книги представляют собой сухую последовательность определений, теорем и доказательств. Такой стиль восходит к «Элементам» Эвклида, и в конце двадцатого века он приобрел силу стандарта. Не будем забывать, однако, что этот монументальный стиль не был доминирующим даже среди греческих математиков. Архимед (Archimedes), к примеру, рассказывал читателям о возникавших на его пути трудностях и тупиках, куда ему случалось забредать, и даже предупреждал их об утверждениях, которыми он пользовался и которые впоследствии оказывались ложными. В этой книге я следую скорее примеру Архимеда, нежели Эвклида, и сделанный мною выбор заметно влияет на способ изложения. Во-первых, исторические комментарии вживлены в текст, а не выделены в отдельные примечания, а их предметом может служить все, что угодно: от истоков теории групп до анекдотов. Во-вторых, алгоритмы записаны на обычном языке, и я не стремлюсь к их оптимизации — если только она не работает на понимание. Пять лет преподавания излагаемого материала убедили меня в том, что программирование предложенных алгоритмов не вызывает сколь-нибудь заметных трудностей у всякого, имеющего необходимую подготовку.

Следует отметить еще одну особенность книги: у каждой важной теоремы и у всякого важного алгоритма есть свое имя. В большинстве своем это классические имена, которые используются в этом качестве уже десяти- или столетиями. Другие изобретены мной. По некоторым из них, как, скажем, *теорема о примитивном корне*, формулировка результата немедленно восстанавливается всяким, знакомым с предметом изложения; для других сделать это труднее. Чтобы облегчить поиск, я

выделил названия *основных* теорем и алгоритмов в отдельный указатель, снабдив их дополнительно кратким описанием[2]. Ссылка на безымянные результаты идет по номерам глав и разделов, в которых они содержатся.

Этот труд представляет собой переработанный вариант книги, опубликованной впервые на португальском языке в 1997 году, и в ее основе лежат лекции, прочитанные студентам-первокурсникам факультета программирования Федерального университета Рио де Жанейро. Я обязан студентам, слушавшим этот курс в течение последних пяти лет, больше, чем я способен выразить словами. Их участие повлияло как на стиль изложения, так и на содержание книги, а их предложения и критика помогли мне исправить ошибки и упростить многие доказательства.

Я особенно благодарен Жонасу де Миранда Гомесу. Именно он первым высказал мысль об издании английского варианта и провел все необходимые переговоры. Без него книга попросту не зародилась бы. Я также чрезвычайно благодарен Амилькару Пачеко и Мартину Холланду за предложения и комментарии.

И, наконец, я выражаю благодарность всем работникам издательства А. К. Петерс, с которыми я сотрудничал при создании этой книги. Их поддержка и спокойствие, даже в моменты, когда у меня опускались руки, помогли мне довести работу до завершения.

Рио де Жанейро, 18 июля 1998 года

[2] В русском переводе основные теоремы и алгоритмы присутствуют в содержании или в предметном указателе. — *Прим. перев.*

Глава 2.

Фундаментальные алгоритмы

Два самых фундаментальных алгоритма — это алгоритм деления и алгоритм Эвклида. Оба они были известны математикам Древней Греции — они содержатся в «Началах» Эвклида, написанных около 300 г. до н.э. Алгоритм деления предназначен для вычисления неполного частного и остатка при делении двух целых чисел. Алгоритм Эвклида вычисляет наибольший общий делитель двух целых чисел. По мере чтения книги Вы убедитесь в их фундаментальности.

§ 2.1. Алгоритмы

Оксфордский словарь английского языка дает следующее определение понятия «*алгоритм*»:

> «процесс или набор правил, обычно выраженный посредством алгебраических обозначений; в настоящее время используется, как правило, в программировании, машинном переводе и лингвистике.»

Если не отходить слишком далеко в сторону, то можно сказать, что алгоритм представляет собой *рецепт* решения задач определенного вида.

Для начала проанализируем детально какой-нибудь простейший рецепт. Пусть мы печем пирог. В хорошей кулинарной книге за названием рецепта следует список используемых продуктов. Затем следуют инструкции, объясняющие, что следует сделать с продуктами, чтобы получился пирог. Инструкции могут иметь вид «просеять, смешать, взбить, выпечь». В конце концов получается готовый результат — пирог, пригодный к употреблению.

Похожим образом устроен и любой другой алгоритм. При описании алгоритма мы должны определить его «*ввод*»и «*вывод*». Ввод соответствует набору продуктов, используемых в рецепте, вывод — результату, который мы хотим получить; в приведенном выше примере таким результатом был пирог. Собственно алгоритм — это набор действий, которые нужно совершить над вводом, чтобы получить вывод.

Предположим, что мы следовали рецепту с должной аккуратностью. Тогда, разумеется, мы ожидаем, что когда печь будет открыта, в ней будет находиться именно пирог, а не ростбиф или печенье. Кроме того, при выборе рецепта мы предполагаем, что пирог испечется за конечное время, желательно не слишком продолжительное. Точно так же мы ожидаем от любого алгоритма, что его результат будет совпадать с ожидаемым выводом. Хотелось бы, чтобы и время работы алгоритма было конечным, желательно не слишком большим. Разумеется, некоторые наборы инструкций могут выполняться бесконечно долго. Вот простой пример: прибавить 1 к целому числу (вводу), затем прибавить 1 к результату и т.д. Поскольку целых чисел бесконечно много, программа с такими инструкциями будет работать бесконечно долго. Разумеется, этот набор инструкций бесполезен.

С другой стороны, алгоритм может работать очень медленно, но при этом приносить весьма ощутимую пользу. Может быть так, что более быстрые алгоритмы неизвестны, или сами правила очень просты и с их помощью можно доказать,

что какая-то задача разрешима. Конечно, *не* всякую задачу можно решить, следуя определенному набору правил. Более удивительно то, что есть математические задачи, не допускающие алгоритмического решения. К несчастью, даже короткое обсуждение этой проблемы увело бы нас слишком далеко в сторону. Подробности можно посмотреть в книге [13].

Заканчивающийся алгоритм приводит к теореме: «при таком-то и таком-то вводе получается такой-то и такой-то вывод». Теоремы часто формулируют в виде «при данном предположении справедливо следующее заключение». Для теоремы, связанной с алгоритмом, ввод алгоритма соответствует предположению теоремы, а вывод — заключению.

Пусть Вас не пугает некоторая неопределенность этих комментариев: мы лишь договариваемся о терминологии. Все станет яснее, когда мы перейдем к приложениям. Итак, алгоритм — это рецепт, набор инструкций, для превращения набора продуктов (ввод) в некоторый результат (вывод). Пусть набор инструкций задан. Как проверить, решает ли он поставленную задачу? Предположим, нам сообщили, что представляют из себя ввод и вывод алгоритма. Теперь необходимо задать следующие вопросы:

- всегда ли при исполнении этих инструкций мы получаем какой-нибудь результат за конечное время?
- совпадает ли результат с ожидаемым?

Вспомнив про кулинарную метафору, мы должны согласиться с тем, что для рецепта пирога на эти вопросы ответить нельзя. Причина в том, что подобные утверждения нужно уметь доказывать до того, как сделано заключение, что данный набор инструкций представляет собой алгоритм. Ключевым словом последней фразы безусловно является «доказательство» — мы проговорились. Под доказательством мы понимаем логическое рассуждение, в основе которого лежат базисные факты, или аксиомы, о которых мы договорились

заранее. Для большинства алгоритмов в качестве аксиом выступают элементарные свойства целых чисел. Разумеется, нет оснований полагать, что правильность рецепта пирога можно в этом смысле доказать.

Происхождение слова *алгоритм* (или *алгорифм*) столь необычно, что ему стоит уделить внимание. Раньше его писали в виде *алгорисм*, и происходит оно из латинизированного арабского имени *Аль-Хорезми*, то есть «рожденный в Хорезме». Так звали арабского математика девятого века Абу Джафара Мохаммеда ибн Мусу (Abu J'afar Mohamed Ben Musa). Именно из его «Краткой книги об исчислении ал-джабра и алмукабалы» арабские числа распространились по всей Европе. *Алгорисм* означает попросту «число», что по-гречески называется «арифмос». Затем, как любезно сообщает нам оксфордский словарь, эти два слова переплелись, и появилось слово *алгорифм*.

Не вполне понятно, каким образом словом *алгоритм* начал называться «рецепт проведения вычислений», однако, похоже, это значение родилось совсем недавно. В английском оно впервые появилось около 1812 года. Однако уже в семнадцатом веке значение этого слова заметно расширилось. Мы видели, что исходное слово «алгоритм» означало число, однако затем так стали называться и вычисления.

Похоже, что первым за пределы арифметики вывел это слово математик и философ Г. В. Лейбниц (G. W. Leibniz). В своем первом докладе о дифференциальном исчислении, опубликованном в 1684 г., Лейбниц называет правила нового исчисления алгоритмами. Столетие спустя оно обрело свое современное значение. Гаусс многократно использует слово *алгоритм* в своих «Арифметических исследованиях», написанных по латыни, обозначая им набор формул, составляющих метод нахождения решений какой-либо арифметической задачи.

Ибн Мусе принадлежит еще по крайней мере один вклад в терминологию современной математики: слово *алгебра* обяза-

но своим происхождением названию его знаменитой книги, о которой говорилось выше.

§ 2.2. Алгоритм деления

Проанализируем алгоритм деления в соответствии со схемой, предложенной в предыдущем параграфе. Нас интересует деление целых чисел, поэтому задача состоит в том, чтобы найти неполное частное и остаток от деления двух положительных целых чисел. При словах «частное» и «остаток» большинству из нас приходит на ум картинка вроде следующей:

$$\begin{array}{r|l} 1234 & 54 \\ \cline{2-2} 108 & 22 \\ \cline{1-1} 154 & \\ 108 & \\ \cline{1-1} 46 & \end{array}$$

В этом примере мы делим 1234 на 54; неполное частное оказалось равным 22, а остаток равен 46. В терминах первого параграфа вводом алгоритма служат делимое и делитель; в приведенном примере они равны соответственно 1234 и 54. Вывод состоит из частного и остатка, значения которых в примере 22 и 46.

В общем случае *ввод* алгоритма деления состоит из двух положительных чисел a и b. Деля a на b, мы получаем при выводе числа q и r, которые связаны с a и b следующим образом:

$$a = bq + r \text{ и } 0 \leqslant r < b.$$

Разумеется, q — это неполное частное, а r — остаток от деления. У этого определения есть простая интерпретация, которую стоит иметь в виду. Допустим, мы хотим разломать полоску шоколада длины a на куски длины b. Алгоритм говорит, что в результате мы получим q кусков длины b и кусочек

меньшей длины r. Эту модель полезно помнить даже при применении теоремы в чисто математическом контексте.

На самом деле шоколадная полоска наводит на простейший алгоритм получения q и r по заданным a и b.

Алгоритм деления

Ввод: натуральные числа a и b.
Вывод: неотрицательные целые числа q и r, для которых выполнено равенство: $a = bq + r$ и $0 \leqslant r < b$.

Шаг 1. Положить $Q = 0$ и $R = a$.
Шаг 2. Если $R < b$, то сообщить: «*частное равно* Q*, а остаток равен* R», и остановиться; в противном случае перейти к шагу 3.
Шаг 3. Если $R \geqslant b$, то вычесть b из R, увеличить Q на 1 и возвратиться к шагу 2.

Такую форму записи алгоритмов мы будем использовать на протяжении всей книги. Для правильного прочтения алгоритмов нужно придерживаться следующих простых соглашений. Заметим, что алгоритм использует две *переменные* Q и R. Имена переменных выбраны такими потому, что по завершении работы алгоритма их значения будут равны неполному частному и остатку от деления[1] a на b. Для вычисления результата шаги 2 и 3 будут повторены несколько раз. Значит они образуют *цикл*. Значения переменных Q и R будут меняться от цикла к циклу. Именно поэтому они и называются переменными! Изменение значений переменных происходит на шаге 3. Инструкция «*вычесть* b *из* R» означает, что переменной R должно быть присвоено новое значение, равное ее значению после окончания предыдущего цикла, уменьшенному на b. Аналогично, инструкция «*увеличить* Q *на* 1» означает,

[1] Термины «частное» и «остаток» в английском языке выглядят как «quotient» и «remainder»; отсюда и обозначения. — *Прим. перев.*

что значение Q после окончания предыдущего цикла следует увеличить на 1.

Предположим, например, что $a > b$. Тогда после первого прохода через шаг 3 мы получим $Q = 1$ и $R = a - b$. Если $a - b \geqslant b$, то, согласно алгоритму, мы должны выполнить шаг 3 еще раз. Проделав это, мы получаем $Q = 2$ и $R = a - 2b$, и т.д. Почему такой процесс не может повторяться бесконечно? Другими словами, почему алгоритм прекращает свою работу? Заметим, что в результате последовательного применения шага 3 мы получаем следующую последовательность значений переменной R:

Начальное значение	1-й цикл	2-й цикл	3-й цикл	...
a	$a - b$	$a - 2b$	$a - 3b$	...

Это убывающая последовательность целых чисел. Поскольку количество чисел между a и 0 конечно, последовательность с необходимостью попадает в число, *меньшее* b. Тогда на шаге 2 работа останавливается, и алгоритм выводит значения переменных R и Q. Вот почему алгоритм всегда завершает работу.

§ 2.3. Теорема деления

В § 2.1 мы говорили о том, что всякому алгоритму соответствует теорема. Сформулируем теорему, отвечающую алгоритму деления.

Теорема деления. *Пусть a и b — натуральные числа. Тогда существует единственная пара неотрицательных целых чисел q и r таких, что*

$$a = bq + r \qquad и \qquad 0 \leqslant r < b.$$

Теорема содержит два утверждения про числа q и r. Во-первых, они существуют, во-вторых, они единственны. Мы

уже знаем, что для данных a и b существуют такие числа q и r, как указано выше. Мы даже знаем, как их вычислить. Однако утверждение о единственности является новым. Что означает единственность пары q и r? Предположим, что мы взяли два числа a и b и предложили их нескольким людям, попросив их найти такие q и r, что выполняются соотношения из теоремы. Обратите внимание, что мы просто просим вычислить эти числа, не накладывая ограничений на метод вычисления. Единственность неполного частного и остатка означает, что *все эти люди найдут одну и ту же пару чисел*. В частности, неважно, каким алгоритмом мы будем пользоваться для подсчета чисел q и r; любые алгоритмы дадут один и тот же результат. Знать это безусловно полезно.

Посмотрим, почему это правда. Пусть a и b — два натуральных числа, которые мы предложили разным людям, скажем, Карлу и Софии, и попросили их найти неполное частное и остаток, удовлетворяющие условиям, сформулированным в теореме. Результатом работы Карла являются числа q и r, а София нашла числа q' и r'. Нам известно только, что

$$a = bq + r \qquad \text{и} \qquad 0 \leqslant r < b;$$

$$a = bq' + r' \qquad \text{и} \qquad 0 \leqslant r' < b.$$

Следует ли отсюда, что $r = r'$ и $q = q'$? Поскольку числа r и r' целые, одно из них не меньше другого, скажем $r' \leqslant r$. Из тождества Карла мы заключаем, что $r = a - bq$, а из тождества Софии — что $r' = a - bq'$. Вычитая эти равенства, получаем

$$r - r' = (a - bq) - (a - bq') = b(q' - q).$$

С другой стороны, оба числа r и r' меньше b. По нашему предположению, $r \geqslant r'$, откуда $0 \leqslant r - r' < b$. Однако $r - r' = b(q' - q)$, поэтому

$$0 \leqslant b(q' - q) < b.$$

Число b положительно, так что на него можно разделить. Значит, $0 \leqslant q' - q < 1$. Но число $q' - q$ целое, поэтому последнее

неравенство выполняется в том и только в том случае, если $q' - q = 0$. Другими словами, $q = q'$, откуда $r = r'$ и единственность неполного частного и остатка доказана.

Подводя итог, мы доказали, что алгоритм деления приводит к теореме, состоящей из *двух* утверждений: неполное частное и остаток от деления двух натуральных чисел всегда *существуют* и они *единственны*. Многие из теорем, которые еще будут обсуждаться в нашей книге, также утверждают существование и единственность некоторых объектов. Наиболее важная из них — теорема о *разложении на простые множители* из главы 3.

§ 2.4. Алгоритм Эвклида

Алгоритм Эвклида предназначен для вычисления наибольшего общего делителя двух натуральных чисел, и мы посвятим начало этого параграфа подробному определению наибольшего общего делителя.

Во-первых, мы говорим, что целое число b *делит* целое число a, если существует еще одно целое число c такое, что $a = bc$. В этом случае мы говорим также, что b является *делителем*, или *множителем* числа a, а a, в свою очередь, — *кратным* числа b. Все это — различные способы сказать одно и то же. Разумеется, определить, является ли b делителем числа a, можно, подсчитав остаток от деления a на b и проверив, равен ли он нулю.

Пусть a и b — натуральные числа. *Наибольший общий делитель* чисел a и b — это наибольшее целое число d, на которое и a, и b делятся; тогда мы пишем $d = \text{НОД}(a, b)$. Если $\text{НОД}(a, b) = 1$, то мы называем числа a и b *взаимно простыми*.

Определение наибольшего общего делителя подсказывает следующий алгоритм его вычисления. Если числа a и b заданы, то найдем все положительные делители числа a и все положительные делители числа b. Выберем все числа, входящие

в оба множества, и возьмем наибольшее из них. Оно и будет наибольшим общим делителем. Эта процедура совсем проста, однако, как мы увидим в следующей главе, она чрезвычайно неэффективна при больших a и b. Проблема состоит в том, что неизвестно ни одного простого алгоритма разложения целых чисел на множители.

К счастью, наибольший общий делитель можно подсчитать и другим, весьма эффективным способом. Эвклид приводит его в предложениях 1 и 2 книги VII своих «Элементов». *Алгоритм Эвклида* действует следующим образом. Разделим a на b с остатком; назовем этот остаток r_1. Если $r_1 \neq 0$, то разделим b на r_1 с остатком; пусть r_2 — остаток второго деления. Аналогично, если $r_2 \neq 0$, то разделим r_1 на r_2 и получим новый остаток r_3. Таким образом, i-ый цикл алгоритма состоит из одного деления с остатком, причем делимое равно остатку, полученному в $(i-2)$-ом цикле, а делитель — остатку, полученному в $(i-1)$-ом цикле. Цикл повторяется до тех пор, пока мы не получим нулевого остатка; *наименьший ненулевой* остаток является наибольшим общим делителем чисел a и b.

Применим алгоритм Эвклида для вычисления наибольшего общего делителя чисел 1234 и 54. Деления с остатком выглядят так:

$$\begin{aligned} 1234 &= 54 \cdot 22 + 46; \\ 54 &= 46 \cdot 1 + 8; \\ 46 &= 8 \cdot 5 + 6; \\ 8 &= 6 \cdot 1 + 2; \\ 6 &= 2 \cdot 3 + 0. \end{aligned}$$

Последний ненулевой остаток равен 2, поэтому

$$\text{НОД}(1234, 54) = 2.$$

Отметим, что неполные частные не принимают непосредственного участия в подсчете наибольшего общего делителя. Опишем теперь алгоритм, следуя модели, предложенной в § 2.1 и § 2.2.

Алгоритм Эвклида

Ввод: натуральные числа a и b, $a \geqslant b$.
Вывод: наибольший общий делитель чисел a и b.

Шаг 1. Положить $A = a$ и $R = B = b$.
Шаг 2. Заменить значение R остатком от деления A на B и перейти к шагу 3.
Шаг 3. Если $R = 0$, то сообщить: «*наибольший общий делитель чисел a и b равен B*», и остановиться; в противном случае перейти к шагу 4.
Шаг 4. Заменить значение A на значение B, значение B на значение R и возвратиться к шагу 2.

Таким образом, для вычисления наибольшего общего делителя нам необходимо лишь выполнить несколько делений с остатком. Но почему наибольший общий делитель совпадает с последним ненулевым остатком в последовательности делений? Да и вообще, почему в последовательности остатков всегда появится нуль? Заметим, что если бы нуль не появлялся, то процедура никогда бы не остановилась.

Начнем со второго вопроса. Тем самым мы докажем, что алгоритм всегда завершает работу. Предположим, что для того, чтобы найти наибольший общий делитель чисел a и b, мы проделали следующие деления с остатком:

$$\begin{aligned}
a &= bq_1 + r_1 \quad \text{и} \quad 0 \leqslant r_1 < b;\\
b &= r_1q_2 + r_2 \quad \text{и} \quad 0 \leqslant r_2 < r_1;\\
r_1 &= r_2q_3 + r_3 \quad \text{и} \quad 0 \leqslant r_3 < r_2;\\
r_2 &= r_3q_4 + r_4 \quad \text{и} \quad 0 \leqslant r_4 < r_3;\\
&\ldots\ldots\ldots\ldots \quad \ldots \quad \ldots\ldots\ldots\ldots
\end{aligned}$$

Забудем на минуту про левый столбец. В правом столбце стоит последовательность остатков. Заметим, что в ней *всякий остаток меньше предыдущего*, а также, что *все остатки неотрицательны*. Переписав неравенства друг за другом, мы по-

лучаем цепочку

$$b > r_1 > r_2 > r_3 \cdots \geqslant 0. \tag{4.1}$$

Поскольку между b и нулем есть лишь конечное число целых чисел, последовательность остатков не может продолжаться бесконечно. Однако в конце ее может стоять только нуль, а значит, алгоритм наверняка остановится.

С помощью рассуждения из предыдущего параграфа можно получить верхнюю оценку на число делений, необходимое для вычисления наибольшего общего делителя. Вернемся к неравенствам (4.1). Каждое число в последовательности строго меньше предыдущего. Поэтому наибольшее возможное значение остатка в каждом делении на единицу меньше значения остатка на предыдущем делении. Если бы в каждом цикле это наибольшее возможное значение достигалось, то для получения нулевого остатка нам потребовалось бы b делений. Ясно, что это и есть наихудший возможный случай. Поэтому при применении алгоритма Эвклида к паре чисел $a \geqslant b$ число делений не превосходит b.

На самом деле, несложно показать, что при $b > 3$ число делений всегда меньше b. Зафиксируем число n. Тогда задачу лучше переформулировать так: для каких наименьших *взаимно простых* a и b вычисление НОД(a, b) требует n делений? Заметим, что для того, чтобы числа a и b были минимально возможными, частные на каждом шаге тоже должны быть минимально возможными. Если теперь предположить, что делитель меньше делимого, то ясно, что наименьшее возможное частное двух целых чисел равно 1. Предположим, что мы выполнили n делений до получения нулевого остатка. Тогда последовательность остатков имеет вид

$$b > r_1 > r_2 > r_3 \cdots \geqslant 0.$$

Мы уже видели, однако, что в наихудшем возможном случае все неполные частные равны 1. Запишем теперь все деления,

начиная с *последнего*. В силу взаимной простоты чисел мы получаем

$$\begin{aligned} r_{n-1} &= 1; \\ r_{n-3} &= r_{n-2} \cdot 1 + 1; \\ r_{n-4} &= r_{n-3} \cdot 1 + r_{n-2}; \\ \dots &\quad \dots \quad \dots \\ a &= b \cdot 1 + r_1. \end{aligned}$$

Вот как выглядит последовательность остатков при $n = 10$:

$$34,\ 21,\ 13,\ 8,\ 5,\ 3,\ 2,\ 1,\ 1,\ 0.$$

Значит, наименьшая пара взаимно простых чисел a и b, для подсчета наибольшего общего делителя которых необходимо 10 делений с остатком, это $a = 34$ и $b = 21$. Заметьте, что хотя число $b = 21$ и наименьшее, все равно оно больше, чем $n = 10$. Приведенная выше последовательность — это начало знаменитой *последовательности Фибоначчи*. Вы снова встретитесь с ней в упражнении 6.

§ 2.5. Доказательство корректности алгоритма Эвклида

Мы показали, что алгоритм обязательно остановится. Действительно, он не может выполнить больше делений с остатком, чем меньшее из двух введенных чисел. Но почему последний ненулевой остаток в точности равен наибольшему общему делителю? Чтобы это понять, нам понадобится один вспомогательный результат из тех, что математики называют *леммами*. Это слово древнегреческого происхождения, и означает оно то, что «предполагается» в доказательстве теоремы.

Лемма. *Пусть a и b — натуральные числа. Предположим, что существуют такие целые числа g и s, при которых $a = bg + s$. Тогда* НОД$(a, b) =$ НОД(b, s).

Мы должны показать справедливость утверждения леммы. Воспользуемся, однако, сначала этой леммой и покажем, что последний ненулевой остаток в алгоритме Эвклида действительно равен наибольшему общему делителю. Применяя алгоритм к целым числам $a \geqslant b > 0$ и предполагая, что остаток после n-го деления равен нулю, имеем

$$\begin{aligned}
a &= bq_1 + r_1 \quad \text{и} \quad 0 \leqslant r_1 < b;\\
b &= r_1q_2 + r_2 \quad \text{и} \quad 0 \leqslant r_2 < r_1;\\
r_1 &= r_2q_3 + r_3 \quad \text{и} \quad 0 \leqslant r_3 < r_2;\\
r_2 &= r_3q_4 + r_4 \quad \text{и} \quad 0 \leqslant r_4 < r_3;\\
&\ldots\ldots\ldots\ldots \quad \ldots \quad \ldots\ldots\ldots\ldots\\
r_{n-4} &= r_{n-3}q_{n-2} + r_{n-2} \quad \text{и} \quad 0 \leqslant r_{n-2} < r_{n-3};\\
r_{n-3} &= r_{n-2}q_{n-1} + r_{n-1} \quad \text{и} \quad 0 \leqslant r_{n-1} < r_{n-2};\\
r_{n-2} &= r_{n-1}q_n \quad \text{и} \quad r_n = 0.
\end{aligned} \tag{5.1}$$

На этот раз мы будем смотреть только на то, что происходит в левом столбце. Последнее равенство означает, что r_{n-2} делится на r_{n-1}. Поэтому наибольший общий делитель этих двух чисел равен r_{n-1}. Другими словами, НОД$(r_{n-2}, r_{n-1}) = r_{n-1}$.

Теперь в действие вступает лемма. Применяя ее к предпоследнему равенству, мы заключаем, что

$$\text{НОД}(r_{n-3}, r_{n-2}) = \text{НОД}(r_{n-2}, r_{n-1}),$$

причем последняя величина, как мы видели, равна r_{n-1}. Повторное применение леммы, на этот раз к предыдущему равенству, дает

$$\text{НОД}(r_{n-4}, r_{n-3}) = \text{НОД}(r_{n-3}, r_{n-2}),$$

что опять равно r_{n-1}. Продолжая действовать таким же обра-

зом до вершины столбца, мы заключаем, что НОД$(a, b) = r_{n-1}$ что и требовалось доказать.

Доказательство корректности алгоритма будет завершено, если мы докажем лемму. Напомним, она утверждает, что если четыре неотрицательных целых числа a, b, g и s связаны соотношением $a = bg+s$, то НОД$(a, b) =$ НОД(b, s). Доказательство легче объяснить, если положить

$$d_1 = \text{НОД}(a, b) \qquad \text{и} \qquad d_2 = \text{НОД}(b, s).$$

Пока мы ничего не сделали, просто присвоили имена наибольшим общим делителям чисел a и b и чисел b и s. Мы хотим доказать, что $d_1 = d_2$. Доказательство проведем в два этапа. Сначала мы покажем, что $d_1 \leqslant d_2$, а затем — что $d_2 \leqslant d_1$. Равенство чисел d_1 и d_2 немедленно следует из этих двух неравенств.

Покажем, что $d_1 \leqslant d_2$; второе неравенство доказывается аналогично, и мы оставляем его в качестве упражнения. Напомним, что $d_1 =$ НОД(a, b). Тогда d_1 делит как a, так и b. Это означает, что существуют такие натуральные числа u и v, что

$$a = d_1 u \qquad \text{и} \qquad b = d_1 v.$$

Подставляя эти значения в равенство $a = bg + s$, получаем $d_1 u = d_1 vg + s$. Другими словами,

$$s = d_1 u - d_1 vg = d_1(u - vg).$$

Но последнее равенство означает, что s делится на d_1.

Подведем итог. По предположению, $d_1 =$ НОД(a, b), поэтому d_1 делит b. Но проведенные вычисления показывают, что d_1 также делит число s. Поэтому d_1 является общим делителем чисел b и s. Однако d_2 — наибольший из таких общих делителей, поэтому $d_1 \leqslant d_2$, что мы и хотели доказать.

Отметим, что в доказательстве существенно используется соотношение $a = bg + s$, аналогичное соотношению в теореме

деления. Однако здесь нам нет необходимости предполагать, что s меньше b; на самом деле s не должно быть и положительным. Значит, предположение о том, что остаток от деления меньше делителя, используется не в доказательстве того, что последний ненулевой остаток является наибольшим общим делителем, а только для того, чтобы доказать, что алгоритм завершает работу.

§ 2.6. Расширенный алгоритм Эвклида

У алгоритма Эвклида, описанного в предыдущем параграфе, есть еще один вариант, более мощный. Мощный в данном случае не означает более быстрый. Достоинство нового варианта в том, что наибольший общий делитель — лишь часть выходных данных. Пусть a и b --- натуральные числа, а d --- их наибольший общий делитель. *Расширенный алгоритм Эвклида* подсчитывает не только d, но и два целых числа α и β таких, что

$$\alpha \cdot a + \beta \cdot b = d. \tag{6.1}$$

Отметим, что (за исключением нескольких тривиальных случаев) если α оказывается положительным, то β — отрицательное, и наоборот.

Лучше всего вычислять эти числа, добавив некоторые инструкции в обычный алгоритм Эвклида так, чтобы d, α и β подсчитывались одновременно. Именно поэтому новая процедура называется *расширенным алгоритмом Эвклида*. Приводимый здесь вариант этого алгоритма принадлежит Кнуту, автору знаменитой книги «Искусство программирования». Алгоритм описан во втором томе этого сочинения, см. [28] ([Д.4]).

Напомним, что алгоритм Эвклида состоит из последовательности делений с остатком. Наибольший общий множитель представляет собой последний ненулевой остаток в этой последовательности. Значит, нам надо найти способ записывать последний ненулевой остаток в виде (6.1).

Суть алгоритма Кнута в том, что нам не следует ждать, пока мы дойдем до последнего ненулевого остатка; вместо этого нам стоит записывать в таком виде каждый из получающихся остатков. Такие действия должны приводить, по-видимому, к большому количеству дополнительной работы. Как мы увидим позднее, это не совсем так.

Предположим, что для вычисления наибольшего общего делителя чисел a и b мы выполнили последовательность делений (5.1). Перепишем ее, сопровождая каждую операцию записью предполагаемого представления остатка:

$$\begin{aligned}
a &= bq_1 + r_1 \quad \text{и} \quad r_1 = ax_1 + by_1;\\
b &= r_1q_2 + r_2 \quad \text{и} \quad r_2 = ax_2 + by_2;\\
r_1 &= r_2q_3 + r_3 \quad \text{и} \quad r_3 = ax_3 + by_3;\\
r_2 &= r_3q_4 + r_4 \quad \text{и} \quad r_4 = ax_4 + by_4;\\
&\ldots\ldots\ldots\ldots \quad \ldots \quad \ldots\ldots\ldots\ldots\\
r_{n-3} &= r_{n-2}q_{n-1} + r_{n-1} \quad \text{и} \quad r_{n-1} = ax_{n-1} + by_{n-1};\\
r_{n-2} &= r_{n-1}q_n \quad \text{и} \quad r_n = 0;
\end{aligned} \tag{6.2}$$

Числа $x_1, \ldots, x_{n-1}$ и $y_1, \ldots, y_{n-1}$ мы и хотим определить. Необходимую информацию удобно свести в таблицу:

остатки	частные	x	y
a	$*$	x_{-1}	y_{-1}
b	$*$	x_0	y_0
r_1	q_1	x_1	y_1
r_2	q_2	x_2	y_2
r_3	q_3	x_3	y_3
$\vdots$	$\vdots$	$\vdots$	$\vdots$
r_{n-2}	q_{n-2}	x_{n-2}	y_{n-2}
r_{n-1}	q_{n-1}	x_{n-1}	y_{n-1}

Отметим прежде всего, что таблица начинается с двух строчек, которым в ней не следовало бы быть. Действительно,

стоящие в первом столбце этих строк числа не являются остатками в каких-либо операциях деления. Мы даем этим строчкам номера -1 и 0, подчеркивая тем самым их «незаконность». Вскоре мы обоснуем их необходимость.

Что же мы хотим сделать? Мы хотим заполнить столбцы x и y. Предположим на минуту, что мы получили таблицу заполненной до некоторой, скажем до $(j-1)$-ой, строки. Для заполнения j-ой строки необходимо прежде всего разделить r_{j-2} на r_{j-1}. В результате получатся r_j и q_j — первые два элемента j-ой строки. Не будем забывать, что $r_{j-2} = r_{j-1}q_j + r_j$ и $0 \leqslant r_j < r_{j-1}$. Таким образом,

$$r_j = r_{j-2} - r_{j-1}q_j. \tag{6.3}$$

Из строчек $j-1$ и $j-2$ мы можем взять значения x_{j-2}, x_{j-1}, y_{j-2} и y_{j-1}. Можно записать

$$r_{j-2} = ax_{j-2} + by_{j-2} \qquad \text{и} \qquad r_{j-1} = ax_{j-1} + by_{j-1}.$$

Подставляя эти значения в (6.3), получаем

$$\begin{aligned} r_j &= (ax_{j-2} + by_{j-2}) - (ax_{j-1} + by_{j-1})q_j = \\ &= a(x_{j-2} - q_jx_{j-1}) + b(y_{j-2} - q_jy_{j-1}). \end{aligned}$$

Поэтому

$$x_j = x_{j-2} - q_jx_{j-1} \qquad \text{и} \qquad y_j = y_{j-2} - q_jy_{j-1}.$$

Заметим, что для вычисления x_j и y_j нам понадобились только частное q_j и данные из двух строк таблицы, непосредственно предшествующих j-ой. Вот почему алгоритм Кнута такой эффективный. Для заполнения очередной строки достаточно знать две строчки, непосредственно предшествующие ей; все остальные строчки хранить не обязательно.

Итак, мы получили рекуррентную процедуру. Все, что необходимо — это научиться ее запускать. Именно для этого мы

и добавили в таблицу две «незаконные» строки. Найти в них значения x и y очень просто. Придавая им тот же смысл, что и в остальных строках, мы должны получить

$$a = ax_{-1} + by_{-1} \qquad \text{и} \qquad b = ax_0 + by_0.$$

Поэтому можно просто положить

$$x_{-1} = 1, \quad y_{-1} = 0, \quad x_0 = 0 \quad \text{и} \quad y_0 = 1,$$

и процедуру можно запускать.

В результате цепочки делений с остатком мы получим равенство $\text{НОД}(a, b) = r_{n-1}$ и вычислим такие целые числа x_{n-1} и y_{n-1}, что

$$d = r_{n-1} = ax_{n-1} + by_{n-1}.$$

Значит $\alpha = x_{n-1}$ и $\beta = y_{n-1}$. Заметим, что, зная α и $d = r_{n-1}$, мы можем найти β по формуле

$$\beta = (d - a\alpha)/b.$$

Поэтому достаточно вычислять только первые три столбца таблицы.

Приведем численный пример. Если $a = 1234$ и $b = 54$, то (полная) таблица выглядит так:

остатки	частные	x	y
1234	*	1	0
54	*	0	1
46	22	$1 - 22 \cdot 0 = 1$	$0 - 22 \cdot 1 = -22$
8	1	$0 - 1 \cdot 1 = -1$	$1 - 1 \cdot (-22) = 23$
6	5	$1 - 5 \cdot (-1) = 6$	$-22 - 5 \cdot 23 = -137$
2	1	$-1 - 1 \cdot 6 = -7$	$23 - 1 \cdot (-137) = 160$
0	3	*	*

Поэтому $\alpha = -7$, $\beta = 160$ и

$$(-7) \cdot 1234 + 160 \cdot 54 = 2.$$

Пришло время разобраться, почему алгоритм дает правильный ответ и почему он завершает свою работу. Как и подразумевается названием, расширенный алгоритм Эвклида представляет собой просто алгоритм Эвклида из предыдущего параграфа, дополненный инструкциями для вычисления значений x и y. Поэтому он останавливается, и наибольший общий делитель составляет часть выходных данных. Кроме того, в каждой строке числа из столбцов x и y удовлетворяют равенству типа (6.1), в котором число d заменено остатком из соответствующей строки. В частности, равенство (6.1) выполняется, если в качестве α и β взять числа из столбцов x и y строки, отвечающей последнему ненулевому остатку. *Расширенный алгоритм Эвклида* приводит к следующей теореме.

Теорема. *Пусть d — наибольший общий делитель натуральных чисел a и b. Тогда существуют такие целые числа α и β, что*

$$\alpha \cdot a + \beta \cdot b = d.$$

Заметим, что пара чисел α, β, упомянутая в теореме, не единственная. На самом деле таких пар бесконечно много. Например, возьмем α и β, для которых $\alpha \cdot a + \beta \cdot b = d$, и рассмотрим какое-нибудь целое k. Тогда, как несложно проверить,

$$(\alpha + kb) \cdot a + (\beta - ka) \cdot b = d.$$

Приложив столько усилий к вычислению α и β, разумно задаться вопросом, для чего эти числа могут понадобиться. Проще всего добраться до ответа, продолжая читать эту книгу. Знание этих чисел необходимо для получения большого числа принципиально важных результатов, включая выбор ключей в системе шифрования RSA.

Упражнения

1. Для каждой приведенной ниже пары целых чисел a, b найдите наибольший общий делитель и такие числа α и β, что

НОД$(a, b) = \alpha \cdot a + \beta \cdot b$:

(1) 14 и 35;
(2) 252 и 180;
(3) 6643 и 2873;
(4) 272, 828, 282 и 3242 (обобщите предварительно все понятия на случай, когда чисел больше двух).

2. Пусть n — натуральное число, большее единицы. Покажите, что

(1) НОД$(n, 2n + 1) = 1$;
(2) НОД$(2n + 1, 3n + 1) = 1$;
(3) НОД$(n! + 1, (n + 1)! + 1) = 1$.

3. Покажите, что для целых чисел a, b и $n > 0$ выполняется равенство

$$b^n - a^n = (b - a)(b^{n-1} + b^{n-2}a + b^{n-3}a^2 + \cdots + ba^{n-2} + a^{n-1}).$$

4. Пусть $n > m$ — натуральные числа, а r — остаток от деления n на m.

(1) Покажите, что остаток от деления $2^n - 1$ на $2^m - 1$ равен $2^r - 1$.
(2) Покажите, что если число r четное, то остаток от деления $2^n + 1$ на $2^m + 1$ равен $2^r + 1$.

Подсказка: для вычисления частного воспользуйтесь упражнением 3; утверждение вытекает из единственности остатка.

5. Пусть $n > m$ — натуральные числа. С помощью упражнения 4 вычислите НОД$(2^{2^n} + 1, 2^{2^m} + 1)$. Этот результат будет использован в упражнении 8 главы 4.

6. Каждое число, начиная с третьего, в последовательности Фибоначчи $1, 1, 2, 3, 5, 8, 13, \ldots$ является суммой двух предыдущих. Обозначая n-ое число последовательности через f_n,

мы получаем

$$f_0 = f_1 = 1 \qquad \text{и} \qquad f_n = f_{n-1} + f_{n-2}.$$

(1) Покажите, что наибольший общий делитель двух последовательных членов последовательности Фибоначчи равен 1.
(2) Сколько делений с остатком необходимо выполнить для подсчета НОД(f_n, f_{n-1})?

7. В этом упражнении мы описываем метод, с помощью которого можно найти решения уравнения $ax+by = c$ с целыми коэффициентами. Другими словами, мы хотим либо указать целые числа x и y, удовлетворяющие этому уравнению, либо показать, что таких чисел нет. Пусть $d = $ НОД(a, b). Тогда $a = da'$ и $b = db'$ для некоторых целых чисел a' и b'. Поэтому

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

Покажите, что если у этого уравнения есть целые решения, то c делится на d.

Если это так, то положим $c = dc'$ и рассмотрим *приведенное уравнение* $a'x + b'y = c'$. Покажите, что любое решение исходного уравнения является решением приведенного уравнения, и наоборот.

Поэтому для того, чтобы найти решения исходного уравнения, достаточно решить приведенное уравнение. Для этого воспользуемся расширенным алгоритмом Эвклида и вычислим целые числа α и β такие, что $\alpha \cdot a + \beta \cdot b = 1$. Покажите, что тогда числа $x = c'\alpha$ и $y = c'\beta$ дают решение приведенного уравнения.

8. Воспользовавшись упражнением 7, напишите программу для решения уравнения $ax+by = c$ в целых числах. Программа должна получать на входе значения a, b и c. На выходе появляется либо целочисленное решение уравнения, либо сообщение

о том, что решений нет. По существу, такая программа реализует расширенный алгоритм Эвклида.

9. Цель настоящего упражнения — выяснить экспериментально, какая часть случайно сгенерированных пар целых чисел состоит из взаимно простых чисел. Программа получает на входе натуральное число m, общее количество генерируемых пар. К каждой их этих пар применяется алгоритм Эвклида, который находит их наибольший общий делитель, а затем подсчитывается число пар, для которых он равен 1. Выходом служит величина

$$\frac{\text{число пар взаимно простых чисел}}{m}.$$

Эта дробь задает меру вероятности того, что случайно выбранная пара[1] состоит из взаимно простых чисел. Чтобы получить хорошее приближение для этой вероятности, программу нужно выполнять при больших значениях m. Прогоните ее десять раз при $m = 10^5$. Какие значения Вы получили? Теоретический анализ дает правильную вероятность $6/\pi^2$, см. [28] ([Д.4]). Насколько полученные Вами экспериментальные значения согласуются с этой величиной?

[1] Термин «случайно выбранная пара» требует точного определения, которого автор избегает. Строгую формулировку задачи можно найти в цитируемой книге Кнута. — *Прим. перев.*

Глава 3.

Разложение на множители

Стратегия «разделяй и властвуй» очень популярна в науке. Например, любое вещество разложимо на составляющие, на атомы. Более того, если свойства атомов хорошо известны, они многое говорят и о свойствах самого вещества.

Нечто похожее происходит и в целых числах. В этом случае роль атомов играют *простые числа*, а любое целое число раскладывается в произведение простых. Разложение служит главным инструментом в доказательстве многих свойств простых чисел. Найти разложение данного числа, однако, не всегда легко. Если число очень велико, то процедура его разложения может потребовать длительного времени, и она предъявляет большие требования к мощности компьютера.

§ 3.1. Теорема о разложении

Начнем со строгого определения главных героев. Целое число p называется *простым*, если $p \neq \pm 1$ и единственными его делителями являются числа ± 1 и $\pm p$. Так, числа $2, 3, 5$ и -7 простые, а число $45 = 5 \cdot 9$ — нет. Почти всюду в книге мы будем использовать определение в несколько более узком смысле, называя простым положительное простое число. Целое чи-

сло, отличное от ± 1 и не простое, называется *составным*, или *разложимым*. Для составного числа n существуют такие целые числа a и b, что $1 < a, b < n$ и $n = ab$. Значит, число 45 составное.

Заметим, что числа ± 1 не являются ни составными, ни простыми. Они относятся к третьей группе — это единственные целые числа, у которых есть целые обратные. В конце этого параграфа мы сможем более убедительно объяснить, почему их не следует считать простыми.

Теорема о разложении на множители. *Всякое целое число $n \geqslant 2$ единственным образом записывается в виде*

$$n = p_1^{e_1} \dots p_k^{e_k},$$

где $p_1, p_2, \dots, p_k$ — простые числа, $1 < p_1 < p_2 < p_3 < \dots < p_k$ и $e_1, \dots, e_k$ — натуральные числа.

Эта теорема настолько важна, что ее иногда называют *основной теоремой арифметики*. Впервые в таком виде она сформулирована Гауссом в § 16 его «Арифметических исследований», что не мешало, однако, его предшественникам неявно использовать ее. Как пишут Харди (Hardy) и Райт (Wright) в своей книге по теории чисел, «Гаусс первым превратил арифметику в систематическую науку», см. [23].

Показатели $e_1, \dots, e_k$ называются *кратностями* простых множителей в разложении числа n. Другими словами, кратностью множителя p_1 в разложении числа n называется *наибольшее* число e_1 такое, что n делится на $p_1^{e_1}$. Отметим также, что у n есть k *различных* простых делителей, однако *общее число* его простых делителей равно $e_1 + \dots + e_k$.

Теорема содержит два различных утверждения. Во-первых, всякое целое число можно представить в виде произведения степеней простых чисел. Во-вторых, простые сомножители и их степени в разложении определены однозначно. Значит, нам нужно доказать две вещи: разложение существует, и оно единственно. Мы докажем их по отдельности. Как мы увидим, су-

ществование разложения доказать несложно, а вот его единственность — гораздо более тонкий факт.

После того, как теорема о разложении сформулирована, нам легче объяснить, почему ± 1 не следует считать простыми. Если включить их в число простых, то разложение на простые сомножители потеряет свойство единственности. Действительно, если число 1 простое, то 2 и $1^2 \cdot 2$ — два различных разложения числа 2 на простые сомножители. Тот же самый трюк с привлечением различных степеней числа 1 (или -1) дает бесконечно много разложений для любого целого числа. С целью избежать этих псевдоразложений (множество которых бесконечно и бессмысленно), мы и исключаем ± 1 из определения простого числа.

§ 3.2. Существование разложения

В этом параграфе мы показываем, что всякое целое число $n \geqslant 2$ может быть записано в виде произведения простых. Для доказательства мы приводим алгоритм, получающий на входе целое число $n \geqslant 2$ и выдающий простые сомножители числа n и их кратности. В качестве предварительного шага построим алгоритм, выходом которого является какой-нибудь простой делитель числа n.

Для поиска простых делителей проще всего воспользоваться следующим алгоритмом. Попробуем разделить n на все целые числа от 2 до $n-1$ подряд. Если одно из них делит n, то число n составное, и мы нашли наименьший из его делителей. В противном случае число n простое. Кроме того, если n составное, то найденный нами делитель обязан быть простым.

Посмотрим, почему справедливо последнее утверждение. Пусть f — целое число, такое что $2 \leqslant f \leqslant n-1$. Предположим, что f — *наименьший* делитель числа n, и пусть $f' > 1$ — делитель числа f. По определению делимости, существуют такие

натуральные числа a и b, что

$$n = f \cdot a \text{ и } f = f' \cdot b.$$

Значит, $n = f' \cdot ab$ и f' также является делителем числа n. Поскольку f — наименьший делитель n, должно выполняться неравенство $f \leqslant f'$, но f' делит f, поэтому $f' \leqslant f$. Эти неравенства означают, что $f = f'$. Тем самым мы доказали, что если число $f' \neq 1$ делит f, то оно совпадает с f. Значит, f простое.

Прежде, чем перейти к подробному описанию алгоритма, следует отметить еще один момент. Мы уже видели, что алгоритм ведет поиск делителей только среди натуральных чисел. Как далеко ему следует забираться? Очевидно, что за $n - 1$ заходить не стоит, делитель числа не может превышать его самого. Однако можно сказать и кое-что посильнее. Действительно, можно не искать делители, превышающие $\sqrt{n}$. Последнее утверждение вновь вытекает из того, что алгоритм ищет *наименьший* делитель числа n, больший чем 1. Поэтому нам необходимо показать только, что *наименьший* делитель $f > 1$ числа n удовлетворяет неравенству $f \leqslant \sqrt{n}$.

Это утверждение легко проверить. Пусть $n = fa$. Поскольку $f > 1$ — наименьший делитель числа n, имеем $f \leqslant a$. Теперь $a = n/f$ и, следовательно, $f \leqslant n/f$, откуда вытекает, что $f^2 \leqslant n$. Другими словами, $f \leqslant \sqrt{n}$, что мы и хотели доказать.

Подвести итог обсуждению можно следующим образом. Алгоритм занимается проверкой (начиная с 2 и пробегая по натуральным числам, не превышающим $\sqrt{n}$) того, делится ли n на очередное число. Для составного числа n таким образом будет найден его наименьший делитель, больший или равный 2. Как мы уже доказали, этот делитель будет обязательно простым. Если в процессе проверки ни один делитель не будет обнаружен, то само n — простое.

Еще одно замечание практического характера. Обозначим через $[\alpha]$ целую часть вещественного числа α. Другими словами, $[\alpha]$ — наибольшее целое число, меньшее или равное α.

Так, $[\pi] = 3$ и $\left[\sqrt{2}\right] = 1$. Заметим, что для целого числа $r \leqslant \alpha$ выполняется неравенство $r \leqslant [\alpha]$. Значит, для реализации описанного выше алгоритма разложения нам достаточно знать только значение[1] $\left[\sqrt{n}\right]$. Процедура подсчета этого значения приведена в § П.1 приложения.

Ниже мы приводим запись алгоритма разложения в соответствии с канонической формой, предложенной в главе 2. Чтобы избежать излишней суеты, будем предполагать, что компьютер умеет вычислять значение $\left[\sqrt{n}\right]$.

Алгоритм разложения путем деления методом проб

Ввод: натуральное число n.
Вывод: натуральное число $f > 1$ — наименьший простой делитель числа n — или сообщение о том, что n простое.

Шаг 1. Положить $F = 2$.
Шаг 2. Если n/F целое, то сообщить: «*F является делителем числа n*», и завершить работу; в противном случае перейти к шагу 3.
Шаг 3. Увеличить F на единицу и перейти к шагу 4.
Шаг 4. Если $F \geqslant \left[\sqrt{n}\right]$, то сообщить: «*$n$ простое*», и завершить работу; в противном случае перейти к шагу 2.

Мы описали способ определить, является ли число $n > 2$ простым, и подсчитать его делитель, если n составное. Разумеется, если n простое, то мы находим и его разложение. Однако, если n составное, то нам хотелось бы найти все его делители и указать их кратности. Для этого достаточно применить описанный алгоритм несколько раз.

Предположим, что в результате применения предыдущего алгоритма к n мы нашли делитель q_1 этого числа. Тогда q_1 —

[1] Строго говоря, даже это необязательно. Достаточно проверять, не превысил ли квадрат очередного проверяемого числа значение n. — *Прим. перев.*

его наименьший простой множитель. Применим тот же алгоритм к числу n/q_1. Предположим, что число n/q_1 составное и q_2 — его наименьший простой делитель. Ясно, что $q_2 \geqslant q_1$. Заметим, однако, что эти делители могут оказаться равными. Такая возможность реализуется, если n делится на q_1^2. Продолжая в том же духе, мы применяем алгоритм к $n/(q_1q_2)$ и т.д. В результате мы получаем последовательность простых чисел

$$q_1 \leqslant q_2 \leqslant q_3 \leqslant \cdots \leqslant q_s,$$

каждое из которых является делителем числа n. Этой последовательности отвечает последовательность частных

$$\frac{n}{q_1} > \frac{n}{q_1q_2} > \frac{n}{q_1q_2q_3} > \cdots .$$

Заметим, что это *строго убывающая* последовательность натуральных чисел, каждое из которых соответствует применению алгоритма деления методом проб к n. Поскольку множество натуральных чисел, меньших n, конечно, после нескольких шагов мы получим полное разложение n на простые множители. Несложно проверить, что последнее число в последовательности частных равно 1, что и служит критерием завершения работы.

Предположим теперь, что мы хотим представить разложение в том виде, в котором оно записано в теореме о разложении. Все простые делители нам известны, осталось лишь подсчитать их кратности. Для этого нужно вычислить, сколько раз каждый простой сомножитель встречается в указанной последовательности простых чисел. Разумеется, подсчет проще всего вести одновременно с их вычислением.

Посмотрим на примере, как работает такой алгоритм. Допустим, мы хотим разложить число $n = 450$. Алгоритм деления методом проб дает первый наименьший простой множитель 2. Повторное применение алгоритма, на этот раз к частному $450/2 = 225$, дает множитель 3. Значит, делитель 3 числа

225 является и делителем числа 450. Применим алгоритм к числу $75 = 225/3$. Вновь наименьшим делителем оказывается 3. Значит, 450 делится на 3^2. Еще два выполнения алгоритма деления методом проб показывают, что 25 делится на 5^2, причем частное равно 1. Поэтому мы нашли полное разложение: $450 = 2 \cdot 3^2 \cdot 5^2$.

§ 3.3. Эффективность алгоритма деления методом проб

Описанный в предыдущем параграфе алгоритм легко понять и запрограммировать, однако он оказывается чрезвычайно неэффективным. Поскольку для поиска разложения мы многократно используем алгоритм деления методом проб, следует оценить эффективность последнего. Проиллюстрируем это на простом, но очень выразительном примере.

При применении алгоритма деления методом проб к натуральному числу $n > 2$, худшим случаем оказывается тот, когда n простое. В этом случае алгоритм выполняет до остановки $\left[\sqrt{n}\right]$ циклов. Для упрощения вычислений предположим, что n простое и в нем не меньше ста цифр. Сколько времени потребуется для подтверждения простоты числа n с помощью алгоритма деления методом проб?

Мы предполагаем, что $n > 10^{100}$, т.е. $\sqrt{n} > 10^{50}$. Значит мы должны повторить цикл по меньшей мере 10^{50} раз. Чтобы прикинуть, сколько это займет времени, допустим, что наш компьютер выполняет 10^{10} делений в секунду. Здесь мы предполагаем, что никаких других операций, кроме операции деления, в цикле нет. Безусловно, это далеко не так, однако сделаем такое допущение. Разделив первое число на второе, мы заключаем, что компьютеру потребуется 10^{40} секунд на проверку простоты числа n. Простой подсчет показывает, что на это уйдет приблизительно 10^{31} лет. По современным стан-

дартам это чересчур большой срок. Чтобы представить его себе, вспомним, что согласно наиболее распространенной точке зрения, Большой Взрыв произошел около $2 \cdot 10^{11}$ лет назад. Дополнительных комментариев не требуется: числа говорят сами за себя.

Означает ли это, что алгоритм деления методом проб бесполезен? Разумеется, нет. Допустим, что у раскладываемого числа есть маленький простой множитель, скажем, меньший, чем 10^6. Тогда алгоритм деления методом проб быстро его отыщет. С другой стороны, если у нас есть основания полагать, что тестируемое число простое, то алгоритм деления методом проб не выглядит наилучшим решением.

Есть много других алгоритмов разложения целых чисел, эффективность которых зависит от типа введенного числа. Так, алгоритм из § 3.2 очень хорош для чисел с маленькими простыми делителями. В следующем параграфе мы изучим алгоритм Ферма, который наиболее эффективен для таких чисел n, у которых есть делитель (не обязательно простой), не сильно превышающий $\sqrt{n}$.

Следует помнить, что эффективный алгоритм разложения на множители случайно выбранного натурального числа неизвестен. Неясно только, действительно ли он не существует или у человечества пока не хватило ума до него додуматься.

§ 3.4. Алгоритм Ферма разложения на множители

Алгоритм из § 3.2 эффективен только в случае, если у числа n, разложение которого мы ищем, есть маленький простой делитель. Насколько этот делитель должен быть мал, зависит от компьютера. В настоящем параграфе мы изучаем алгоритм, который эффективен, когда у n есть делитель (не обязательно простой), незначительно превосходящий $\sqrt{n}$. Идею алгоритма

придумал Ферма, и она требует гораздо большей изобретательности, чем алгоритм деления методом проб.

Предположим для начала, что n нечетное. Если бы оно было четным, то 2 было бы его делителем. Ключевая идея алгоритма состоит в том, чтобы попробовать представить n в виде $n = x^2 - y^2$, где x, y — неотрицательные целые числа. Если такие числа найдены, то

$$n = x^2 - y^2 = (x - y)(x + y).$$

Значит, $x - y$ и $x + y$ являются делителями числа n.

Чтобы отвлечься от посторонних деталей, будем предполагать, что компьютер умеет вычислять целую часть числа $\sqrt{n}$.

Проще всего применять алгоритм Ферма к полным квадратам. Если $n = r^2$ для некоторого целого числа r, то r является делителем n. Тогда $x = r$ и $y = 0$.

С другой стороны, если $y > 0$, то

$$x = \sqrt{n + y^2} > \sqrt{n}.$$

Алгоритм Ферма разложения на множители

Ввод: нечетное натуральное число n.
Вывод: множитель числа n или сообщение о том, что n простое.

Шаг 1. Положить $x = \lceil \sqrt{n} \rceil$. Если $n = x^2$, то x является делителем числа n, и работа алгоритма останавливается; в противном случае увеличить x на 1 и перейти к шагу 2.
Шаг 2. Если $x = (n + 1)/2$, то число n простое, и работа алгоритма останавливается; в противном случае вычислить $y = \sqrt{x^2 - n}$.
Шаг 3. Если число y целое (т.е., если $[y]^2 = x^2 - n$), то n раскладывается в произведение $(x + y)(x - y)$, и работа алгоритма останавливается; в противном случае увеличить x на 1 и перейти к шагу 2.

Как показывает следующий пример, этот алгоритм очень прост в применении. Допустим, мы хотим разложить на множители число $n = 1\,342\,127$. Сначала переменной x присваивается целая часть числа $\sqrt{n}$. В примере она равна $x = 1158$. Однако

$$x^2 = 1158^2 = 1\,340\,964 < 1\,342\,127.$$

Поэтому мы должны увеличить x на 1. Мы продолжим этот процесс до тех пор, пока число $\sqrt{x^2 - n}$ не станет целым или не будет выполняться равенство $x = (n + 1)/2$. Заметим, что в нашем примере $(n + 1)/2 = 671064$. Значения переменных x и y после завершения каждого цикла приведены в таблице.

x	$\sqrt{x^2 - n}$
1159	33,97...
1160	58,93...
1161	76,11...
1162	90,09...
1163	102,18...
1164	113

Таким образом, на шестом цикле мы получили целое число. Значит, искомые числа $x = 1164$ и $y = 113$. Соответствующие множители равны

$$x + y = 1277 \qquad \text{и} \qquad x - y = 1051.$$

§ 3.5. Доказательство корректности алгоритма Ферма

Теперь мы должны доказать, что алгоритм Ферма выполняет свою задачу, и что он всегда завершает работу. При доказательстве удобно разделить случай составного и случай простого числа n на входе. В первом случае мы должны показать,

что существует натуральное число x, такое что $\lceil\sqrt{n}\rceil \leqslant x < (n+1)/2$ и $\sqrt{x^2-n}$ целое число. Это означает, что если n составное, то алгоритм всегда находит делитель, меньший x, прежде, чем x становится равным $(n+1)/2$. А если n простое, то мы должны проверить, что число $\sqrt{x^2-n}$ не может оказаться целым при $x < (n+1)/2$.

Предположим, что n можно представить в виде произведения $n = ab$, где $a \leqslant b$. Мы хотим найти такие целые числа x и y, что $n = x^2 - y^2$. Другими словами,

$$n = ab = (x-y)(x+y) = x^2 - y^2.$$

Поскольку $x-y \leqslant x+y$, разумно попытаться положить $a = x-y$ и $b = x+y$. Решая эту систему уравнений относительно двух неизвестных, мы получаем

$$x = \frac{b+a}{2} \qquad \text{и} \qquad y = \frac{b-a}{2}.$$

И действительно, простое вычисление показывает, что

$$\left(\frac{b+a}{2}\right)^2 - \left(\frac{b-a}{2}\right)^2 = ab = n. \tag{5.1}$$

Отметим, что x и y должны быть целыми, а это значит, что оба числа $b+a$ и $b-a$ должны быть четными. Именно поэтому мы и требуем, чтобы n было нечетным; тогда каждое из чисел a и b, будучи делителем n, также нечетно, а значит и $b+a$, и $b-a$ четные. Если число n четное, то алгоритм может работать неправильно. Если, например, $n = 2k$ для нечетного k, то алгоритм никогда не завершит свою работу.

Если n простое, то единственные возможные значения a и b — это $a = 1$, $b = n$. Таким образом, $x = (n+1)/2$, и это наименьшее значение x, для которого число $\sqrt{x^2-n}$ целое. Рассмотрим теперь, что происходит, если n составное. Если $a = b$, то алгоритм находит делитель на шаге 1. Значит, мы можем предполагать, что n составное и не является полным

квадратом. Другими словами, $1 < a < b < n$. Мы утверждаем, что тогда алгоритм останавливается, так как

$$\left[\sqrt{n}\right] < \frac{a+b}{2} < \frac{n+1}{2}. \tag{5.2}$$

Начнем с доказательства неравенств.

Правое неравенство говорит, что $a+b < n+1$. Заменяя n на ab и вычитая из обеих частей $b+1$, мы приходим к неравенству $a-1 < ab-b$. Однако $a > 1$, поэтому обе части неравенства можно разделить на $a-1$. В результате мы получаем $1 < b$. Это рассуждение показывает, что неравенство $1 < b$ эквивалентно неравенству $a+b < n+1$. Поскольку $1 < a < b$ по предположению, мы показали, что $(a+b)/2 < (n+1)/2$.

Рассмотрим теперь левое неравенство. Отметим для начала, что поскольку $\left[\sqrt{n}\right] \leqslant \sqrt{n}$, нам достаточно доказать, что $\sqrt{n} \leqslant (a+b)/2$. Ясно, что последнее неравенство выполняется тогда и только тогда, когда $n \leqslant (a+b)^2/4$. Однако формула (5.1) дает

$$\frac{(a+b)^2}{4} - n = \frac{(b-a)^2}{4},$$

и правая часть неотрицательна. Тем самым мы доказали неравенство $(a+b)^2/4 - n \geqslant 0$, эквивалентное исходному.

Вернемся к алгоритму. Напомним, что значение переменной x вначале равно $\left[\sqrt{n}\right]$, а затем оно увеличивается на 1 при каждом выполнении цикла. Поэтому из неравенства (5.2) вытекает, что если число n составное, то алгоритм дойдет до $(a+b)/2$ раньше, чем до $(n+1)/2$. Однако при $x = (a+b)/2$ получаем

$$y^2 = \left(\frac{a+b}{2}\right)^2 - n = \left(\frac{b-a}{2}\right)^2.$$

Таким образом, достигнув этого значения, алгоритм завершит работу, а его вывод будет состоять из множителей a и b. Поэтому если n составное, то алгоритм всегда останавливается при некотором $x < (n+1)/2$, вычислив два делителя этого числа.

Заметим, что данное составное число n можно представить в виде $n = ab$, где $1 < a < b < n$, возможно, несколькими различными способами. Какое из разложений находит алгоритм Ферма? Поиск начинается при $x = \lceil \sqrt{n} \rceil$, и x увеличивается на каждом шаге. Значит, найденные алгоритмом делители a и b таковы, что разность

$$\frac{a+b}{2} - \lceil \sqrt{n} \rceil$$

— наименьшая из возможных.

Алгоритм Ферма сообщает нам кое-что важное о системе шифрования RSA. Напомним, что безопасность системы RSA зависит от сложности разложения на множители числа n, являющегося произведением двух простых. Если нам удастся разложить n, то шифр будет взломан. Алгоритм деления методом проб мог бы создать иллюзию, что выбрав большие простые сомножители, мы могли бы добиться того, что n сложно разложить. Однако это *не* так. Если множители большие, но их разность мала, то n очень легко разложить на множители алгоритмом Ферма. Мы вернемся к этому вопросу в главе 12.

§ 3.6. Одно фундаментальное свойство простых чисел

Для доказательства единственности разложения целого числа в произведение простых нам понадобится следующее фундаментальное свойство простых чисел. Мы доказываем это свойство в настоящем параграфе, а § 3.7 и § 3.8 посвящены некоторым его приложениям. Начнем с леммы, которая послужит первым применением расширенного алгоритма Эвклида.

Лемма. *Пусть a, b и c натуральные числа, причем a и b взаимно просты. Тогда:*

(1) *если произведение ac делится на b, то c делится на b;*

(2) *если c делится на a и на b, то c делится на ab.*

Докажем сначала утверждение (1). По предположению, числа a и b взаимно просты, т.е. НОД$(a, b) = 1$. Тогда результатом работы расширенного алгоритма Эвклида служат такие числа α и β, что

$$a\alpha + b\beta = 1.$$

Перейдем теперь к «абракадабре» доказательства. Умножив обе части последнего равенства на c, получим

$$\alpha ac + \beta bc = c. \tag{6.1}$$

Ясно, что второе слагаемое в левой части делится на b, но то же справедливо и для первого слагаемого. Действительно, оно делится на ac, а поэтому, по предположению, и на b. Таким образом, вся сумма в левой части делится на b, а поскольку она равна c, то первое утверждение доказано.

Выведем теперь (2) из утверждения (1). Раз c делится на a, то существует такое натуральное t, что $c = at$. Однако c делится и на b, и поэтому из (1) следует, что t делится на b, так как a и b взаимно просты. Значит, $t = bk$ для некоторого целого k. Поэтому число

$$c = at = a(bk) = (ab)k$$

делится на ab, что и утверждалось в п. (2).

Эта лемма будет использоваться очень часто, начиная с доказательства следующего свойства простых чисел, которое сформулировано в виде предложения 30 книги VII эвклидовых «Начал». Это свойство настолько важное, что мы дадим ему имя. Будем называть его *фундаментальным свойством простых чисел.*

Фундаментальное свойство простых чисел. *Если произведение натуральных чисел $a \cdot b$ делится на простое число p, то либо a делится на p, либо b делится на p.*

Фундаментальное свойство доказывается с помощью леммы. По предположению, ab делится на p. Если a делится на p,

то доказательство завершено. Предположим, что a не делится на p. Поскольку число p простое, это означает, что НОД$(a, p) = 1$. Тогда из первого утверждения леммы вытекает (ab делится на p, но a и p взаимно просты), что b делится на p.

§ 3.7. Греки и иррациональности

В этом параграфе мы рассматриваем одно из приложений фундаментального свойства простых чисел, доказанного в § 3.6. Мы хотим показать, что если p простое, то число $\sqrt{p}$ иррациональное. Доказательство будет первым в длинной цепочке *доказательств от противного*.

Идея этого метода доказательства очень проста, и мы часто пользуемся им в повседневной жизни. Вот достаточно безыскусный пример. Пусть Вам нужен файл, который находится на одной из двух дискет — синей или красной. К несчастью, Вы не помните, на какой именно, а меток на дискетах нет. Что Вы делаете? Вставляете одну из дискет, скажем синюю, в дисковод, и просматриваете ее оглавление. Если нужного файла там нет, то он находится на другой дискете. Если говорить более формально, то Вы предполагаете, что нужный файл находится на синей дискете. Обнаружив, что это не так, Вы заключаете, что предположение было неверным, и что файл, следовательно, находится на красной дискете.

Причина, по которой мы предполагаем, что такая стратегия будет работать, состоит в том, что никакое утверждение не может быть одновременно истинным и ложным. Так, если файл находится на одной из двух дискет, и если его нет на голубой дискете, то он должен быть на красной. Разумеется, в повседневной жизни редко встречаются столь простые ситуации. Например, Ваша уверенность в том, что файл находится на одной из двух дискет, может оказаться ошибочной. Или, хуже того, Вы могли походя стереть нужный файл. К счастью, в математике такой беспорядок редкость.

Посмотрим, как можно применить предложенную стратегию к доказательству того, что число $\sqrt{p}$ иррационально. Кстати, что означает в этом контексте слово «иррациональный»? Иногда можно услышать, что иррациональное — это что-то, чего нельзя понять. Однако здесь мы имеем в виду всего лишь отрицание *не рациональное*, т.е. «не являющееся отношением». Согласно Оксфордскому словарю английского языка, *отношение* представляет собой

> «численное соотношение между двумя аналогичными величинами, измеряемое количеством повторений одной из этих величин в другой.»

Это определение почти в точности повторяет определение Эвклида из книги V «Начал». К сожалению, такое определение не позволяет узнать, что такое отношение, если Вы еще не знаете, что это такое. В этом смысле оно похоже на знаменитое эвклидово определение точки: «нечто, что не имеет частей». К счастью, нам надо знать только, что иррациональное число это вещественное число, *не представимое в виде дроби*. Итак, перед нами вопрос, к которому метод *доказательства от противного* вполне приложим. Мы хотим проверить, что $\sqrt{p}$ *не* дробь? Давайте предположим, что это не так, и придем к противоречию. Если нам это удастся, то мы докажем иррациональность числа $\sqrt{p}$.

В проведении доказательства следует соблюдать осторожность. Напомним, что мы предполагаем (в надежде прийти к противоречию), что $\sqrt{p}$ является дробью. Другими словами, мы предполагаем, что существуют натуральные числа a и b такие, что

$$\sqrt{p} = \frac{a}{b}. \tag{7.1}$$

Более того, можно считать, что дробь записана в приведенном виде, т.е. НОД$(a, b) = 1$. В таком виде можно записать каждую дробь: для этого достаточно произвольную дробь сократить на наибольший общий делитель числителя и знамена-

теля. Предположение о взаимной простоте числителя и знаменателя существенно — оно облегчает поиск противоречия.

Чтобы иметь дело только с целыми числами, возведем обе части равенства (7.1) в квадрат. Получим

$$p = \frac{a^2}{b^2}, \text{ т.е. } b^2 \cdot p = a^2. \tag{7.2}$$

Значит, a^2 делится на p. Согласно фундаментальному свойству простых чисел, это означает, что a делится на p. Поэтому существует такое число c, что $a = pc$. Подставляя последнее выражение в (7.2), получаем

$$b^2 \cdot p = p^2 \cdot c^2.$$

Сокращая на p, мы заключаем, что b^2 должно делиться на p. Повторное использование фундаментального свойства простых чисел дает, что b делится на p. Значит, и a, и b делятся на p. Однако это невозможно, поскольку НОД$(a, b) = 1$. Тем самым, мы пришли к ожидаемому противоречию, и $\sqrt{p}$ не может быть дробью. Значит, число $\sqrt{p}$ иррационально.

У вопроса о существовании иррациональных чисел долгая и яркая история. Согласно греческому историку Геродоту (Herodotus), геометрия зародилась в Египте, где фараон раздавал подданным прямоугольные участки земли под годовую ренту. Если Нил смывал часть участка, то необходимо было вызывать землемера для определения, какая часть утеряна. Плата владельца участка сокращалась пропорционально потерянной площади.

Египтяне интересовались только практическими измерениями площади и другими подобными вычислениями, поэтому они неявно полагали все числа дробями. На передний план иррациональные числа выдвинулись в Древней Греции в результате развития более теоретического подхода к геометрии.

Считается, что иррациональные числа были открыты в философской школе (или секте), основанной Пифагором. Разви-

тие геометрии чрезвычайно интересовало пифагорейцев, поскольку они полагали, что числа (под которыми они подразумевали целые числа и дроби) лежат в основе мироздания. Можно представить себе, как ужаснулись они, поняв, что имеются отношения величин, не выражаемые никакой дробью. Говорят, что Хипас из Метапонтума (Hypasus of Metapontum) был изгнан из секты за обнародование этого секрета. Решив, что этого недостаточно, пифагорейцы даже воздвигли ему гробницу, чтобы продемонстрировать, что он для них умер!

Открытие иррациональных чисел с неизбежностью вскоре распространилось среди философов. Платон утверждает в диалоге «Театет», что Феодор Сиренский (Theodorus of Cirene) доказал иррациональность чисел $\sqrt{3}, \ldots, \sqrt{17}$. К сожалению, он ничего не говорит о методе доказательства.

Приведенное выше доказательство иррациональности числа $\sqrt{p}$ было известно грекам. В главе 23 книги I своей «Первой аналитики» Аристотель (Aristotle) пишет, что

> «диагональ квадрата несоизмерима с его стороной, так как если предполагать их соизмеримость, то нечетные числа равны четным.»

Это чрезвычайно сжатая форма доказательства иррациональности числа $\sqrt{2}$. Более развернутое доказательство содержится в предложении 117 книги X эвклидовых «Начал».

§ 3.8. Единственность разложения

Пришло время доказать, что представление натурального числа в виде, указанном в § 3.1, единственно. Мы докажем единственность от противного, воспользовавшись фундаментальным свойством простых чисел.

Предположим, напротив, что существуют натуральные числа, большие 2, допускающие больше одного разложения. Пусть

n — *наименьшее* натуральное число, у которого есть по крайней мере два различных разложения. Пусть

$$n = p_1^{e_1} \dots p_k^{e_k} = q_1^{r_1} \dots q_s^{r_s}, \tag{8.1}$$

где $p_1 < \dots < p_k$ и $q_1 < \dots < q_s$ — простые, а $e_1, \dots, e_k$, $r_1, \dots, r_s$ — натуральные числа. Кроме того, мы предполагаем, что эти два разложения различны. Заметим, что такое может произойти по двум причинам. Во-первых, в одном из разложений могут присутствовать простые множители, которых нет в другом. Во-вторых, даже если набор простых множителей в обоих случаях одинаков, их кратности могут быть различными. К счастью, неважно, какая именно из этих возможностей реализуется в разложении (8.1).

Исследуя левое разложение, мы заключаем, что n делится на p_1. Но $n = q_1^{r_1} \dots q_s^{r_s}$. Многократное применение *фундаментального свойства простых чисел* показывает, что один из сомножителей в $q_1^{r_1} \dots q_s^{r_s}$ должен делиться на p_1, а значит одно из чисел q_i должно делиться на p_1. Но простое число делится на другое простое, только если они равны. Значит $p_1 = q_j$ для некоторого j, $1 \leqslant j \leqslant s$.

Поэтому в правом разложении числа n можно заменить q_j на p_1:

$$\begin{aligned} n = p_1^{e_1} \dots p_k^{e_k} &= q_1^{r_1} \dots q_j^{r_j} \dots q_s^{r_s} \\ &= q_1^{r_1} \dots p_1^{r_j} \dots q_s^{r_s}. \end{aligned}$$

Теперь на p_1 можно сократить, поскольку оно входит в оба разложения в положительной степени. В результате получим

$$p_1^{e_1 - 1} \dots p_k^{e_k} = q_1^{r_1} \dots p_1^{r_j - 1} \dots q_s^{r_s},$$

т.е. два разложения на множители некоторого нового натурального числа, которое мы обозначим через m. Однако эти разложения не могут быть различными. Действительно, мы

выбрали в качестве n *наименьшее* положительное число с двумя различными разложениями, а $m = n/p_1 < n$. Раз полученные разложения совпадают, то $j = 1$, т.е. $p_1 = q_1$, а кроме того $k = s$. Далее,

$$p_2 = q_2, \quad p_3 = q_3, \quad , \ldots \text{ и } p_k = q_k,$$

и кратности каждого простого числа в обоих разложениях тоже равны,

$$e_1 - 1 = r_1 - 1, \quad e_2 = r_2, \quad , \ldots \text{ и } e_k = r_k.$$

Однако из этих равенств вытекает, что разложения в (8.1) также одинаковы, и мы пришли к противоречию. Значит, представление натурального числа в виде, указанном в теореме из § 3.1, единственно.

Преодолев все препятствия в доказательстве единственности разложения, мы должны признать, что большинство людей просто не могут себе представить возможности наличия нескольких разложений. Значит вновь математики доказывают нечто, что всем остальным совершенно очевидно.

Истина, однако, противоположна. Причина, по которой единственность разложения на простые множители для нас очевидна, состоит в том, что мы начинаем изучение целых чисел с изучения разложений, и в очень раннем возрасте. Поэтому вся выработанная нами интуиция основана на этом факте. Это все равно, что сказать, что эвклидова геометрия единственно правильная. Последнее утверждение заведомо неверно, и в наше время теории относительности и черных дыр никакому образованному человеку не придет на ум высказать что-либо подобное.

Если посмотреть на историю математики последнего столетия, то в ней найдется огромное количество примеров «числовых систем», элементы которых допускают разложение на неприводимые. Такое разложение, как правило, не единственно. Самый знаменитый контрпример связан с *великой теоре-*

мой Ферма. Так называют сформулированное Ферма утверждение о том, что если

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

для некоторой степени $n \geqslant 3$, то $xyz = 0$. На своем экземпляре диофантовой «Арифметики» Ферма пометил, что у него есть поистине замечательное доказательство этого факта, однако поля слишком узки, чтобы его привести.

Очевидная стратегия доказательства основана на том, чтобы полностью разложить разность $z^n - y^n$. Для этого необходимо ввести комплексные числа; тогда

$$x^n - y^n = (z - y)(z - \zeta y) \cdots (z - \zeta^{n-1} y),$$

где $\zeta = \cos(2\pi/n) + i \sin(2\pi/n)$. Оказывается, что необходимое нам множество комплексных чисел устроено очень похоже на целые числа. Всякий его элемент раскладывается в произведение неприводимых, т.е. таких, которые нельзя разложить на множители. Однако для большинства значений n разложение *не единственно*, что и служит основным препятствием к получению на этом пути простого доказательства.

Предполагается, что «доказательство» Ферма содержало аналогичную ошибку. Ферма мог попасться в ловушку, полагая, что разложение на множители в множестве комплексных чисел, с которым он работал, единственно. Было бы неудивительно, если бы Ферма пал жертвой такой ошибки. Как мы уже говорили в § 3.1, только Гаусс сформулировал утверждение о единственности разложения на множители в том явном виде, которым мы пользуемся и по сей день. Даже после выхода «Исследований» Гаусса, Куммер (Kummer) предложил доказательство теоремы Ферма, похожее на изложенное выше, не понимая, в чем заключается проблема, пока один из коллег не указал ему на ошибку. Не желая оставаться побежденным, Куммер разработал метод, позволяющий обходить неединственность разложения на множители. В результате он

смог доказать великую теорему Ферма для большого круга новых простых n.

Великая теорема Ферма была в конце концов доказана Уайлсом в 1995 г. Он следовал путем, предложенным лишь в последние 10 лет, предшествовавшие его работе. Этот путь основан на теории эллиптических кривых, в которых Уайлс является экспертом. Более раннюю историю теоремы Ферма можно изучить по книге [15] ([Д.12]). Хорошее элементарное введение в идеи, лежащие в основе доказательства Уайлса, можно найти в [20].

Упражнения

1. Существуют ли такие натуральные числа x, y и z, что $2^x \cdot 3^4 \cdot 26^y = 39^z$?

2. Пусть k — натуральное число, $k > 1$. Покажите, что все числа в последовательности

$$k! + 2, k! + 3, \ldots, k! + k$$

составные. Используйте этот факт для доказательства существования отрезков последовательных составных чисел произвольной длины.

3. С помощью алгоритма Ферма найдите делители следующих чисел: 175 557, 455 621 и 731 021.

4. Какие из приводимых ниже утверждений являются верными:

(1) Число $\sqrt{6}$ иррациональное.
(2) Сумма иррационального числа и дроби всегда иррациональна.
(3) Сумма двух иррациональных чисел всегда иррациональна.
(4) Число $\sqrt{2} + \sqrt{3}$ рациональное.

5. Покажите, что если число n составное, то число

$$R(n) = \frac{10^n - 1}{9} = \underbrace{111\ldots11}_{n \text{ раз}}$$

тоже составное. Такие числа называются *повторяющимися единицами*.

Подсказка: если n делится на k, то $R(n)$ делится на $R(k)$.

6. Пусть число $n > 0$ составное и p — его *наименьший* простой делитель. Найдите все возможные значения n, для которых

(1) $p \geqslant \sqrt{n}$;
(2) НОД$(6n + 7, 3n + 2)$ делится на $p - 4$.

7. *Наименьшее общее кратное* НОК(a, b) двух натуральных чисел a и b это наименьшее положительное число, которое делится как на a, так и на b. Положим

$$a = p_1^{e_1} p_2^{e_2} \cdots p_k^{e_k} \text{ и } b = p_1^{r_1} p_2^{r_2} \cdots p_k^{r_k},$$

где $p_1 < p_2 < \cdots < p_k$, а показатели $e_1, \ldots, e_k$ и $r_1, \ldots, r_k$ неотрицательны. Заметим, что мы *не* предполагаем, что в разложении обоих чисел на простые множители участвуют одинаковые простые числа; если, например, a делится на p_1, а b нет, то мы полагаем $r_1 = 0$. Покажите, что единственными простыми делителями чисел НОД(a, b) и НОК(a, b) являются $p_1, \ldots, p_k$ и найдите их кратности в соответствующих разложениях.

8. Натуральное число n называется *совершенным*, если сумма всех его делителей (включая 1 и само n) равна $2n$. Например, числа 6 и 28 совершенные. Пусть s — такое натуральное число, что $2^{s+1} - 1$ простое.

(1) Покажите, что делители числа $2^s(2^{s+1} - 1)$ образуют две геометрические прогрессии со знаменателем 2,

первая из которых начинается с 1, а вторая — со знаменателем $2^{s+1}-1$.

(2) Вычислите сумму этих делителей и покажите, что число $2^s(2^{s+1}-1)$ совершенное.

Выше сформулировано предложение 36 книги IX эвклидовых «Начал». Такие совершенные числа иногда называют *эвклидовыми*.

Цель следующих двух упражнений состоит в доказательстве того, что все *четные* совершенные числа эвклидовы, т.е. имеют вид $2^s(2^{s+1}-1)$, где $2^{s+1}-1$ простое число. Это утверждение было доказано Эйлером, однако сама статья была опубликована в 1849 году, через много лет после его смерти. Приводимое ниже доказательство взято из [14]. Интересно отметить, что все известные совершенные числа просты, а значит представимы в виде, известном уже Эвклиду. Было доказано, что если *нечетные* совершенные числа существуют, то каждое из них больше, чем 10^{300}, и имеет не менее восьми простых делителей.

9. Обозначим через $S(n)$ сумму всех делителей натурального числа n, включая 1 и n.

(1) Покажите, что число r простое, если и только если $S(r)=r+1$.

(2) Покажите, что число n совершенное, если и только если $S(n)=2n$.

(3) Пусть b_1 и b_2 — взаимно простые натуральные числа. Покажите, что b_1b_2 делится на d, если и только если d представимо в виде $d=d_1d_2$, где $d_1=$ $=\text{НОД}(d,b_1)$, а $d_2=\text{НОД}(d,b_2)$.

(4) С помощью утверждения (3) покажите, что если b_1 и b_2 *взаимно просты*, то $S(b_1b_2)=S(b_1)S(b_2)$.

10. Всякое четное число n можно записать в виде $n=2^st$, где $s\geqslant 1$, а число t нечетное. Предположим, что n совершенное.

(1) Подставьте $n = 2^s t$ в формулу $S(n) = 2n$ и, воспользовавшись утверждением (4) упражнения 9, покажите, что $S(t)$ должно делиться на 2^{s+1}.
(2) Выведите из (1), что $S(t) = 2^{s+1}q$ для некоторого натурального q. Покажите, что $t = (2^{s+1} - 1)q$.
(3) Мы хотим доказать от противного, что $q = 1$. Предположим, что $q > 1$. Тогда из (2) следует, что у q по крайней мере три различных делителя, а именно $1, q$ и t. Поэтому $S(t) \geqslant 1 + q + t$. Покажите, что $S(t) = 2^{s+1}q = t + q$ и получите искомое противоречие.
(4) Из (3) вытекает, что $q = 1$. Подставив это значение в формулу для t, мы получаем $t = 2^{s+1} - 1$ и $S(t) = 2^{s+1}$. Значит, $S(t) = t + 1$ и из утверждения (1) упражнения 9 вытекает, что t простое число.

Окончательно мы заключаем, что $n = 2^s(2^{s+1} - 1)$, причем второй сомножитель простой.

11. Обозначим через $d(n)$ число положительных делителей натурального числа n. Число n называется *сильно составным*, если $d(m) < d(n)$ для всех $m < n$. Напишите программу, которая по заданному натуральному числу r выдает все сильно составные числа, меньшие r. С помощью Вашей программы составьте список всех сильно составных чисел, меньших 5000. Что можно сказать о простых делителях этих чисел, рассматривая разложения элементов списка на множители? Сильно составные числа были введены и изучены знаменитым индийским математиком Шринивасой Рамануджаном (Srinivasa Ramanujan), см. [40].

12. Напишите программу, реализующую алгоритм Ферма разложения на множители. Входом программы должно служить любое положительное число, меньшее 2^{32}, а выходом — его разложение в произведение двух множителей или сообще-

ние о том, что оно простое. Не забудьте, что с четными числами алгоритм Ферма работает неправильно, поэтому входное число нужно сначала проверить на нечетность. Это упражнение начинает серию, которая заканчивается упражнением 8 главы 12.

Глава 4.

Простые числа

В первых двух главах мы изучали некоторые свойства простых чисел, без которых много не докажешь, и два алгоритма, без которых много не вычислишь. Содержание этой главы более явно связано с нашей конечной целью — RSA-криптосистемой. В самом деле, для уверенности в безопасности реализации RSA мы должны уметь подбирать большие простые числа: по два для каждого пользователя. В настоящей главе мы приступаем к решению этой задачи. Сначала рассмотрим простые числа, получающиеся из полиномиальных, экспоненциальных и прайморіальных формул. Важным следствием изучения прайморіальной формулы будет доказательство бесконечности множества простых чисел. Глава заканчивается обсуждением решета Эратосфена — старейшего из известных методов нахождения простых чисел и прародителя всех современных решет.

§ 4.1. Полиномиальная формула

Представление большинства людей о «формуле простых чисел» можно было бы выразить в следующем определении. Функция $f : \mathbb{N} \longrightarrow \mathbb{N}$ называется *формулой простых чисел*, если

$f(m)$ — простое число для каждого $m \in \mathbb{Z}$. Как мы увидим, это определение слишком амбициозно. Вместо «формулы простых чисел» мы будем искать функции, среди значений которых часто встречаются простые числа. Поскольку многочлен — простейшая из возможных функций, стоит начать с вопроса: существует ли полиномиальная формула простых чисел?

Как следует из определения, данного выше, многочлен

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0$$

с целыми коэффициентами a_n, a_{n-1}, $\ldots$, a_1, a_0 является *формулой простых чисел*, если число $f(m)$ простое для каждого положительного целого m. Поэкспериментируем с многочленом $f(x) = x^2 + 1$. Начнем с вычисления $f(x)$ для нескольких натуральных значений переменной x. Результат представлен в следующей таблице.

x	$f(x)$	Простое?
1	2	да
2	5	да
3	10	нет
4	17	да
5	26	нет
6	37	да
7	50	нет
8	65	нет
9	82	нет
10	101	да

Заметим, что если x нечетно, то $f(x)$ четно. Таким образом, $f(x)$ — всегда четное, а значит и составное число для нечетных значений x (за исключением $x = 1$, так как $f(1) = 2$). Значит, если $x > 1$ и $f(x)$ — простое, то x обязательно четное. Следовательно, если бы числа $f(x)$ были простыми при каждом четном x, то многочлен $f(2x)$ был бы формулой простых чисел. К сожалению, это не так; например, $f(8) = 65$ —

составное число. Итак, многочлен $f(x) = x^2 + 1$ не дает формулу простых чисел в смысле данного выше определения. Конечно, это только один пример, и мы могли бы надеяться, что нам не повезло с выбором многочлена. Но следующий результат показывает, что дело не в этом.

Теорема. *Для данного многочлена $f(x)$ существует бесконечно много положительных целых чисел m, при которых число $f(m)$ составное.*

Мы докажем эту теорему только для многочленов степени 2. Общий случай доказывается аналогично, только формулы будут более громоздкими, и усилия, затрачиваемые для их понимания, могут легко затмить ключевые идеи.

Пусть $f(x) = ax^2 + bx + c$ — многочлен с целыми коэффициентами a, b и c. Мы можем предполагать, что $a > 0$. Это означает, что $f(x)$ положителен для достаточно больших значений переменной x. Если $f(x)$ составное число для каждого натурального x, то доказывать нечего. Заметим, что такое на самом деле случается; например, если $f(x) = 4x$. Таким образом, мы можем предполагать: существует такое натуральное число m, что $p = f(m)$ — простое.

Пусть h — произвольное натуральное число. Вычислим значение $f(m + hp)$. Возникает резонный вопрос: откуда появилось $m + hp$? Наилучший ответ на него заключен в выкладках, приведенных ниже. Мы хотим найти

$$f(m + hp) = a(m + hp)^2 + b(m + hp) + c.$$

Раскрывая квадрат и собирая вместе члены, содержащие p, получаем

$$f(m + hp) = (am^2 + bm + c) + p(2amh + aph^2 + bh).$$

Заметим, что выражение в первой скобке равно $f(m) = p$, так что

$$f(m + hp) = p(1 + 2amh + aph^2 + bh). \qquad (1.1)$$

Формула (1.1) склоняет нас к предположению о разложимости числа $f(m+hp)$. Действительно, оно равно произведению p на натуральное число, что означает окончание доказательства. К сожалению, в этом рассуждении есть ошибка. Число $f(m+hp)$ будет простым только в том случае, если выражение в скобках в правой части формулы (1.1) не равно 1. Поэтому нам нужно найти такое значение h, при котором

$$1 + 2amh + aph^2 + bh > 1.$$

Последнее неравенство эквивалентно

$$2amh + aph^2 + bh > 0.$$

Поскольку число h положительно (по предположению), требуемое неравенство выполнено только если

$$2am + aph + b > 0, \quad \text{т.е.} \quad h > \frac{-b - 2am}{ap}.$$

Заметим, что $-b-2am$ может быть положительным числом в том случае, если b отрицательно и меньше, чем $-2am$.

Что же мы доказали? Мы показали, что если $f(x) = ax^2 + +bx+c$ — многочлен с целыми коэффициентами и положительным a, и число $f(m) = p$ — простое, то $f(m + hp)$ является составным для всех $h > (-b - 2am)/ap$. В частности, найдется бесконечно много натуральных значений x, при которых $f(x)$ — составное число.

Как мы уже отмечали, похожее доказательство работает и для многочленов произвольной фиксированной степени. Конечно, вычисление значения $f(m + ph)$ не такое компактное, но главная трудность связана с определением нижней границы для h. Поскольку у нас был квадратный многочлен, эта граница легко получалась из линейного неравенства. В общей ситуации, когда мы работаем с многочленом степени n, нижняя граница для h определяется из неравенства, содержащего

многочлен степени $n-1$. Эта трудность проиллюстрирована в упражнении 1, где рассматривается случай кубического многочлена. Если же степень многочлена больше 3, то простой формулы, выражающей нижнюю границу для h, нет. В такой ситуации мы были бы рады показать, что нижняя граница в принципе существует, даже если не сможем выписать для нее точную формулу. Доказательство существования нижней границы требует применения элементов математического анализа, и мы не будем его приводить. Доказательство этой теоремы можно найти в [41] ([Д.14]).

Доказанная теорема означает, что ответ на вопрос, сформулированный в начале параграфа, отрицателен. Однако мы рассматривали только многочлены от одной переменной. Неожиданным образом существуют многочлены от нескольких переменных, все положительные значения которых — простые числа. Проблема в том, что эти многочлены во многом неопределенны, так что их использование для нахождения простых чисел не очень практично. Примеры смотри в [41] ([Д.14]).

§ 4.2. Экспоненциальные формулы: числа Мерсенна

Есть две экспоненциальные формулы огромной исторической важности. Обе изучались математиками XVII и XVIII веков, в особенности Ферма и Эйлером. Вот эти формулы:

$$M(n) = 2^n - 1 \qquad \text{и} \qquad F(n) = 2^{2^n} + 1,$$

где n — натуральное. Числа из первой формулы называются *числами Мерсенна*, а из второй — *числами Ферма*.

Вопрос о том, при каких значениях n числа Мерсенна просты, восходит к математикам античной Греции. В пифагорейском мистицизме число называлось *совершенным*, если оно

равняется полусумме своих положительных делителей. Например, делители числа 6 — это 1, 2, 3 и 6. Складывая их, получаем:

$$1 + 2 + 3 + 6 = 12 = 2 \cdot 6.$$

Следовательно, 6 — совершенное число. Конечно, никакое простое число не будет совершенным. Действительно, делители простого числа p — это 1 и p, и $1 + p < 2p$, поскольку $p > 1$.

Эвклид знал, что число $2^{n-1}(2^n - 1)$ совершенно, если $2^n - 1$ простое. Нетрудно показать, что все *четные* совершенные числа имеют такой вид, но доказан этот факт был только Эйлером в восемнадцатом веке. Доказательство упомянутых результатов можно найти в упражнениях 8, 9 и 10 главы 3. Формула Эвклида сводит задачу о поиске четных совершенных чисел к нахождению простых чисел Мерсенна.

Задача о нахождении совершенных чисел, имеющая своим истоком туманный мистицизм пифагорейцев, может показаться крайне странной некоторым, живущим в конце двадцатого века. Однако, факт остается фактом: несмотря на то, что проблема стоит около 2500 лет, у нее все еще нет удовлетворительного решения. Например, неизвестно, обязано ли совершенное число быть четным, хотя к настоящему времени не найдено ни одного нечетного. Конечно, возраст этой проблемы бросает труднопреодолимый вызов всем, кто любит числа. Более того, ее сложность может означать, что она относится к глубочайшим свойствам целых чисел. Это делает ее даже еще более важной с точки зрения математиков.

Как мы упоминали во введении, Марэн Мерсенн был священником и математиком-любителем семнадцатого века. Числа вида $2^n - 1$ обязаны своим именем утверждению Мерсенна о том, что они просты, в случае

$$n = 2,\ 3,\ 5,\ 7,\ 13,\ 17,\ 19,\ 31,\ 67,\ 127,\ \text{и}\ 257,$$

и являются составными для всех остальных 44 положительных простых n, меньших 257.

Первое важное замечание: Мерсенн рассматривал значения функции $2^n - 1$ только при простых n. Действительно, если n — составное, то такое же и $M(n)$. Предположим, что $n = rs$, $1 < r$ и $s < n$, тогда

$$M(n) = 2^n - 1 = 2^{rs} - 1 = (2^r - 1)(2^{r(s-1)} + 2^{r(s-2)} + \cdots + 2^r + 1).$$

Следовательно, если r делит n, то $M(r)$ делит $M(n)$. Второй важный момент заключается в том, что обратное неверно. Иначе говоря, если n — простое, то $M(n)$ не обязано быть простым. Мы видим из списка Мерсенна, что $M(11)$ должно быть составным. Это легко проверить:

$$M(11) = 2047 = 23 \cdot 89.$$

Как часто бывало в то время, Мерсенн не привел доказательства своего утверждения, что дало повод для сомнений в его истинности и оставило широкое поле деятельности для математиков. В поисках простых чисел Мерсенна принял участие и Эйлер. В 1732 году он нашел два «новых простых» числа: $M(41)$ и $M(47)$, отсутствовавших в списке Мерсенна. Позже выяснилось, что в этом случае Эйлер был не прав. Первую ошибку в списке Мерсенна нашли Первузэн (Pervusin) и Зилхоф (Seelhov) в 1886. Они обнаружили, что число $M(61)$ простое, хотя его и нет в списке. Другие ошибки были найдены в последующие годы. Сейчас известно, что кроме $M(61)$, в списке пропущены простые числа $M(89)$ и $M(107)$, и присутствуют составные числа $M(67)$ и $M(257)$.

При доказательстве простоты чисел Мерсенна, Ферма использовал метод разложения на множители, который будет описан в § 10.1. В наше время используется гораздо более эффективный *тест Люка–Лемера*, изучаемый в § 10.4. С помощью этого теста в 1998 году было показано, что число Мерсенна $M(3\,021\,377)$ просто. Оно состоит из 1 819 050 знаков и является наибольшим из простых чисел, известных к моменту издания этой книги.

§ 4.3. Экспоненциальные формулы: числа Ферма

История чисел Ферма очень схожа с историей чисел Мерсенна. Ферма знал, что если 2^m+1 простое, то m должно быть степенью двойки. Поэтому, интересуясь простыми, необходимо смотреть только на числа вида $2^{2^n}+1$. В письме, адресованном шевалье Френиклю (Frenicle), другому математику-любителю, Ферма выписал эти числа для $n = 0,\ 1,\ \ldots,\ 6$:

$$3;\ 5;\ 17;\ 257;\ 65\,537;\ 4\,294\,967\,297 \text{ и } 18\,446\,744\,073\,709\,551\,617.$$

Затем он предположил, что все числа вида $2^{2^n}+1$ простые. Как ни странно, Ферма, кажется, не пытался разлагать на множители эти числа методом, аналогичным использованному им для разложения чисел Мерсенна. Если бы он применил этот метод, то увидел бы, что число $F(5)$ составное. Необходимую проверку Эйлер сделал столетие спустя. Мы будем изучать его метод в § 10.2.

Интересно также, что Френикль не обнаружил ошибки Ферма. В конце концов, он был слишком занят попытками разложения чисел Мерсенна на множители. Френикль не мечтал затмить Ферма-математика, но тон его корреспонденции наводит на мысль, что он очень хотел найти ошибку в работе Ферма. Тем не менее, кажется, он был согласен с Ферма в справедливости его гипотезы.

В отличие от чисел Мерсенна, о которых известно, что они являются богатым источником больших простых чисел, среди чисел Ферма простых известно очень мало. Фактически, все известные простые числа Ферма — это $F(0), \ldots, F(4)$ — список, неизменный со времен Ферма. Конечно, вычислять числа Ферма при «больших» значениях n очень трудно. В конце концов, формула, описывающая эти числа — двойная экспонента, т.е. экспонента от экспоненты.

В двух предыдущих параграфах мы немного познакомились с историей самых известных чисел, описываемых экспоненциальными формулами. Доказательства упомянутых результатов мы оставим до главы 10. В настоящий момент нам придется довольствоваться знанием того, что числа Мерсенна — неисчерпаемый источник очень больших простых чисел.

Следует указать, что предложенный Ферма метод разложения чисел Мерсенна весьма прост для объяснения и нетруден при доказательстве. Но есть более элементарное рассуждение, которое нашел бы и сам Ферма; оно требует только нескольких удачных отождествлений (см. [8]). Несмотря на это, мы откладываем изучение метода Ферма до девятой главы. К тому моменту мы будем владеть основными понятиями и теоремами теории групп, которые позволят нам дать более короткое и прозрачное обоснование метода Ферма. В качестве бесплатного приложения мы сможем использовать те же самые идеи в ряде других случаев, один из которых — метод Эйлера определения делителей чисел Ферма.

Один из фундаментальных принципов развития математики заключается в том, что важные частные задачи нередко решаются только после развития общих методов и абстрактных теорий, выявляющих связи и аналогии между результатами, которые раньше считались имеющими мало общего. Эти связи, в свою очередь, часто указывают на неожиданные приложения новых методов.

§ 4.4. Прайморіальная формула

Напомним, что факториалом натурального числа n называется произведение всех положительных целых чисел, не превосходящих n. Аналогично мы определяем *прайморіал* $p^{\#}$ простого $p > 0$ как произведение всех простых чисел, меньших или равных p. Например, $2^{\#} = 2$ и $5^{\#} = 2 \cdot 3 \cdot 5 = 30$. Заметим,

что если q — следующее после p простое число, то

$$q^{\#} = p^{\#} q.$$

Мы хотим рассмотреть числа вида $p^{\#} + 1$. Чтобы понять, зачем, посмотрите на таблицу:

p	$p^{\#}$	$p^{\#} + 1$
2	2	3
3	6	7
5	30	31
7	210	211
11	2 310	2 311

Все числа в правом столбце таблицы — простые! Может ли это быть просто совпадением? Если этот вопрос вселяет в вас надежду на то, что все числа вида $p^{\#} + 1$ простые, вам бы следовало попытаться заполнить следующую строку таблицы. На самом деле,

$$13^{\#} + 1 = 30\,031 = 59 \cdot 509$$

является составным числом.

Однако, хотя $p^{\#} + 1$ не всегда простое, мы можем показать, что оно не имеет делителей, меньших или равных p. Воспользуемся методом «от противного». Предположим, что $q \leqslant p$ — простой делитель числа $p^{\#} + 1$. Поскольку $p^{\#}$ — произведение всех простых чисел вплоть до p, q должно также делить $p^{\#}$. Следовательно, q делит разность

$$(p^{\#} + 1) - p^{\#} = 1.$$

Значит, $q = 1$, что противоречит его простоте. В итоге мы получаем: наименьший делитель числа $p^{\#} + 1$ должен быть больше p.

Это наблюдение может навести на мысль о следующем алгоритме получения больших простых чисел. Пусть мы знаем

все простые числа вплоть до p. Вычисляем $p^{\#} + 1$. Если оно простое, то все сделано. Если нет, то найдем его наименьший простой делитель; он должен быть больше, чем p. В любом случае, мы нашли простое число, большее p.

Описанный подход плох по нескольким причинам. Наиболее очевидная из них — необходимость разложения на простые множители числа $p^{\#} + 1$. Даже для сравнительно небольших значений p прайморіал $p^{\#}$ огромен и разложение его на множители весьма проблематично.

С другой стороны, если повезет, число $p^{\#} + 1$ может оказаться простым. А как мы увидим позже, существуют вполне приемлемые способы проверки простоты больших чисел, которые не используют разложения на множители. Простое число, представимое в такой форме, называется *прайморіально простым*. Конечно, остается наивный подход к проверке простоты, заключающийся в систематических попытках подобрать подходящий делитель числа. Как мы видели в § 3.3, этот алгоритм весьма неэффективен. Специальный алгоритм тестирования простоты чисел вида $p^{\#} + 1$ будет изучаться в главе 11. Несмотря на то, что он очень удобен, пока найдено только 16 прайморіально простых, наибольшее из которых соответствует $p = 24\,027$ и насчитывает 10 387 знаков.

Итак, прайморіальная формула не дает очень уж удобного пути к отысканию больших простых чисел; к счастью, на ней свет клином не сошелся.

§ 4.5. Бесконечность множества простых чисел

Истинная причина, по которой мы так долго и детально разбирались с прайморіальной формулой, состоит в том, что она дает нам самое быстрое доказательство следующего фундаментального результата.

Теорема. *Простых чисел бесконечно много.*

Доказательство, которое мы здесь приводим, можно найти в «Элементах» Эвклида как предложение 20 книги IX. Доказываем «от противного». Предположим, множество простых чисел конечно. Это означает, что существует наибольшее простое число; скажем, p. Другими словами, мы предполагаем, что все числа, бóльшие p, — составные. Однако, как мы видели в предыдущем параграфе, число $p^{\#}+1$ не может иметь простых делителей меньших или равных p. Из предположения и последнего утверждения вытекает, что у $p^{\#}+1$ нет простых делителей, т.е. оно само является простым числом. А поскольку $p^{\#}+1>p$, мы получаем противоречие с предположением: p — наибольшее простое число. Итак, простых чисел должно существовать бесконечно много.

Было найдено много других доказательств бесконечности ряда простых чисел. Доказательство Эйлера (1737 года) носит особый характер. Оно оказалось тем семенем, из которого позже выросли многие достижения, поэтому мы приведем его здесь практически полностью. Подобно доказательству Эвклида, оно тоже идет «от противного». Итак, предположим, что существует только конечное число простых чисел, и пусть p — наибольшее из них. Рассмотрим произведение множителей

$$P=\left(\frac{1}{1-1/2}\right)\left(\frac{1}{1-1/3}\right)\left(\frac{1}{1-1/5}\right)\cdots\left(\frac{1}{1-1/p}\right), \qquad (5.1)$$

по одному для каждого простого числа. Естественно, это произведение равно некоторому положительному вещественному числу. Более того, аккуратно перемножая члены произведения, можно показать, что

$$P=1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\cdots, \qquad (5.2)$$

где теперь у нас есть свое слагаемое для каждого натурального числа. Это равенство мы докажем чуть позже, а сейчас

попытаемся привести его к противоречию. Хотя число слагаемых в выражении (5.2) бесконечно, его сумма еще могла бы оказаться конечным числом; например, бесконечная сумма $1+1/2+1/2^2+1/2^3+1/2^4+\cdots$, как сумма бесконечной геометрической прогрессии со знаменателем $1/2$, равна $\frac{1}{1-1/2}=2$. Но нетрудно увидеть, что сумма, соответствующая P, не может быть равной ни одному вещественному числу. Для начала заметим, что

$$\begin{aligned}
\frac{1}{3}+\frac{1}{4} &\geqslant 2\cdot\frac{1}{4}=\frac{1}{2}\\
\frac{1}{5}+\frac{1}{6}+\frac{1}{7}+\frac{1}{8} &\geqslant 4\cdot\frac{1}{8}=\frac{1}{2}\\
\cdots &\quad \cdots\\
\frac{1}{2^{n-1}+1}+\cdots+\frac{1}{2^n} &\geqslant 2^{n-1}\frac{1}{2^n}=\frac{1}{2}.
\end{aligned}$$

Следовательно,

$$P>1+\frac{1}{2}+\frac{1}{3}+\frac{1}{4}+\frac{1}{5}+\frac{1}{6}+\cdots+\frac{1}{2^n}\geqslant n\cdot\frac{1}{2}=\frac{n}{2}$$

для любого данного натурального n. Таким образом, P больше любого наперед заданного числа, поэтому оно не может быть каким-то вещественным числом. Полученное противоречие доказывает, что простых чисел бесконечно много.

Теперь вернемся к доказательству равенства (5.2). Фактически, нам нужно показать, что произведение (5.1) совпадает с бесконечной суммой (5.2). Напомним, что дробь $\frac{1}{1-q}$ (при $q\in(0;1)$) можно интерпретировать как сумму бесконечной геометрической прогрессии со знаменателем q:

$$\frac{1}{1-q}=1+q+q^2+q^3+q^4+\cdots+q^n+\cdots.$$

Таким образом, равенство (5.1) можно переписать в виде:

$$P=\left(1+\tfrac{1}{2}+\tfrac{1}{2^2}+\cdots\right)\cdot\left(1+\tfrac{1}{3}+\tfrac{1}{3^2}+\cdots\right)\cdots\left(1+\tfrac{1}{p}+\tfrac{1}{p^2}+\cdots\right).$$

Раскроем скобки в этом произведении. Такую операцию можно сделать двумя способами. Один из них заключается в *последовательном* перемножении скобок: сначала раскрываем произведение первых двух скобок, затем полученный результат умножаем на третью, и т.д. Это привычный способ, но длинный и малоэффективный. Действительно, мы не знаем точное количество перемножаемых скобок, да и каждая скобка представляет собой бесконечную сумму.

Другой способ, менее привычный, но более грамотный, основан на простом наблюдении: если бы у нас хватило терпения и сил раскрыть скобки первым методом, то мы получили бы бесконечную сумму произведений $\frac{1}{2^{r_1}}\frac{1}{3^{r_2}}\frac{1}{5^{r_3}}\cdots\frac{1}{p^{r_s}}$ ($r_i \geqslant 0$), по *одному* сомножителю из *каждой* скобки (значение показателя $r_i = 0$ говорит о том, что из i-ой скобки в качестве сомножителя мы берем 1). Мы не будем объяснять это наблюдение более подробно, поскольку такого сорта утверждения легче осознать самостоятельно, нежели понять их доказательство.

Итак, число P из равенства (5.1) представляет собой бесконечную сумму слагаемых вида

$$\frac{1}{2^{r_1}3^{r_2}5^{r_3}\cdots p^{r_s}}, \quad r_1 \geqslant 0,\ r_2 \geqslant 0, \ldots, r_s \geqslant 0. \tag{5.3}$$

Упорядочим эти слагаемые по убыванию. Самым большим будет член $\frac{1}{2^0 3^0 5^0 \cdots p^0} = 1$ затем идет $\frac{1}{2^1 3^0 5^0 \cdots p^0} = 1/2$, потом $\frac{1}{2^0 3^1 5^0 \cdots p^0} = 1/3$, после него — $\frac{1}{2^2 3^0 5^0 \cdots p^0} = 1/4$, и т.д. Как видите, мы получили начало бесконечной суммы (5.2). Нам осталось показать, что число P на самом деле совпадает с этой суммой. Для этого необходимо проверить две вещи:

(а) для любого натурального числа n среди дробей (5.3) найдется такая, что

$$\frac{1}{n} = \frac{1}{2^{r_1}3^{r_2}5^{r_3}\cdots p^{r_s}};$$

(б) каждая дробь вида $\frac{1}{n}$ встречается в бесконечной сумме, представляющей P, только один раз.

Начнем, как ни странно, с проверки второго утверждения (оно несколько проще). Предположим, что нашлись две *разные* дроби, равные $\frac{1}{n}$, т.е.

$$\frac{1}{n} = \frac{1}{2^{r_1}3^{r_2}5^{r_3}\cdots p^{r_s}} = \frac{1}{2^{k_1}3^{k_2}5^{k_3}\cdots p^{k_s}}.$$

Поскольку числители всех этих дробей равны 1, мы получаем равенство знаменателей:

$$n = 2^{r_1}3^{r_2}5^{r_3}\cdots p^{r_s} = 2^{k_1}3^{k_2}5^{k_3}\cdots p^{k_s}.$$

Как первое произведение в этом равенстве, так и второе, целиком состоит из простых чисел, т.е. мы двумя *разными способами* разложили натуральное число n на простые сомножители, что противоречит основной теореме арифметики. Значит, $r_1 = k_1,\ r_2 = k_2,\ \ldots,\ r_s = k_s$, т.е. дроби совпадали изначально. Таким образом, утверждение (б) проверено.

Перейдем к утверждению (а). По основной теореме арифметики любое натуральное число n представляется в виде произведения простых сомножителей: $n = p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_m^{\alpha_m}$. С другой стороны, знаменатель каждой из дробей (5.3) — это произведение степеней *всех* простых чисел: $2^{r_1} \cdots p^{r_s}$, причем их показателями могут быть любые неотрицательные целые числа. Таким образом, чтобы приравнять дробь вида (5.3) числу $\frac{1}{n}$, достаточно выбрать подходящие показатели r_i.

«Ага! — скажете Вы, — Вот я и поймал педанта-лектора! Как же могут быть равными дроби

$$\frac{1}{p_1^{\alpha_1} \cdot p_2^{\alpha_2} \cdots p_m^{\alpha_m}} \qquad \text{и} \qquad \frac{1}{2^{r_1} \cdots p^{r_s}},$$

если в первой из них присутствуют только *некоторые* простые числа, в то время как во второй — *все*?» Не спешите с выводами. Напомню, что показатели r_i могут обращаться в нуль. Вот мы и возьмем нулевые показатели для всех тех простых чисел, которых нет в знаменателе левой дроби. Итак,

утверждение (а) также проверено. То есть мы доказали, что произведение (5.1) и бесконечная сумма (5.2) равны одному и тому же числу P.

Все было бы замечательно, если бы не мелкое жульничество, которое было допущено в этом рассуждении. Оно заключается в приглашении: «упорядочим эти слагаемые по убыванию». Дело в том, что оно относится к бесконечной сумме, неявно подразумевая, что эта сумма не зависит от порядка слагаемых. К сожалению, в общем виде такое утверждение просто неверно. Однако, если, как и в нашем случае, суммируются неотрицательные вещественные числа, то результат суммирования даже бесконечного числа слагаемых не зависит от порядка. Доказательство этого факта выходит за рамки нашего учебника. Заинтересованный читатель может прочесть об этом в [Д.6]. Более того, там же он может познакомиться с удивительной теоремой Римана, которая утверждает, что в некоторых бесконечных суммах можно так поменять порядок слагаемых, что в результате получится любое, наперед заданное число.

Вот теперь мы полностью привели доказательство Эйлера бесконечности простых чисел.

Конечно, если бы множество простых чисел было ограничено, жизнь была бы проще, но мир стал бы скучнее. Тот факт, что простых чисел бесконечно много, ставит много интересных проблем. Например, что можно сказать об их распределении? Растет или убывает «плотность» простых чисел, когда мы переходим ко все бóльшим и бóльшим числам? Существует ли возможность измерить эту «плотность»? Наилучший способ точно сформулировать проблему о распределении простых чисел состоит в использовании π-функции. Для вещественного положительного числа x обозначим через $\pi(x)$ количество простых чисел, не превосходящих x. Хорошая оценка функции $\pi(x)$ — важная задача теории чисел.

Упомяните при математиках о распределении простых чисел, и вы тут же услышите имя Римана. Идеи, рожденные до-

казательством Эйлера бесконечности ряда простых чисел, легли в основу работы Б. Римана, ставшей фундаментальным трудом, посвященным функции $\pi(x)$ и родственным вопросам. Эта статья, опубликованная в 1895 году, содержит много интересных очаровательных результатов, часть которых приведена без доказательства. К сожалению, Риман умер от туберкулеза семь лет спустя, не закончив детальную проработку своих доказательств. Эту работу взвалило на свои плечи несколько математиков, в первую очередь Адамар (Hadamard).

Одним из результатов усилий Адамара по заполнению пробелов, оставленных Риманом, стало доказательство знаменитой *теоремы о простых числах*, которая говорит, что

$$\lim_{x\to\infty} \frac{\pi(x)\ln x}{x} = 1,$$

где $\ln x$ — логарифм по основанию e (т.е. натуральный логарифм). Этот результат даже старше работы Римана; его сформулировал Гаусс в качестве гипотезы. Доказан же он был в 1896 году независимо друг от друга Адамаром и де ля Валле-Пуссеном (de la Vallé-Poussin).

Нестрого говоря, теорема о простых числах утверждает, что для очень больших значений x число $\pi(x)$ приблизительно равно $x/\ln(x)$. Но аппроксимация будет хорошей только если x поистине велико. Например, если $x = 10^{16}$, то разность

$$\pi(x) - \left[\frac{x}{\ln x}\right] = 7\,804\,289\,844\,393$$

будет величиной порядка 10^{13}. Поскольку в этом случае $x/\ln x$ имеет порядок 10^{14}, то ошибка действительно значительна. Есть много других простых функций, дающих неплохую аппроксимацию функции $\pi(x)$ при больших x. Одна из них изучается экспериментальным образом в упражнении 11. Подробное обсуждение распределения простых чисел можно найти в [23] и [24] ([Д.10]). История теоремы о простых числах содержится в [7].

§ 4.6. Решето Эратосфена

Решето Эратосфена — это старейший из известных способов выписывания простых чисел. В отличие от методов, обсуждавшихся в предыдущих параграфах, он не использует никакой специальной функции. Эратосфен (Erathostenes) был греческим математиком, родившимся около 284 года до н.э. Он владел многими отраслями знаний, однако современники не считали его выдающимся специалистом ни в одной из них. Они прозвали его «Бета» (вторая буква греческого алфавита) и «Пентатлосом»[1]. Вот уже 2300 лет мы пользуемся его работами, а полученные им прозвища являются лишним подтверждением величия древнегреческой математики.

В своей «Арифметике», опубликованной около 100 года н.э., Никомах из Герасы (Nicomachus of Gerasa) вводит решето Эратосфена следующим образом:

> «Метод для получения этих [простых чисел] называется по Эратосфену решетом, так как мы берем все нечетные числа, смешанные беспорядочно вместе, и выбрасывая из них, как неким инструментом, или решетом, мы отделяем в первую очередь неразложимые, а во вторую составные посредством их самих.»

Для дальнейшего знакомства с высказыванием Никомаха о решете смотри [47].

Итак, решето получило свое имя потому, что когда оно применяется к списку натуральных чисел, составные числа просеиваются, а простые задерживаются. Посмотрим, как оно работает.

Прежде всего, цель решета — определить все положительные простые числа, меньшие некоторой верхней границы $n > 0$,

[1] Пентатлос (Pentatlos) в переводе с греческого языка означает «пятый». — *Прим. перев.*

которую мы предполагаем целой. Чтобы использовать метод решета, как это делал Эратосфен (имея только карандаш и бумагу), мы поступаем следующим образом. Сначала выписываем все нечетные целые между 3 и n. Причина, по которой мы не трогаем четные числа, заключается в том, что, кроме 2, среди них нет простых чисел.

Теперь мы начинаем просеивать список. Первое число в нем 3. Начиная со следующего числа в списке (это 5), мы вычеркиваем из него каждое третье число. Проделав это до конца, мы вычеркнем все числа из списка, кратные 3 и бóльшие самой тройки.

Теперь выберем наименьшее число из списка, превосходящее 3, которое еще не было вычеркнуто. Таким будет 5, а следующее за ним число — 7. Вычеркиваем каждое пятое число из нашего списка, начиная с 7. Таким образом, все числа, кратные 5, будут вычеркнуты. Продолжаем эту процедуру, пока не дойдем до n. Заметим, что если мы собираемся вычеркивать каждое p-ое число, то всегда надо начинать отсчет с числа $p+2$, даже когда это число было вычеркнуто на предыдущих шагах.

Например, для $n = 41$ список нечетных чисел выглядит так:

3 5 7 9 11 13 15 17 19 21
23 25 27 29 31 33 35 37 39 41

Вычеркнув каждое третье число начиная с 5, мы получим

3 5 7 ~~9~~ 11 13 ~~15~~ 17 19 ~~21~~
23 25 ~~27~~ 29 31 ~~33~~ 35 37 ~~39~~ 41

Теперь мы вычеркиваем каждое пятое число, начиная с 7, что дает

3 5 7 ~~9~~ 11 13 ~~15~~ 17 19 ~~21~~
23 ~~25~~ ~~27~~ 29 31 ~~33~~ ~~35~~ 37 ~~39~~ 41

Мы должны бы теперь вычеркнуть каждое седьмое число, начиная с 9. Но если мы это сделаем, то никакие новые числа

не отсеются. Далее нам нужно бы вычеркнуть каждое одиннадцатое число, начиная с 13, но это опять не даст никакого эффекта. На самом деле, ни одно из чисел, оставшихся в списке после вычеркивания каждого пятого, не будет позже вычеркнуто ни на каком этапе просеивания. Итак, положительные нечетные простые числа, не превосходящие 41, это

$$3 \quad 5 \quad 7 \quad 11 \quad 13 \quad 17 \quad 19 \quad 23 \quad 29 \quad 31 \quad 37 \quad 41.$$

В разобранном примере есть пара важных обстоятельств, которые нужно взять на заметку. Во-первых, хотя нам следовало повторять процесс вычеркивания вплоть до граничного числа n (41 в нашем примере), мы избавились от всех составных чисел к тому моменту, когда отсеяли кратные 5, и все остальные просеивания оказались излишними. Во-вторых, некоторые числа вычеркивались больше, чем один раз. Такое произошло, например, с 15. Первый раз оно было вычеркнуто, когда мы отсеивали кратные 3. Но 15 также делится на 5, поэтому оно было вычеркнуто снова при отсеивании кратных 5.

Посмотрим, как можно увеличить эффективность решета в свете этих двух замечаний. Начнем со второго из них, т.е. посмотрим, можем ли мы так все организовать, чтобы каждое число вычеркивалось только один раз? К сожалению, ответ на этот вопрос отрицателен: нет хорошего способа добиться этого. Хотя кое-что в указанном направлении сделать все-таки можно.

Предположим, что мы отсеиваем числа, кратные какому-то простому p. Учитывая наше описание решета, нам следовало бы вычеркивать каждое p-ое число, начиная с $(p+2)$ — следующего за p числа в списке. Простейшее усовершенствование — это начинать процесс вычеркивания не с $p+2$, а с наименьшего числа, кратного p, *которое не делится на простое число, меньшее* p. Найдем его. Положительные числа, кратные p, записываются в виде kp, где k — натуральное. Если $k < p$, то kp делится на число, меньшее p, а именно на k. Значит, первое

кратное p число, которое не делится на простое число, меньшее p, есть p^2. Так что достаточно вычеркивать каждое p-ое число, начиная с p^2. Однако необходимо обратить внимание на то, что даже после этого нововведения останутся числа, которые будут вычеркиваться не единожды.

Что касается другого замечания, можем ли мы закончить отсев раньше, чем дойдем до n-го шага? На этот раз ответ положителен, он следует из того, что мы только что сделали. Допустим, например, что мы вычеркиваем каждое p-ое число. Как мы только что видели, первое число, подлежащее вычеркиванию, равно p^2. Но если $p^2 > n$, такого числа нет в списке и мы можем забыть о нем. В итоге, нам нужно вычеркивать каждое p-ое до тех пор, пока $p \leqslant \sqrt{n}$. Поскольку p — целое, это равносильно тому, что $p \leqslant \left[\sqrt{n}\right]$. В примере, разобранном выше, $\left[\sqrt{41}\right] = 6$. Вот почему отбрасывание кратных 3 и 5 было достаточно для вылавливания всех составных чисел из списка.

Сейчас мы должны обсудить компьютерную реализацию решета. Список нечетных чисел представляется одномерным массивом, который также принято называть вектором. Напомним, что с каждой ячейкой вектора ассоциируется два числа. Одно из них — значение ячейки, а другое идентифицирует ее местоположение в векторе. Например, в векторе

$$\begin{array}{ccccccccc} (& a & b & c & d & e & f & g &) \\ & & \uparrow & & & & & & \end{array}$$

величина ячейки, отмеченной стрелкой, равна b, а ее индекс — двум, поскольку она стоит на втором месте в векторе.

Вернемся к решету Эратосфена. Предположим, что мы хотим найти все простые числа, не превосходящие нечетного целого n. Сначала мы должны построить вектор с $(n-1)/2$ ячейками, по одной для каждого нечетного целого $2j+1$. Ячейки будут принимать одно из двух возможных значений: 1 или 0. Если значение ячейки равно 0, то нечетное число, ею предста-

вленное, было вычеркнуто на каком-то предыдущем шаге процесса просеивания. Итак, начальное значение каждой ячейки равно 1, поскольку еще нет никаких вычеркнутых чисел. Чтобы «вычеркнуть» число $2j+1$, нужно заменить 1, стоящую в j-ой ячейке вектора, на 0. Конечно, эта ячейка могла быть «вычеркнута» на предыдущем шаге решета. В этом случае ее значение уже равно 0 и не будет меняться в процессе выполнения следующих шагов алгоритма.

Теперь мы приведем более или менее подробную версию алгоритма для решета Эратосфена, который был описан выше. Она содержит оба усовершенствования, которые мы обсуждали, т.е. каждое p-ое число вычеркивается начиная с p^2, а алгоритм заканчивается, как только p превысит $\sqrt{n}$.

Решето Эратосфена

Ввод: нечетное натуральное n.
Вывод: список всех нечетных положительных простых чисел, меньших или равных n.
Шаг 1. Начинаем с создания вектора $\mathbf{v}$ с $(n-1)/2$ ячейками, каждой из которых присвоено значение 1, и полагаем $P=3$.
Шаг 2. Если $P^2 > n$, выписываем все числа $2j+1$, для которых значение j-ой ячейки вектора равно 1 и останавливаемся; в противном случае переходим к шагу 3.
Шаг 3. Если значение ячейки вектора $\mathbf{v}$ с номером $\frac{P-1}{2}$ равно 0, увеличиваем P на 2 и возвращаемся к шагу 2; в противном случае переходим к шагу 4.
Шаг 4. Присваиваем новой переменной T значение P^2; замещаем нулем значение ячейки вектора $\mathbf{v}$ под номером $\frac{T-1}{2}$ и увеличиваем T на $2P$; повторяем эти два шага до тех пор, пока $T \leqslant n$, затем увеличиваем P на 2 и возвращаемся к шагу 2.

Заметим, что на последнем шаге мы увеличиваем T на $2P$, а не на P, как можно было бы ожидать. Мы делаем это потому, что вектор $\mathbf{v}$ представляет набор нечетных чисел, так что как T, так и P нечетны. Поэтому, если мы вычеркиваем

каждое p-ое число, то число, которое будет вычеркнуто после T, есть $T + 2P$.

Вам может показаться, что существует простое изменение описанной выше процедуры, которое ускорит алгоритм. Способ, с помощью которого мы избавляемся от нежелательных составных чисел в векторе, состоит в замене 1, стоящей в соответствующей ячейке, на 0. Но почему, если мы не заботимся о составных числах, просто не выбросить их из вектора? К сожалению, мы не можем этого сделать. Неприятность в том, что метод, которым мы узнаем кратно ли число, соответствующее данной ячейке, некоторому p, зависит от номера ячейки. Другими словами, числа, кратные p, встречаются в каждой p-ой позиции вектора. Если мы удалим некоторые числа из списка, то алгоритм, который мы описали, перестанет работать.

Подобно всем алгоритмам, решето Эратосфена имеет ограничения. Например, оно неэффективно при поиске очень больших простых чисел. Напомним, однако, что цель этого алгоритма — поиск *всех простых* меньших, чем определенная верхняя граница. Ясно, что такой поиск невыполним, если граница слишком велика.

Суммируя ограничения, накладываемые назначением алгоритма, отметим два его слабых места: решето требует уйму компьютерной памяти и должно проделать слишком много витков цикла. К его достоинствам можно отнести простоту программирования и, кроме того, нам не нужно вычислять отдельные делители.

Упражнения

1. Пусть a, b, c и d — целые числа, причем $a > 0$. Рассмотрим многочлен $f(x) = ax^3 + bx^2 + cx + d$ степени 3. Предположим, что найдется такое натуральное число m, что $f(m) = p > 0$ простое. Найдите натуральное значение h, для которого $f(m + hp)$ — составное число.

2. Используя алгоритм деления методом проб из § 3.2, найдите все простые делители числа $p^{\#} + 1$ для

(1) $p = 17$;

(2) $p = 13$.

Нечетное простое число можно записать либо как $4n + 1$, либо как $4n + 3$. Другими словами, возможные остатки от деления этого числа на 4 — это 1 или 3. Например 3, 7, 11 и 19 имеют вид $4n + 3$, в то время, как 5 и 13 — вид $4n + 1$. Цель упражнений 3–7 — доказательство бесконечности множества простых чисел вида $4n + 3$. Это факт верен также и для простых вида $4n + 1$, но его доказательство не столь элементарно; оно может быть найдено в [23].

3. Покажите, что произведение двух целых чисел вида $4n+1$ имеет тот же вид.

4. Покажите, что каждое нечетное простое число имеет вид либо $4n + 1$, либо $4n + 3$.

5. Верно ли, что произведение двух чисел вида $4n+3$ тоже имеет вид $4n + 3$?

6. Предположим, что $3 < p_1 < \cdots < p_k$ — простые числа вида $4n + 3$. Используя упражнение 3, покажите, что число $4(p_1 \cdot p_2 \cdots p_k)+3$ должно делиться на простое число вида $4n+3$, не принадлежащее множеству $\{3,\ p_1,\ \ldots,\ p_k\}$.

7. Основываясь на предыдущем упражнении, покажите, что существует бесконечно много простых чисел вида $4n + 3$.

8. Мы говорили в упражнении 5 главы 2, что если $n > m$ — натуральные числа, то $\text{НОД}(F(n), F(m)) = 1$. Иначе говоря, два различных числа Ферма не могут иметь общий делитель. Используя этот факт, дайте другое доказательство бесконечности множества простых чисел.

9. Докажите, что если p, $p + 2$ и $p + 4$ — положительные простые числа, то $p = 3$.

10. Пусть f — квадратный многочлен. Напишите программу поиска целого числа n, меньшего 100, для которого $f(n)$ — простое число. Исходными данными программы будут коэффициенты a, b и c многочлена $f(x) = ax^2 + bx + c$. Эти коэффициенты предполагаются целыми, но могут быть как положительными, так и отрицательными. Программа будет вычислять $f(n)$ для всех неотрицательных целых n, меньших 100, и выбирать из них простые числа. Чтобы сделать это, мы сначала должны применить решето Эратосфена для выявления всех простых чисел, меньших, чем $\max\{|f(0)|, |f(100)|\}$. Заметим, что необходимо наложить разумные ограничения на $|a|$, $|b|$ и $|c|$. В противном случае $f(x)$ может выйти за границы области целых чисел, поддерживаемой языком программирования, которым Вы пользуетесь. Примените эту программу к каждому из следующих многочленов:

(1) $f(x) = x^2 + 1$;
(2) $f(x) = x^2 - 69x + 1231$;
(3) $f(x) = 2x^2 - 199$;
(4) $f(x) = 8x^2 - 530x + 7681$.

Второй многочлен — это вариант знаменитого примера, опубликованного Л. Эйлером в 1772 году.

11. Мы упоминали в § 4.5, что есть несколько формул, аппроксимирующих $\pi(x)$, число простых чисел, не превосходящих x. Например, в качестве следствия из теоремы о простых числах мы получили, что $x/\ln x$ приблизительно равно $\pi(x)$ для больших x. Но в этом случае, чтобы ошибка была незначительной, число x должно быть чудовищно большим. В этом упражнении мы экспериментально изучаем формулу, дающую лучшую аппроксимацию при малых x. Вот эта формула:

$$S(x) = \frac{x}{\ln x}\left(1 + \left[\sum_{k=0}^{12} a_k (\ln\ln x)^k\right]^{-\frac{1}{4}}\right),$$

где ln обозначает натуральный логарифм (по основанию e), и

$$\begin{aligned}
a_0 &= 229\,168,50747390, & a_1 &= -429\,449,7206839,\\
a_2 &= 199\,330,41355048, & a_3 &= 28\,226,22049280,\\
a_4 &= 0,\ a_5 = 0, & a_6 &= -34\,712,81875914,\\
a_7 &= 0, & a_8 &= 33\,820,10886195,\\
a_9 &= -25\,379,82656589, & a_{10} &= 8\,386,14942934,\\
a_{11} &= -1\,360,44512548, & a_{12} &= 89,14545378.
\end{aligned}$$

Напишите программу, вычисляющую $\pi(x)$ по данному натуральному x, положив в ее основу решето Эратосфена. Примените эту программу для определения разности $\pi(x) - S(x)$ при $x = 11; 100; 1000; \ldots; 9000$ и $10\,000$. Сравните полученный результат с соответствующими значениями $\pi(x) - x/\ln x$. Какой можно сделать вывод?

12. Мы видели, что нечетное простое число может быть записано или как $4n+1$, или как $4n+3$. Более того, как следует из упражнения 7, простых чисел вида $4n+3$ бесконечно много. Верно также и то, что остальных простых чисел тоже бесконечно много, хотя доказать это сложнее (см. комментарий перед упражнением 3). Цель этого упражнения — экспериментально установить, какой из этих типов простых чисел чаще встречается. Пусть x — положительное вещественное число, $\pi_1(x)$ — число положительных простых чисел вида $4n+1$, не превосходящих x, а $\pi_3(x)$ — аналогичное число простых вида $4n+3$. Исходя из решета Эратосфена, напишите программу, вычисляющую $\pi_1(x)$ и $\pi_3(x)$ по данному натуральному x. Используйте ее для определения $\pi_1(x)$, $\pi_3(x)$ и $\frac{\pi_1(x)}{\pi_3(x)}$ для $x = 100k$, где $1 \leqslant k \leqslant 10^5$. Известно, что $\lim\limits_{x\to\infty} \frac{\pi_1(x)}{\pi_3(x)} = 1$. Подтверждают ли ваши данные этот результат?

13. Адаптируйте программу из упражнения 12 для определения наименьшего числа x, при котором $\pi_1(x) > \pi_3(x)$.

Численные данные, доступные в начале столетия, спровоцировали некоторых математиков сделать вывод о том, что за исключением малых значений x неравенство $\pi_1(x) < \pi_3(x)$ должно быть всегда верным. Истина выяснилась в 1914 году, когда Дж. Е. Литлвуд (J. E. Littlewood) показал, что существуют такие бесконечные последовательности $x_1, x_2, \ldots$ и $y_1, y_2, \ldots$ положительных вещественных чисел, для которых

$$\lim_{i\to\infty}(\pi_1(x_i)-\pi_3(x_i)) = \infty \qquad \text{и} \qquad \lim_{i\to\infty}(\pi_1(y_i)-\pi_3(y_i)) = -\infty.$$

Мораль очевидна: рискованно делать обобщения на основании численных данных.

Глава 5.

Арифметика остатков

Большинство алгоритмов, описанных в предыдущих главах, проверяют делимость чисел непосредственным делением, убеждаясь, что в остатке действительно получается 0. Однако сейчас изобретены более эффективные методы, одним из которых (в главе 10) доказано, что $5 \cdot 2^{23\,473} + 1$ — множитель числа $F(23\,471)$. Так как эти числа слишком большие, мы полагаем, что даже проверка справедливости этого утверждения непосредственным делением займет очень много времени. Можно ли оценить насколько велико число $F(23\,471)$? Пользуясь логарифмами, легко показать, что в нем более, чем 10^{7063} знаков! То есть количество знаков в $F(23\,471)$ больше числа элементарных частиц в видимой части вселенной. Не приходится и говорить о проверке непосредственным делением того факта, что $5 \cdot 2^{23\,473} + 1$ — множитель числа $F(23\,471)$. Как же тогда это сделать?

Выход из этой дилеммы заключается в использовании *арифметики остатков*, являющейся темой данной главы. При решении вопросов делимости техника арифметики остатков играет основную роль и, как мы увидим в главе 8, она полезна также при вычислениях, имеющих отношение к феномену периодичности.

Основные идеи арифметики остатков были известны очень давно, но систематическую разработку ее аппарата впервые осуществил Гаусс в начале своих «Арифметических исследований»(см. [17]). В наше время к арифметике остатков обычно подходят с точки зрения *отношения эквивалентности*, которое мы подробно рассматриваем в первом параграфе.

§ 5.1. Отношение эквивалентности

Арифметику остатков лучше всего вводить с помощью отношения эквивалентности. Поскольку такие отношения будут играть важную роль как в этой главе, так и далее, стоит подробно разобрать это базисное понятие.

Пусть X — конечное или бесконечное множество. *Отношением* на X называется правило, по которому «сравниваются» его элементы. Это неформальное определение, но его вполне достаточно для наших целей. Заметим, что для определения отношения мы должны четко задать само множество; другими словами, нам должно быть ясно, какие элементы нужно сравнивать.

Рассмотрим несколько примеров. На множестве целых чисел есть много простых отношений, вроде «равно», «не равно», «меньше, чем», «меньше или равно». На множестве цветных мячей у нас есть отношение «*тот же цвет*». Последний пример, ввиду своей конкретности, хорош для запоминания в качестве модельного случая. Кстати, мы предполагаем, что каждый мяч из множества окрашен только в один цвет, пестрые мячи мы не рассматриваем.

Отношение эквивалентности — это отношение весьма специфичного вида. Возвращаясь к общим определениям, предположим, что X — множество, в котором было определено отношение. Удобно зафиксировать какой-нибудь символ для обозначения эквивалентности, обычно употребляют значок «$\sim$». С этого момента «$\sim$» будет *отношением эквивалентности*,

если для всех $x, y, z \in X$ выполнены следующие свойства:

(1) $x \sim x$;
(2) если $x \sim y$, то $y \sim x$;
(3) если $x \sim y$ и $y \sim z$, то $x \sim z$.

Первое свойство называется *рефлексивностью*. Оно говорит, что когда мы имеем отношение эквивалентности, любой элемент эквивалентен сам себе. Это свойство верно для равенства целых чисел: любое целое число равно самому себе. Но оно не выполнено для отношения «$<$». Поэтому «$<$» на множестве $\mathbb{Z}$ не является отношением эквивалентности.

Второе свойство называется *симметричностью*. Отношение «$<$» на множестве целых чисел не симметрично. Действительно, $2 < 3$, в то время как неравенство $3 < 2$ ложно. С другой стороны, отношение «$\leqslant$» на $\mathbb{Z}$ рефлексивно, но не симметрично.

Третье — свойство *транзитивности*. На множестве целых чисел отношения «равно», «меньше, чем», «меньше или равно», — транзитивны. А вот «не равно» этим свойством не обладает. Действительно, $2 \neq 3$ и $3 \neq 2$, но из этих неравенств не следует $2 \neq 2$. Добавим, что «$\neq$» симметрично, но не рефлексивно.

Мы предусмотрительно привели примеры отношений, которые не удовлетворяют этим свойствам, потому что это единственный путь к пониманию их действительного смысла. Именно владение примерами и контрпримерами обеспечивает успех в усвоении новых понятий. В примерах отношения эквивалентности нет недостатка. Равенство целых чисел, очевидно, удовлетворяет всем свойствам, выписанным выше. Отношение «тот же цвет» на множестве цветных мячей — еще один простой и, пожалуй, самый яркий пример. Среди примеров отношения эквивалентности на множестве многоугольников находятся такие отношения, как «одинаковое число сторон» и «одна и та же площадь».

Отношение эквивалентности используют для классификации элементов данного множества, группируя их в подмножества по принципу схожести свойств. Естественное разбиение множества, индуцированное отношением эквивалентности, называется разбиением на классы эквивалентности. Пусть на множестве X задано отношение эквивалентности $\sim$, и x — элемент этого множества. *Классом эквивалентности* элемента x называется подмножество в X, состоящее из всех элементов, эквивалентных x относительно $\sim$. Обозначив класс эквивалентности элемента x символом $\overline{x}$, можно записать:

$$\overline{x} = \{y \in X : \ y \sim x\}.$$

Приведем простой пример. Обозначим символом $\mathcal{M}$ множество цветных мячей с отношением эквивалентности «тот же цвет». Класс эквивалентности красного мяча в $\mathcal{M}$ состоит из *всех* красных мячей, содержащихся в $\mathcal{M}$.

Одно из свойств классов эквивалентности настолько важно, что мы назовем его *основным принципом* классов эквивалентности. Принцип гласит, что *любой элемент класса эквивалентности — хороший представитель всего класса*. Иначе говоря, зная один элемент из класса эквивалентности, можно немедленно восстановить этот класс полностью. Этот факт бросается в глаза, когда мы имеем дело с множеством $\mathcal{M}$ цветных мячей и отношением «тот же цвет». Предположим, Вам говорят, что в картонной коробке находятся все элементы одного класса эквивалентности множества $\mathcal{M}$. Увидев один элемент из этого множества (допустим, это синий мяч), Вы немедленно заключаете, что в коробке лежит класс эквивалентности всех синих мячей $\mathcal{M}$. Проще и быть не может!

Вернемся к абстрактному множеству X с отношением эквивалентности $\sim$. Основной принцип говорит, что если y — элемент из класса эквивалентности x, то классы эквивалентности x и y совпадают. То же самое можно выразить короче:

$$\text{если} \quad x \in X \quad \text{и} \quad y \in \overline{x}, \quad \text{то} \quad \overline{x} = \overline{y}.$$

Докажем это непосредственно из определяющих свойств отношения эквивалентности. Если $y \in \overline{x}$, то, по определению класса эквивалентности, $y \sim x$. Ввиду симметричности, $x \sim y$. Но если $z \in \overline{x}$, то и $z \sim x$. Тогда свойство транзитивности влечет $z \sim y$, т.е. $z \in \overline{y}$. Мы доказали включение: $\overline{x} \subseteq \overline{y}$. Похожее рассуждение доказывает обратное включение: $\overline{y} \subseteq \overline{x}$. Вероятно, это все может показаться несколько педантичным. Но основной принцип — такой источник неразберихи и ошибок, что нам не стоит жалеть усилий на прояснение его точного смысла. Кроме того, полезно осознать, что он непосредственно следует из определения отношения эквивалентности. Кстати о педантичности: вы поняли, что свойство $x \in \overline{x}$ вытекает из рефлексивности?

Основной принцип приводит к важнейшему свойству отношения эквивалентности. Как и раньше, пусть X — множество с отношением эквивалентности $\sim$; тогда

(1) X — объединение своих классов эквивалентности относительно $\sim$; и

(2) два разных класса эквивалентности не могут иметь общего элемента.

Первое утверждение следует из часто упоминаемого факта: класс эквивалентности элемента x содержит сам этот элемент. Для доказательства второго предположим, что элементы $x, y, z \in X$ и $z \in \overline{x} \cap \overline{y}$. Так как $z \in \overline{x}$, то по основному принципу $\overline{z} = \overline{x}$. Аналогично $\overline{z} = \overline{y}$. Так что $\overline{x} = \overline{y}$. Заметим, что свойства (1) и (2) означают, что множество X разбито на непересекающиеся подмножества, классы эквивалентности. Другими словами, мы имеем дело с *разбиением* множества X.

Множество, составленное из классов эквивалентности множества X относительно отношения эквивалентности $\sim$, имеет специальное название: *фактормножество* X по отношению $\sim$. Отметим, что элементы фактормножества — это подмножества в X. Поэтому фактормножество *не* является подмножеством в X, будьте внимательны!

Закончим этот параграф примером, в котором проявляется наконец истинная природа дробей. Из чего состоит дробь? Когда Вы на нее смотрите, то видите два числа, одно из которых (знаменатель) должно быть ненулевым. Конечно, Вы ее, вероятно, воспринимаете как частное. Но если на Вас надавить, Вы можете попытаться выбрать более легкий выход и сказать, что дробь в действительности — пара чисел, одно из которых не равно нулю. Однако, такое определение некорректно.

В математике две пары равны, если они имеют одинаковые первый и второй элементы. Так, пары (2,4) и (1,2) неравны. Но дроби $2/4$ и $1/2$ равны; так что дроби — не пары чисел.

Что же такое дроби? Это элементы фактормножества! Рассмотрим множество $\mathcal{Q}$ пар целых (a, b) с $b \neq 0$. На стандартном жаргоне $\mathcal{Q} = \mathbb{Z} \times (\mathbb{Z} \setminus \{0\})$. Две пары (a, b) и (a', b') целых чисел можно теперь называть эквивалентными, если $ab' = a'b$. Легко проверить, что это отношение эквивалентности, а дробь — класс эквивалентности множества $\mathcal{Q}$ относительно этого отношения. Следовательно, a/b означает не пару (a, b), а бесконечное множество всех пар из $\mathcal{Q}$, эквивалентных (a, b). Итак, множество $\mathbb{Q}$ рациональных чисел — это фактормножество множества $\mathcal{Q}$ по только что определенному отношению эквивалентности.

Представьте себе на минуту, что Вы до сих пор ничего о дробях не слышали и Вам придется исходить из описания, сделанного выше. Если Вам теперь скажут, что нужно вычислять с дробями, Вы почувствуете, что имеете вескую причину для паники: Вы же только что выучили, что дробь — это бесконечное множество. Мысль о прибавлении к одному бесконечному множеству другого бесконечного множества внушает легкое беспокойство. Именно в этот момент приходит на помощь основной принцип. Вам не нужно заботиться о бремени всего бесконечного множества; нужно знать только один элемент из него. Этот элемент расскажет Вам обо всем, что не-

обходимо знать о целом классе эквивалентности. Более того, Вас устроит любой элемент класса.

Итак, Вы можете оперировать с $1/2$ как обычно, так же, как если бы это была пара чисел. Вы вспоминаете, что дробь — это класс эквивалентности, только когда (в процессе вычислений) оказывается, что дробь можно сократить. В этот момент вы заменяете одного представителя класса эквивалентности на другой для упрощения вычислений.

Зачем мы сделали такое длинное отступление о дробях? В следующем параграфе определятся отношение эквивалентности на множестве $\mathbb{Z}$, а фактормножество этого отношения играет абсолютно фундаментальную роль в этой книге. Как и в случае дробей, классы эквивалентности будут бесконечны, а нам предстоит делать вычисления с ними. Но теперь Вы знаете, что нет причин для волнения.

§ 5.2. Сравнения

Проанализируем хорошо знакомые 24-часовые часы в свете наработок предыдущего параграфа. Когда некто говорит «один час», мы не можем понять, подразумевает ли говорящий это время сегодня, вчера, или завтра. Поэтому «один час» это не момент времени, а класс эквивалентности таких моментов. Объясним более подробно. Сначала разделим временной континуум на равные интервалы, скажем *часы*. После этого определим отношение эквивалентности: два момента, отличающиеся на эти 24 равных интервала, эквивалентны. Теперь один час — это класс эквивалентности моментов по специальному отношению эквивалентности. С одной стороны, такой подход усложняет очевидное, но с другой, крайне полезно попрактиковаться с феноменом цикличности.

Теперь мы будем рассматривать похожее отношение эквивалентности, определенное на множестве целых чисел. Выберем натуральное число n, которое с этого момента будет фик-

сировано. Оно называется периодом, или *модулем* отношения, которое мы собираемся определить.

Построим отношение эквивалентности на множестве $\mathbb{Z}$, декларируя, что каждое n-ое число (начиная с 0) входит в один и тот же класс эквивалентности. Иначе говоря, любые два целых числа, отличающиеся друг от друга на кратное n, эквивалентны. Более формально, два целых числа a и b *сравнимы по модулю* n, если $a - b$ делится на n; в таком случае мы пишем:

$$a \equiv b \pmod{n}.$$

Приведем несколько числовых примеров. Для модуля $n = 5$

$$10 \equiv 0 \pmod{5} \qquad \text{и} \qquad 14 \equiv 24 \pmod{5}.$$

Выберем другой модуль, скажем $n = 7$; в этом случае

$$10 \equiv 3 \pmod{7} \qquad \text{и} \qquad 14 \equiv 0 \pmod{7}.$$

Заметим, что числа, сравнимые по некоторому модулю, не обязательно сравнимы по другому модулю. Так, 21 сравнимо с 1 по модулю 5; но эти числа *не* сравнимы по модулю 7, потому что разность $21 - 1 = 20$ не кратна 7.

Теперь нужно проверить, что сравнимость по модулю n является отношением эквивалентности. Начнем с рефлексивности. Для ее проверки мы должны показать, что любое целое число a удовлетворяет сравнению $a \equiv a \pmod{n}$. Это будет так, если $a - a$ кратно n. Но $a - a = 0$ кратно любому целому числу. Значит, сравнение по модулю n рефлексивно.

Следующим идет свойство симметричности. Пусть для некоторых целых a и b справедливо сравнение $a \equiv b \pmod{n}$. Это означает, что $a - b = kn$ для какого-то целого k. Умножая это равенство на -1, получаем

$$b - a = -(a - b) = (-k)n,$$

что также кратно n. Поэтому $b \equiv a \pmod{n}$, и мы доказали, что сравнение по модулю n симметрично.

Наконец, транзитивность. Предположим, что $a \equiv b \pmod{n}$ и $b \equiv c \pmod{n}$, где a, b и c — целые числа. По определению, эти сравнения говорят, что разности $a - b$ и $b - c$ кратны n. Но при сложении кратных n получается число, делящееся на n. Поэтому $(a - b) + (b - c) = (a - c)$ кратно n. Другими словами, $a \equiv c \pmod{n}$, как мы и хотели показать. Убедившись в справедливости этих трех свойств для сравнения по модулю n, мы делаем вывод, что оно — отношение эквивалентности.

Множество, которое будет притягивать бóльшую часть нашего внимания на протяжении этой главы — это фактормножество $\mathbb{Z}$ по отношению сравнения по модулю n. Оно называется *множеством вычетов по модулю* n и обозначается символом $\mathbb{Z}_n$. Из определения фактормножества следует, что элементы $\mathbb{Z}_n$ — подмножества в $\mathbb{Z}$, т.е. классы эквивалентности в $\mathbb{Z}$ сравнимых чисел по модулю n. Мы хотим отождествить между собой элементы этих классов.

Пусть $a \in \mathbb{Z}$. Класс a образован всеми целыми b, для которых $b - a$ кратно n, т.е. $b - a = kn$ для некоторого $k \in \mathbb{Z}$. Таким образом, класс эквивалентности числа a описывается формулой:

$$\overline{a} = \{a + kn | \, k \in \mathbb{Z}\}.$$

Заметим, что $\overline{0}$ состоит из всех чисел, кратных n, и каждый класс эквивалентности — бесконечное множество.

Мы видели, что в $\overline{a}$ бесконечно много элементов, каждый из которых хорошо представляет класс целиком. Отсюда резонный вопрос: можно ли простым способом найти наименьшее натуральное число, представляющее $\overline{a}$? Ответ — да. Нам нужно только поделить a на n. Обозначим через r остаток, а через q неполное частное этого деления; тогда

$$a = nq + r \qquad \text{и} \qquad 0 \leqslant r \leqslant n - 1.$$

Следовательно, разность $a - r = nq$ кратна n. Поэтому $a \equiv r \pmod{n}$. Число r называется *вычетом* a по модулю n.

На самом деле мы доказали больше, чем обещали. Действительно, мы показали, что любое целое число сравнимо с одним из целых, лежащим между 0 и $n-1$. В частности, фактормножество $\mathbb{Z}_n$ имеет не более n классов $\overline{0}, \ldots, \overline{n-1}$. Для уверенности в том, что оно имеет ровно n различных классов сопряженности по модулю n, нам нужно проверить, что никакие два из выписанных классов не могут быть равными. Каждый класс сопряженности представляется неотрицательным целым числом меньшим n. Если бы классы совпадали, их представители были бы сравнимы по модулю n, т.е. разность двух разных неотрицательных целых чисел, меньших n, делилась бы на n. Этого не может быть. Так что классы $\overline{0}, \ldots, \overline{n-1}$ на самом деле различны. Итак,

$$\mathbb{Z}_n = \left\{\overline{0}, \overline{1}, \ldots, \overline{n-1}\right\}.$$

Говорят, что класс эквивалентности $\overline{a}$ записан в *приведенном виде*, если $0 \leqslant a \leqslant n-1$. Как и в случае дробей, класс сопряженности удобно представлять именно в таком виде. Этому есть две причины. Первая: меньший представитель класса выбрать легче, чем бóльший. Вторая: если два класса записаны в приведенном виде, то очень легко определить, равны они или нет (они будут равны тогда и только тогда, когда их представители совпадают). Разумеется, утверждение неверно, если классы записаны не в приведенном виде.

Это все очень хорошо, но кажется довольно абстрактным. Хорошо бы иметь наглядную геометрическую картинку, изображающую множество $\mathbb{Z}_n$. Вспомним геометрическое изображение множества $\mathbb{Z}$. Большинство людей представляют себе целые числа как точки, равномерно расположенные вдоль прямой линии. Где-то на этой прямой стоит точка, изображающая 0, так что отрицательные целые числа находятся слева от нее, а положительные — справа. Сравнение по модулю n отождествляет между собой все числа, отличающиеся на kn ($k \in \mathbb{Z}$), и поэтому, подходя к n, мы должны вернуться опять к 0.

Если прямую целых чисел представить себе как гибкую проволоку, то можно взять точку, помеченную n, и склеить ее с 0, что даст окружность. Продолжая наматывать прямую целых чисел на получившуюся окружность, Вы увидите, что числа, сравнимые по модулю n, попадут в одни и те же точки на окружности. В связи с этим $\mathbb{Z}_n$ можно представлять себе как окружность, на которой через равные промежутки отмечены n классов сопряженности.

§ 5.3. Арифметика остатков

Геометрическая картинка, приведенная в конце предыдущего параграфа, помогает нам дать простое описание сложения в $\mathbb{Z}_n$. Представьте себе n классов эквивалентности $\mathbb{Z}_n$ цифрами, отмеченными на циферблате часов. Предположим, что $\overline{0}$ находится в верхней точке круга (что соответствует 12-ти часам), в то время как остальные классы распределены вдоль границы циферблата через равные промежутки. Пусть эти часы имеют только одну стрелку, закрепленную в центре циферблата, и мы можем указывать ею на наши классы.

Хотелось бы превратить эти «часы» в прибор для вычисления сумм в $\mathbb{Z}_n$. Бесспорно, «часы» для арифметики остатков — это то же самое, что счет на пальцах для нормальной арифметики. Последний, между прочим, имеет длинную и выдающуюся историю. В средние века ему обучали в монастырских школах, а одна из первых английских книг по арифметике, «Основа наук» Роберта Рекорда (Robert Recorde), содержала целый параграф, объясняющий «искусство счета руками». Так что мы в хорошей компании.

Итак, предположим, что нужно сложить $\overline{a}$ и $\overline{b}$, два класса из $\mathbb{Z}_n$. Будем предполагать, что оба класса записаны в приведенной форме, т.е. a и b неотрицательны и меньше n. Процесс вычисления $\overline{a}+\overline{b}$ следующий. Поместим стрелку «часов» в точ-

ку, помеченную $\overline{a}$, затем передвинем ее по часовой стрелке на b мест. Теперь стрелка будет показывать сумму $\overline{a}+\overline{b}$.

Численный пример. Предположим, что Вам нужно к $\overline{4}$ прибавить $\overline{5}$ в $\mathbb{Z}_8$. Поместите стрелку «часов» на $\overline{5}$ и передвиньте ее вперед на 4 места. Когда Вы будете это делать, стрелка минует $\overline{0}$ и остановится на $\overline{1}$. Значит $\overline{4}+\overline{5}=\overline{1}$ в $\mathbb{Z}_8$.

К сожалению, наш прибор будет слишком медленно работать при больших n (как и счет на пальцах, который трудноосуществим при слишком больших числах). Поэтому требуется более математизированный способ вычисления сумм в $\mathbb{Z}_n$. Реально, он достаточно прост. Пусть $\overline{a}$ и $\overline{b}$ — классы в $\mathbb{Z}_n$, которые нам предстоит сложить. Операция определяется формулой:

$$\overline{a}+\overline{b}=\overline{a+b}.$$

Интерпретация этой формулы требует некоторой осторожности. Слева в ней стоит сумма двух классов $\mathbb{Z}_n$, а справа мы имеем класс, соответствующий сумме двух целых чисел. Итак, сложение классов определяется в терминах операции, которую мы хорошо знаем: сложение целых чисел.

Вернемся к примеру сложения в $\mathbb{Z}_8$, которое мы выполнили с помощью прибора. Мы хотим к $\overline{4}$ прибавить $\overline{5}$. Учитывая формулу, сначала сложим 4 и 5; их сумма, очевидно, равна 9. Отсюда следует, что $\overline{4}+\overline{5}=\overline{9}$. На первый взгляд кажется, что новый результат суммирования отличается от полученного с помощью прибора. Но не забывайте, что $9-1=8$, т.е. $\overline{9}=\overline{1}$.

Последний пример указывает на одну важную проблему. Мы видели, что класс эквивалентности может быть представлен любым своим элементом; это основной принцип из §5.1. Но складывая два класса, мы сначала суммируем их представителей, а затем берем соответствующий класс. Как же мы можем быть уверенными в том, что при выборе каких-то других представителей, результирующий класс останется прежним? Для уверенности в том, что Вы уловили суть, рассмотрим еще

раз сумму $\overline{5}$ и $\overline{4}$ в $\mathbb{Z}_8$. Следуя сформулированному выше правилу, мы нашли, что их сумма равна $\overline{9}$. Однако, $\overline{13} = \overline{5}$ и $\overline{12} = \overline{4}$, и формула говорит нам, что если сложить $\overline{13}$ и $\overline{12}$, то получится $\overline{25}$. Так, поначалу может показаться, что выбирая различные представители классов, мы получаем разные суммы. Но это только кажется. На основании того, что $25 - 9 = 16$ делится на 8, мы заключаем: $\overline{25} = \overline{9}$ в $\mathbb{Z}_8$.

Один из путей решения проблемы мог бы состоять в записи классов в приведенном виде перед их сложением. Это неудобно и совсем необязательно. Как подсказывает разобранный выше пример, результат суммирования не зависит от выбора представителей классов. Это очень важно и должно быть проверено в деталях. Пусть $\overline{a}$ и $\overline{b}$ — два класса в $\mathbb{Z}_n$. Пусть $\overline{a} = \overline{a'}$ и $\overline{b} = \overline{b'}$. Мы хотим показать, что $\overline{a+b} = \overline{a'+b'}$. Но равенство $\overline{a} = \overline{a'}$ означает, что $a - a'$ кратно n и то же самое верно для $b - b'$. Сумма кратных n снова кратна n, откуда число

$$(a - a') + (b - b') = (a + b) - (a' + b')$$

должно быть кратно n. Следовательно, $\overline{a+b} = \overline{a'+b'}$, что и требовалось доказать.

Вычитание классов определяется аналогичным образом и совсем нетрудно. Посмотрим, как должно определяться умножение. Пусть $\overline{a}$ и $\overline{b}$ — классы в $\mathbb{Z}_n$. Определение сложения классов советует написать

$$\overline{a} \cdot \overline{b} = \overline{ab}.$$

Как и в случае суммы, мы должны быть уверены, что определение умножения не зависит от выбора представителей классов. Итак, предположим, что $\overline{a} = \overline{a'}$ и $\overline{b} = \overline{b'}$. Нужно проверить, что $\overline{ab} = \overline{a'b'}$. Равенство $\overline{a} = \overline{a'}$ влечет: $a - a'$ кратно n; скажем, $a = a' + rn$ для некоторого целого r. Аналогично $b = b' + sn$ для некоторого целого s. Умножая a на b, получим

$$ab = (a' + rn)(b' + sn) = a'b' + (a's + rb' + srn)n.$$

Значит $ab - a'b'$ кратно n и $\overline{ab} = \overline{a'b'}$.

Зная теперь, как складываются и перемножаются классы, любопытно выяснить, так же ли ведут себя эти операции, как их тезки в $\mathbb{Z}$. Пусть $\overline{a}$, $\overline{b}$ и $\overline{c}$ — классы в $\mathbb{Z}_n$. Сложение классов обладает следующими свойствами:

$$(\overline{a}+\overline{b})+\overline{c}=\overline{a}+(\overline{b}+\overline{c});$$

$$\overline{a}+\overline{b}=\overline{b}+\overline{a};$$

$$\overline{a}+\overline{0}=\overline{a};$$

$$\overline{a}+\overline{-a}=\overline{0}.$$

Элемент $\overline{-a}$ называется *противоположным* $\overline{a}$. Заметим, что если $\overline{a}$ записан в приведенном виде, то приведенным видом класса $\overline{-a}$ будет $\overline{n-a}$. Умножение классов обладает следующими свойствами:

$$(\overline{a}\cdot\overline{b})\cdot\overline{c}=\overline{a}\cdot(\overline{b}\cdot\overline{c});$$

$$\overline{a}\cdot\overline{b}=\overline{b}\cdot\overline{a};$$

$$\overline{a}\cdot\overline{1}=\overline{a}.$$

Итак, каждое свойство умножения соответствует свойству сложения, за одним исключением: существование противоположного элемента. Мы вернемся к этому вопросу в § 5.7, где обсудим деление классов.

Есть еще свойство *дистрибутивности*

$$\overline{a}\cdot(\overline{b}+\overline{c})=\overline{a}\cdot\overline{b}+\overline{a}\cdot\overline{c}.$$

Эти свойства довольно легко следуют из их аналогов для сложения и умножения целых, поэтому мы опустим доказательства. Последовательному читателю не составит труда найти их самостоятельно.

Ну хорошо. Операции в $\mathbb{Z}_m$, очевидно, ведут себя так же, как их двойники в $\mathbb{Z}$. Однако слепо полагаться на эту аналогию довольно опасно, поскольку может сформироваться ложное чувство уверенности, приводящее к ошибкам. Действительно, одно ключевое свойство целых чисел не выполняется для сравнений по модулю n. Следующий пример проясняет этот факт. Рассмотрим классы $\overline{2}$ и $\overline{3}$ в $\mathbb{Z}_6$. Они оба отличны от нуля, однако

$$\overline{2} \cdot \overline{3} = \overline{6} = \overline{0},$$

т.е. произведение двух ненулевых элементов из $\mathbb{Z}_6$ может равняться нулю. Такого, конечно, не бывает в $\mathbb{Z}$.

В качестве важного следствия этого примера получаем, что в $\mathbb{Z}_n$ не всегда можно сокращать на ненулевой элемент. Другими словами, если $\overline{a} \neq \overline{0}$, то *не* всегда справедливо рассуждение:

$$\overline{a} \cdot \overline{b} = \overline{a} \cdot \overline{c} \implies \overline{b} = \overline{c}.$$

Так, например, $\overline{2} \cdot \overline{3} = \overline{2} \cdot \overline{0}$, но $\overline{3} \neq \overline{0}$. Мы вернемся к этому вопросу в § 5.7, рассмотрев сначала некоторые приложения арифметики остатков.

§ 5.4. Критерий делимости

Большинство людей помнят из начальной школы, что число делится на 3, если сумма цифр в его десятичной записи делится на 3. Но почему это верно? Мы легко можем это доказать, используя сравнения по модулю 3. Напомним, что число делится на 3 тогда и только тогда, когда оно сравнимо с 0 по модулю 3. Значит, на языке арифметики остатков критерий делимости на 3 утверждает, что число сравнимо с 0 по модулю 3, если и только если то же самое справедливо для суммы его цифр. Последнее утверждение мы сейчас и докажем.

Пусть a — целое число и $a_0, a_1, \ldots, a_n$ — его цифры в десятичной записи. Иначе говоря,

$$a = a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_1 10 + a_0,$$

где $0 \leqslant a_i \leqslant 9$ для $i = 0, \ldots, n$. В предыдущем параграфе мы показали, что умножение не зависит от выбора представителей классов по модулю n. Поскольку $10 \equiv 1 \pmod 3$, независимость от выбора говорит нам, что 10^k сравнимо с 1^k по модулю 3 для любого положительного целого k. Другими словами, любая степень 10 имеет вычет 1 по модулю 3. Значит,

$$a \equiv a_n + a_{n-1} + \cdots + a_1 + a_0 \pmod 3.$$

Отсюда немедленно следует, что $a \equiv 0 \pmod 3$ тогда и только тогда, когда $a_n + a_{n-1} + \cdots + a_1 + a_0 \equiv 0 \pmod 3$. А это как раз то, что нужно было доказать.

Заметим, что все проделанные вычисления останутся верными после замены 3 на 9, потому что $10 \equiv 1 \pmod 9$. Значит, целое число делится на 9, если и только если на 9 делится сумма его цифр в десятичной записи.

Применим похожие аргументы к другому числу, например, к 11. Снова будем предполагать, что

$$a = a_n 10^n + a_{n-1} 10^{n-1} + \cdots + a_1 10 + a_0,$$

где $a_n, a_{n-1}, \ldots, a_1, a_0$ — цифры записи числа a. Поскольку $10 \equiv -1 \pmod{11}$, то

$$10^k \equiv -1^k \pmod{11}$$

будет либо 1 (если k четно), либо -1 (если k нечетно). Поэтому

$$a \equiv a_n(-1)^n + a_{n-1}(-1)^{n-1} + \cdots + a_2 - a_1 + a_0 \pmod{11}.$$

Говоря человеческим языком, число делится на 11 тогда и только тогда, когда на 11 делится альтернированная сумма его

цифр. Например, 3 443 делится на 11, потому что $3-4+4-3=0$ делится на 11.

Критерии делимости на 2 и 5 слишком очевидны, чтобы приводить их доказательства. Таким образом мы нашли простые критерии делимости на все простые числа от 2 до 11, исключая 7. Разберемся, что произойдет в случае применения того же подхода к 7.

Мы уже знаем из предыдущего примера, что та часть рассуждений, которая зависит от модуля, состоит в вычислении степеней 10. На этот раз $10 \equiv 3 \pmod 7$, а степени 3 не так легко вычисляются, как степени 1 или -1. Попытаемся найти эти степени при малых показателях. Все сравнения, приведенные ниже, сделаны по модулю 7.

$$\begin{aligned}
10^2 &\equiv 3^2 \equiv 2\\
10^3 &\equiv 10\cdot 10^2 \equiv 3\cdot 2 \equiv 6 \equiv -1\\
10^4 &\equiv 10\cdot 10^3 \equiv (-1)\cdot 3 \equiv 4\\
10^5 &\equiv 10\cdot 10^4 \equiv 3\cdot 4 \equiv 5\\
10^6 &\equiv 10\cdot 10^5 \equiv 3\cdot 5 \equiv 1
\end{aligned}$$

Заметим, что последний вычет равен $10^0 = 1$. Это означает, что вычеты будут циклически, с периодом 6, повторяться. Эти вычисления показывают, что критерий делимости на 7 несколько более сложен для запоминания, чем на 3 и 11. Поскольку мы уже довольно далеко продвинулись в работе над этим критерием, точно сформулируем его в простом случае. Предположим, что $a = a_2 10^2 + a_1 10 + a_0$, где $0 \leqslant a_0, a_1, a_2 \leqslant 9$. Используя вычеты степеней 10, вычисленные только что, мы имеем

$$a \equiv a_2 10^2 + a_1 10 + a_0 \equiv 2a_2 + 3a_1 + a_0 \pmod 7.$$

Таким образом, a делится на 7 тогда и только тогда, когда на 7 делится выражение $2a_2 + 3a_1 + a_0$. Например, 231 делится на 7 ввиду делимости $2\cdot 2 + 3\cdot 3 + 1 = 14$ на 7.

§ 5.5. Степени

Во многих приложениях мы будем встречаться со следующей задачей: пусть a, k и n — натуральные числа; найти остаток от деления a^k на n. Если k очень большое, то может оказаться невозможным даже пересчитать знаки в a^k, как в примере из начала этой главы. Однако мы можем упростить эту проблему, используя арифметику остатков.

Начнем с простого примера. Предположим, что мы хотим найти остаток от деления 10^{135} на 7. Мы видели в предыдущем параграфе, что $10^6 \equiv 1 \pmod 7$. Деля 135 на 6, мы найдем $135 = 6 \cdot 22 + 3$. Полученное равенство дает следующие сравнения по модулю 7:

$$10^{135} \equiv (10^6)^{22} \cdot 10^3 \equiv (1)^{22} \cdot 10^3 \equiv 6.$$

Значит остаток от деления 10^{135} на 7 равен 6.

Вычисления не всегда оказываются такими легкими. Например, какой остаток получится при делении 3^{64} на 31? Вычислив несколько степеней 3 по модулю 31, мы быстро найдем, что $3^3 \equiv -4 \pmod{31}$. Вместо вычисления более высоких степеней в надежде, что одна из них станет снова 1, воспользуемся уже имеющейся информацией. Так как $64 = 3 \cdot 21 + 1$, мы получаем сравнение по модулю 31:

$$3^{64} \equiv (3^3)^{21} \cdot 3 \equiv (-4)^{21} \cdot 3 \equiv -(2)^{42} \cdot 3.$$

Мы еще не получили искомый остаток, а только степень 2, которая лежит между нами и нашей целью. Удачно, что $2^5 \equiv 1 \pmod{31}$. Так как $45 = 8 \cdot 5 + 2$, мы находим

$$3^{64} \equiv -(2)^{42} \cdot 3 \equiv -(2^5)^8 \cdot 2^2 \cdot 3 \equiv -12 \pmod{31}.$$

Но $-12 \equiv 19 \pmod{31}$, так что остаток от деления 3^{64} на 31 равен 19.

Были бы вычисления легче, продолжай мы находить степени 3 до тех пор, пока не встретится 1? Ответ — нет, в чем

Вы сами можете убедиться, попытавшись найти наименьшее натуральное r, для которого $3^r \equiv 1 \pmod{31}$.

Предположим теперь, что мы хотим найти остаток от деления 6^{35} на 16. В этом случае бесполезно пытаться отыскать наименьшую степень 6, сравнимую с 1 по модулю 16: ее просто нет. Действительно,

$$6^4 \equiv 2^4 \cdot 3^4 \equiv 0 \cdot 3^4 \equiv 0 \pmod{16},$$

откуда

$$6^{35} \equiv 6^4 \cdot 6^{31} \equiv 0 \pmod{16}.$$

Эти примеры иллюстрируют некоторые трюки, используемые для облегчения подсчета вычетов степеней по модулю n. Другие приемы появятся в последующих главах. Конечно, компьютеру не нужны никакие трюки. Это не говорит о том, что компьютер не использует арифметику остатков для таких вычислений; на самом деле, он ее использует. Быстрый алгоритм, вычисляющий степени по модулю n, можно найти в § П.2 приложения. Его можно использовать для доказательства делимости $F(23\,471)$ на $5 \cdot 2^{23\,473} + 1$ — достижение, которое раньше казалось невыполнимой задачей.

§ 5.6. Диофантовы уравнения

Ниже мы используем сравнения для доказательства отсутствия решений у некоторых Диофантовых уравнений. *Диофантово уравнение* — это полиномиальное уравнение с несколькими неизвестными и целыми коэффициентами. Примеры: $3x - 2y = 1$, $x^3 + y^3 = z^2$, $x^3 - 117y^3 = 5$. Когда говорят о решениях диофантова уравнения, обычно имеют в виду целочисленные решения. Эти уравнения названы по имени греческого математика Диофанта из Александрии, жившего около 250 г. н.э. В своей «Арифметике» Диофант подробно

обсуждает проблему поиска решений неопределенных уравнений. Однако он искал рациональные решения, а не целые, что мы обычно делаем сегодня.

Поскольку эти уравнения зависят от нескольких переменных, они могут иметь бесконечно много решений. Например, для любого целого k числа $x = 1+2k$ и $y = 1+3k$ удовлетворяют уравнению $3x - 2y = 1$. Уравнение $x^3 + y^3 = z^3$ — частный случай великой Теоремы Ферма, о которой рассказывалось во введении и конце второй главы. Как мы видели, если x, y и z — целые числа, удовлетворяющие этому уравнению, то одно из них должно быть равно нулю. Это специальный случай теоремы, впервые доказанный Эйлером в 1770 году. Вспомните, что Ферма сформулировал свою великую Теорему на полях принадлежащей ему копии Диофантовой «Арифметики».

Уравнение $x^3 - 117y^3 = 5$ имеет более недавнюю и простую историю. В статье, написанной в 1969 году, Льюис (D. J. Lewis) показал, что это уравнение не может иметь более 18 целочисленных решений. Два года спустя Р. Финкельштейн (R. Finkelstein) и Х. Лондон (H. London) наконец доказали, что это уравнение не имеет ни одного целого решения. Их доказательство коротко: оно занимает только страницу 111 четырнадцатого тома «Канадского Математического Бюллетеня», но отнюдь не элементарно. Однако, в 1973 году Халтер-Кох (F. Halter-Koch) и Удреско (V. St. Udresco) независимо дали доказательство отсутствия у этого уравнения решений, которое использует только сравнения по модулю 9. Именно его мы сейчас подробно опишем.

Доказательство «от противного». Предположим, что уравнение $x^3 - 117y^3 = 5$ имеет целочисленное решение, т.е. существуют такие целые x_0 и y_0, что $x_0^3 - 117y_0^3 = 5$. Так как все числа в этом выражении целые, мы можем рассмотреть его по модулю 9. Число 117 делится на 9, поэтому

$$x_0^3 \equiv x_0^3 - 117y_0^3 \equiv 5 \pmod 9.$$

Следовательно, если это уравнение имеет целое решение (x_0, y_0),

то $x_0^3 \equiv 5 \pmod 9$. Возможно ли это? Чтобы разобраться, напомним: каждое целое число по модулю 9 имеет вычет, лежащий между 0 и 8. Значит нам будет достаточно вычислить кубы по модулю 9 каждого из этих вычетов.

класс по модулю 9	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$	$\overline{4}$	$\overline{5}$	$\overline{6}$	$\overline{7}$	$\overline{8}$
куб по модулю 9	$\overline{0}$	$\overline{1}$	$\overline{8}$	$\overline{0}$	$\overline{1}$	$\overline{8}$	$\overline{0}$	$\overline{1}$	$\overline{8}$

Простой взгляд на таблицу показывает, что вычет куба целого числа по модулю 9 может быть равен только 0, 1 или 8. В частности, нет такого целого x_0, куб которого сравним с 5 по модулю 9: $x_0^3 \equiv 5 \pmod 9$. Следовательно, $x^3 - 117y^3 = 5$ не может иметь целочисленных решений, что и требовалось доказать.

Можно извлечь такую мораль из этого примера: *первое доказательство теоремы часто оказывается и не самым простым, и не самым элегантным*. Так бывает потому, что первое доказательство обычно находится исследователем, «вспахивающим целину» неизученного. Со временем связь между новыми методами и ближайшими областями становится яснее, что делает возможным находить более короткое, простое и прямое доказательство, чем первое. Пример, который мы разобрали, довольно наивен, но, конечно, он не лишает законной силы сделанного нами вывода. Как однажды сказал математик А. С. Безикович (A. S. Besicovitch), «репутация математика держится на том, сколько он дал некрасивых доказательств».

§ 5.7. Деление по модулю n

Настало время вернуться к проблеме деления классов в $\mathbb{Z}_n$. Но сначала рассмотрим тот же самый вопрос в более знакомой ситуации. Пусть a и b — вещественные числа. Один из способов деления a на b состоит в умножении a на $1/b$. Число $1/b$

называется обратным к b и однозначно определяется как решение уравнения $b \cdot x = 1$. С практической точки зрения этот способ не облегчает дела, потому что для вычисления $1/b$ мы все равно должны разделить 1 на b. Однако с теоретической точки зрения иногда бывает лучше рассуждать об обратном элементе, нежели о делении. Наконец, $1/b$ существует только если $b \neq 0$, поскольку уравнение $0 \cdot x = 1$ не имеет решений. Держа эти замечания в уме, вернемся к $\mathbb{Z}_n$.

Как всегда, n — фиксированное натуральное число. Предположим, что $\overline{a} \in \mathbb{Z}_n$. Назовем $\overline{\alpha} \in \mathbb{Z}_n$ *обратным* к $\overline{a}$, а сам элемент $\overline{a}$ *обратимым*, если в $\mathbb{Z}_n$ справедливо равенство: $\overline{a} \cdot \overline{\alpha} = \overline{1}$. Ясно, что $\overline{0}$ не имеет обратного в $\mathbb{Z}_n$. К сожалению, $\overline{0}$ может быть не единственным элементом в $\mathbb{Z}_n$, не имеющим обратного. Необходимо рассмотреть этот момент очень подробно.

Предположим, что $\overline{a} \in \mathbb{Z}_n$ имеет обратный элемент $\overline{\alpha}$, и посмотрим, что мы с этого будем иметь. Из уравнения

$$\overline{a} \cdot \overline{\alpha} = \overline{1}$$

следует, что $a\alpha - 1$ делится на n. Иначе говоря,

$$a\alpha + kn = 1 \tag{7.1}$$

для некоторого целого k. Равенство (7.1) влечет: НОД$(a, n) = 1$. Таким образом мы заключаем, что если $\overline{a}$ имеет обратный элемент в $\mathbb{Z}_n$, то НОД$(a, n) = 1$.

Верно ли обратное? Для ответа предположим, что a — такое целое число, для которого НОД$(a, n) = 1$. Равенство (7.1) подсказывает применить расширенный алгоритм Евклида к числам a и n. Этот алгоритм даст нам такие целые числа α и β, что

$$a\alpha + n\beta = 1.$$

Полученное уравнение эквивалентно тождеству

$$\overline{a} \cdot \overline{\alpha} = \overline{1}$$

в $\mathbb{Z}_n$. Значит класс $\overline{a}$, найденный с помощью расширенного алгоритма Евклида, является обратным к $\overline{a}$ в $\mathbb{Z}_n$. Итак, если НОД$(a, n) = 1$, то $\overline{a}$ обратим в $\mathbb{Z}_n$. Подведем итог в следующей теореме.

Теорема обратимости. *Класс $\overline{a}$ обратим в $\mathbb{Z}_n$ тогда и только тогда, когда целые a и n взаимно просты.*

Рассуждения, приведенные выше, являются конструктивным доказательством теоремы обратимости, в том смысле, что они представляют собой процедуру для проверки существования обратного элемента и его вычисления, если он есть. Конечно, эта процедура — непосредственное применение расширенного алгоритма Евклида. Например, обладает ли $\overline{3}$ обратным в $\mathbb{Z}_{32}$? И если «да», то чему он равен? Применяя расширенный алгоритм Евклида к числам 32 и 3, мы найдем, что НОД$(3, 32) = 1$, и

$$3 \cdot 11 - 32 = 1.$$

Так что обратный элемент существует. Переписывание этого уравнения по модулю 32 приводит к равенству: $\overline{3} \cdot \overline{11} = \overline{1}$. Итак, $\overline{11}$ — обратный элемент к $\overline{3}$ в $\mathbb{Z}_{32}$.

Множество обратимых элементов в $\mathbb{Z}_n$, обозначающееся символом $U(n)$, играет ключевую роль в главах 9, 10 и 11. По теореме обратимости мы имеем:

$$U(n) = \left\{ \overline{a} \in \mathbb{Z}_n | \text{ НОД}(a, n) = 1 \right\}.$$

Для простого числа p множество $U(p)$ вычислить очень легко, потому что в этом случае условие НОД(a, p) равносильно такому: p не делит a. Поскольку это справедливо для всех натуральных чисел, меньших p, имеем $U(p) = \mathbb{Z}_p \setminus \{\overline{0}\}$.

К сожалению, такое простое описание множества $U(p)$ верно только для простого p. Если n — составное и $1 < k < n$ — его множитель, то НОД$(k, n) = k \neq 1$, так что $\overline{k}$ не обра-

тим в $\mathbb{Z}_n$. Два простых примера.

$$U(4) = \left\{\overline{1}, \overline{3}\right\} \qquad \text{и} \qquad U(8) = \left\{\overline{1}, \overline{3}, \overline{5}, \overline{7}\right\}.$$

Ключевое свойство множества $U(n)$ состоит в том, что оно содержит произведение любых своих элементов. Более точно, если $\overline{a}$ и $\overline{b}$ — обратимые классы в $\mathbb{Z}_n$, то произведение $\overline{a} \cdot \overline{b}$ тоже обратимо в $\mathbb{Z}_n$. Это утверждение является очень важным для главы 9, так что проверим его детально. Пусть $\overline{\alpha}$ — обратный элемент для $\overline{a}$, а $\overline{\beta}$ — для $\overline{b}$ в $\mathbb{Z}_n$. Тогда обратным к произведению $\overline{a} \cdot \overline{b}$ будет $\overline{\alpha} \cdot \overline{\beta}$. Действительно,

$$(\overline{a} \cdot \overline{b})(\overline{\alpha} \cdot \overline{\beta}) = (\overline{a} \cdot \overline{\alpha})(\overline{b} \cdot \overline{\beta}) = \overline{1} \cdot \overline{1} = \overline{1}.$$

Мы еще будем периодически возвращаться к множеству $U(n)$ в последующих главах.

Вернемся к проблеме делимости $\overline{a}$ на $\overline{b}$ в $\mathbb{Z}_n$, с которой мы начинали этот параграф. Прежде всего нам нужно узнать, имеет ли $\overline{b}$ обратный элемент в $\mathbb{Z}_n$. Если имеет, то мы найдем его с помощью расширенного алгоритма Евклида. Пусть, например, это будет $\overline{\beta}$. Чтобы разделить $\overline{a}$ на $\overline{b}$, мы вычислим произведение $\overline{a} \cdot \overline{\beta}$. Разделим, в качестве примера, $\overline{2}$ на $\overline{3}$ в $\mathbb{Z}_8$. После применения алгоритма Евклида к 3 и 8 мы имеем: НОД$(3, 8) = 1$ и $\overline{3}$ сам себе обратен. Поэтому результат деления $\overline{2}$ на $\overline{3}$ в $\mathbb{Z}_8$ равен $\overline{6}$.

Применим результаты этого параграфа к решению линейных сравнений в $\mathbb{Z}_n$. *Линейное сравнение* — это уравнение вида

$$ax \equiv b \pmod{n}, \tag{7.2}$$

где $a, b \in \mathbb{Z}$. Если бы это было линейное уравнение над вещественными числами, нам следовало бы поделить его на a. Попытаемся использовать ту же идею здесь. Предполагая, что НОД$(n, a) = 1$, мы заключаем (по теореме обратимости), что

существует такой $\alpha \in \mathbb{Z}$, что $a\alpha \equiv 1 \pmod{n}$. Умножая обе стороны уравнения (7.2) на α, получаем:

$$x \equiv \alpha a x \equiv \alpha b \pmod{n},$$

и уравнение решено. Например, для решения сравнения $7x \equiv$ $\equiv 3 \pmod{15}$ мы сначала ищем обратный элемент к 7 по модулю 15. Так как $15 - 2 \cdot 7 = 1$, обратным к 7 по модулю 15 будет $-2 \equiv 13 \pmod{15}$. Умножая сравнение $7x \equiv 3 \pmod{15}$ на 13, получаем

$$x \equiv 13 \cdot 3 \equiv 39 \equiv 9 \pmod{15},$$

что является искомым решением.

Заметим, метод, применяемый к решению линейных сравнений, показывает, что в случае НОД$(a, n) = 1$ сравнение $ax \equiv$ $\equiv b \pmod{n}$ имеет одно и только одно решение по модулю n. Иначе говоря, несмотря на бесконечность числа целых, удовлетворяющих данному уравнению, все они сравнимы друг с другом по модулю n. Замечание может показаться банальным, но это не так. В самом деле, высказывание не верно, если НОД$(a, n) \neq 1$. Уравнение $2x \equiv 1 \pmod{8}$, например, не имеет решений вовсе. Мы вернемся к этому вопросу в начале главы 8.

Упражнения

1. Какие из отношений на множестве $\mathbb{Z}$, выписанных ниже, являются отношениями эквивалентности?

(1) $a \sim b$, когда НОД$(a, b) = 1$;

(2) для фиксированного целого $n > 0$ отношение определяется правилом: $a \sim b$ тогда и только тогда, когда НОД$(a, n) =$ НОД(b, n).

2. Найдите вычет числа a по модулю n, если

(1) $a = 2\,351$ и $n = 2$;
(2) $a = 50\,121$ и $n = 13$;
(3) $a = 321\,671$ и $n = 14$.

3. Найдите вычеты каждой из следующих степеней:

(1) $5^{20} \pmod 7$;
(2) $7^{1001} \pmod{11}$;
(3) $81^{119} \pmod{13}$;
(4) $13^{216} \pmod{19}$.

4. Найдите остаток от деления $1000!$ на 3^{300}.

5. Определите $U(n)$ и найдите обратные к каждому из его элементов при $n = 4,\ 11$ и 15.

6. Решите линейные сравнения:

(1) $4x \equiv 3 \pmod 4$;
(2) $3x + 2 \equiv 0 \pmod 4$;
(3) $2x - 1 \equiv 7 \pmod{15}$.

7. Покажите, что всякий элемент из $U(34)$ является степенью $\overline{3}$.

8. Покажите, что диофантово уравнение $x^2 - 7y^2 = 3$ не имеет целочисленных решений.

9. Покажите, что $p = 274\,177 = 1\,071 \cdot 2^8 + 1$ — простой делитель числа Ферма $F(6)$.

Указание: Сначала вычислите $1\,071^8$ по модулю p. Для этого заметьте, что $1\,071 = 7 \cdot 9 \cdot 17$ и найдите восьмую степень каждого из этих множителей по модулю p, а затем перемножьте их. Далее, так как $p = 1\,071 \cdot 2^8 + 1$, мы имеем:

$$(1\,071 \cdot 2^8)^8 \equiv 1 \pmod p.$$

С другой стороны,

$$(1\,071 \cdot 2^8)^8 \equiv 1\,071^8 \cdot 2^{64} \pmod p.$$

Замените $1\,071^8$ в последней формуле на его вычет и сравните с предыдущим уравнением. Тот факт, что p делит $F(6)$ вытекает отсюда как по волшебству.

Следующие три задачи — близкие родственники. Мы вернемся к ним в контексте теста на простоту для чисел Мерсенна (см. § 10.4).

10. Рассмотрим множество чисел вида $a + b\sqrt{3}$, где a и b целые. Оно обычно обозначается символом $\mathbb{Z}[\sqrt{3}]$. Поскольку элементы множества $\mathbb{Z}[\sqrt{3}]$ — вещественные числа, их можно складывать и перемножать. Покажите, что для $\alpha, \beta \in \mathbb{Z}[\sqrt{3}]$ их сумма $(\alpha + \beta)$ и произведение $(\alpha\beta)$ тоже лежат в $\mathbb{Z}[\sqrt{3}]$.

11. Определим отношение на $\mathbb{Z}[\sqrt{3}]$ следующим образом. Зафиксируем целое число n и возьмем $\alpha, \beta \in \mathbb{Z}[\sqrt{3}]$. Скажем, что $\alpha \equiv \beta \pmod{n}$, если и только если найдется такое число $\gamma \in \mathbb{Z}[\sqrt{3}]$, что $\alpha - \beta = n\gamma$. Покажите, что это отношение является отношением эквивалентности на $\mathbb{Z}[\sqrt{3}]$.

12. Для фиксированного натурального n обозначим через $\mathbb{Z}_n[\sqrt{3}]$ фактормножество $\mathbb{Z}[\sqrt{3}]$ по отношению эквивалентности из предыдущего упражнения. Предположим, что $\widetilde{\alpha}$ и $\widetilde{\beta}$ — классы из $\mathbb{Z}_n[\sqrt{3}]$. Проверьте, что правила

$$\widetilde{\alpha} + \widetilde{\beta} = \widetilde{\alpha + \beta},$$
$$\widetilde{\alpha} \cdot \widetilde{\beta} = \widetilde{\alpha \cdot \beta}$$

задают корректно-определенные операции на $\mathbb{Z}_n[\sqrt{3}]$.

13. В этом упражнении мы покажем, что число вида $4n+3$ не может быть представлено как сумма двух квадратов целых чисел.

(1) Покажите, что квадрат любого целого числа сравним с 0 или 1 по модулю 4.
(2) Используя (1), покажите, что для целых x и y сумма $x^2 + y^2$ сравнима с 0, 1 или 2 по модулю 4.

(3) Учитывая (2), докажите, что целое число вида $4n+3$ не может равняться сумме двух квадратов целых чисел.

Этот результат — частный случай теоремы, сформулированной Ферма в письме Робервалю (1640 год). Ферма знал также, что никакое из простых чисел вида $4n+1$ не может быть записано в виде суммы двух квадратов. За подробностями обратитесь к [49] ([Д.13]). Смотрите также упражнение 14 главы 6.

14. Напишите программу на основе алгоритма вычисления степеней по модулю n, описанного в § П.2 приложения. Исходные данные состоят из трех натуральных чисел: a, k и n; а выводиться должен вычет числа a^k по модулю n. Этот алгоритм будет основой всех приложений следующих глав.

Глава 6.
Индукция и Ферма

Теперь, зная базисные факты арифметики остатков, мы готовы вернуться к изучению простых чисел. Основной результат этой главы — очень полезная теорема, впервые доказанная Ферма. Она непосредственно следует из более сильной теоремы теории групп, которую мы изучим в главе 9. Здесь же мы, следуя указаниям Ферма, даем ее прямое доказательство методом *математической индукции*. С описания такого метода доказательства мы и начнем.

§ 6.1. Ханой! Ханой!

Думали ли Вы когда-нибудь над головоломкой под названием «Ханойские башни»? Она состоит из трех деревянных стержней, закрепленных на деревянной основе и некоторого числа деревянных дисков (из шести в моем комплекте). В центре каждого диска есть отверстие для нанизывания на стержень. Обозначим стрежни **A**, **B** и **C**. Диаметры дисков различны; в начале игры все они располагаются на стержне **A** в порядке убывания размеров: самый большой внизу, а самый маленький — наверху башни.

Цель задачи: переместить всю башню со стержня **A** на стержень **C**, используя **B** как перевалочный пункт, подчиняясь при этом следующим правилам:

(1) за один ход можно переместить только один диск;

(2) бóльший диск нельзя класть поверх меньшего.

Обратите внимание! По первому правилу с любого стержня за один ход можно снять и переместить только верхний диск. Поэтому, если убрать из головоломки перевалочный стержень **B**, задача станет неразрешимой.

Стоит попытаться решить головоломку самостоятельно, чтобы освоиться с ней. Практически это можно сделать довольно быстро. Но вопрос, который мы хотим поставить, выходит за рамки стандартной игры: можем ли мы найти формулу минимального числа ходов, требуемых для перемещения всей башни из n дисков со стержня **A** на стержень **C**? Естественно, мы предполагаем, что диски перемещаются согласно правилам. Поставленный вопрос имеет важнейшее значение для тех, кто готов поверить следующему рассказу.

Под высоким куполом одного индийского храма находятся три шпиля, густо усеянные алмазами, как пчела бархатным ворсом. В момент сотворения мира Бог поместил на один из них 64 диска чистого золота: наибольший — внизу, а остальные — сверху, так, что получилась башня. Он дал задание главному жрецу храма переставить диски согласно правилам, сформулированным выше. Когда вся башня из 64 дисков будет полностью перенесена на один из оставшихся шпилей, Бог вернется и положит конец миру. Итак, чтобы узнать, когда наступит конец света, Вам достаточно решить задачу о минимальном числе перемещений 64 дисков.

Оставив в стороне схоластические аспекты проблемы, вернемся к ее «деревянной» постановке, с которой мы начали. Если в игре только один диск, его достаточно просто перенести со стержня **A** на стержень **C**. При этом мы не нарушаем никаких правил, и головоломка решена. В этом случае достаточно

одного перемещения. Теперь предположим, что у нас есть два диска. Сначала мы перенесем меньший диск на стержень **B**; затем бóльший диск можно переставить на стержень **C**; и наконец, меньший — на **C**, так что он окажется сверху большого диска. Итак, для решения головоломки с двумя дисками достаточно трех ходов. Если у Вас есть такая головоломка, то неплохо было бы подсчитать количество перемещений, которое Вы сделаете, переставляя башню из четырех, а потом из пяти дисков.

Теперь поразмышляем над общим случаем головоломки из n дисков. Рассуждение станет легче для восприятия, если мы опишем его в виде диалога между учеником и учителем.

Учитель: Предположим, что диски пронумерованы сверху вниз числами 1, 2, ... , n, так что наименьший диск имеет номер 1 (и находится вверху), а наибольший — n (он внизу стержня). Что нужно сделать, чтобы можно было передвинуть диск n?

Ученик: Чего?

Учитель: Мы хотим передвинуть диск n, но все остальные диски стоят на нем. Что же нам делать?

Ученик: Переставить все диски, которые стоят на самом большом?

Учитель: Это так, нам следует перенести $n-1$ дисков, которые лежат на n-ом. Но нельзя забывать, что мы хотим переложить все диски на стержень **C**, а диск n при этом должен оказаться в самом низу стержня. Куда же нам деть остальные $n-1$?

Ученик: На стержень **B**?

Учитель: На стержень **B**. Однако возникают проблемы.

Ученик: Они всегда есть.

Учитель: Правила! По первому из них мы можем менять

положение только одного из дисков; а по второму — диски должны быть нанизаны на стержень **B** в порядке убывания их размеров. Как нам передвинуть $n-1$ меньших диска с **A** на **B**?

Ученик: Нам следует передвигать только один диск за один ход, не нарушая правил.

Учитель: А более точно?

Ученик: Было бы здорово, я полагаю, чтобы задача с $n-1$ дисками была уже решена. Тогда можно было бы передвинуть башню из $n-1$ дисков со стержня **A** на **B**.

Учитель: А что будет служить перевалочным пунктом?

Ученик: Может, это **C**?

Учитель: Точно. Подведем итог. Нам нужно переставить диск n на **C**. Но мы можем это сделать только после перемещения на стержень **B** $(n-1)$ дисков, которые находятся над ним. Сделаем это, играя с $n-1$ меньшими дисками, используя **C** в качестве перевалочного пункта. Таким образом весь набор из $n-1$ дисков будет передвинут со стержня **A** на **B**. Осуществив это, мы свободно перенесем n-ый диск на **C**.

Ученик: Как же мы потом передвинем оставшиеся $n-1$ диски на **C** так, чтобы они правильно легли поверх диска n? Мне кажется, что нам нужно опять поиграть в $n-1$ диски, чтобы передвинуть их с **B** на **C**?

Учитель: Такая игра может занять много времени, но это именно тот путь, который приводит к победе. Заметим, что нам пришлось решить головоломку с $n-1$ дисками *дважды*. В первый раз мы переместили башню, состоящую из $n-1$ дисков с **A** на **B** (используя **C** в качестве перевалочного). Это нам освободило n-ый диск, и мы переместили его на **C**. Наконец, мы еще раз сыграли в игру с $n-1$ дисками, перемещая их со стержня **B** на **C** (используя **A** как перевалочный). В результате все n дисков оказались нанизанными на стержень **C**, и ни одно правило не было нарушено.

Ученик: Ух!

Учитель: Это еще не конец!

Ученик: Не конец? О, Боже!

Учитель: Мы хотели найти наименьшее число ходов, необходимое для решения головоломки, разве нет?

Ученик: Полагаю, все еще хотим.

Учитель: Чтобы облегчить дальнейшие рассуждения, обозначим через $T(n)$ минимальное число ходов, требуемых для решения головоломки с n дисками. Как мы увидели, для ее решения нам сначала нужно передвинуть $n-1$ дисков, лежащих над n-ым. Сколько ходов нам потребуется для этого?

Ученик: Чтобы переместить их на стержень **B**, нам нужно решить головоломку с $n-1$ дисками, не так ли? И нам, я полагаю, потребуется по крайней мере $T(n-1)$ ходов.

Учитель: Поскольку мы перенесли $n-1$ меньших дисков на стержень **B**, стержень **C** остался пустым. Поэтому теперь можно передвинуть последний n-ый диск на **C**. Много ли нам нужно ходов для этого?

Ученик: Один?

Учитель: Естественно; и сколько теперь всего затрачено ходов от начала игры?

Ученик: Уф, $T(n-1)+1$?

Учитель: Что дальше?

Ученик: Нам еще нужно перетащить $n-1$ меньших дисков с **B**, где они сейчас находятся, на **C**, накрыв ими диск n.

Учитель: Так, а сколько передвижений потребуется для достижения этой цели?

Ученик: Не меньше, чем $T(n-1)$, разумеется, потому что это минимальное число ходов в головоломке с $n-1$ дисками.

Учитель: Итак, мы видим, что с момента начала игры всего нам предстоит сделать $T(n-1)+1+T(n-1) = 2T(n-1)+1$ ходов. Более того, если мы внимательно посмотрим на проведенные рассуждения, мы убедимся, что нет никакой возможности полностью решить головоломку за меньшее число ходов. Так что такое $T(n)$?

Ученик: Минимальное число ходов, требуемое для решения головоломки по перемещению n дисков со стержня **A** на **C**.

Учитель: Да, но как нам вычислить $T(n)$, предполагая, что мы уже знаем $T(n-1)$?

Ученик: $T(n) = 2T(n-1) + 1$?

Учитель: Отлично! Теперь подсчитаем минимальное число ходов, необходимых для решения задачи с шестью дисками.

Конечный результат диалога — формула

$$T(n) = 2T(n-1) + 1.$$

Заметим, что эта формула *не говорит* нам чему равно $T(n)$. Для вычисления $T(n)$ мы должны сначала найти $T(n-1)$. Значит, $T(n)$ вычисляется многократным применением формулы. Например, для вычисления $T(6)$, нам сначала необходимо найти $T(1), T(2), \ldots, T(5)$. Поскольку, как мы уже усвоили, $T(1) = 1$, то

$$T(2) = 2T(1) + 1 = 3.$$

Действуя так же, мы имеем

$$T(3) = 7, \quad T(4) = 15, \quad T(5) = 31 \quad \text{и} \quad T(6) = 63.$$

Следовательно, для решения моей головоломки, состоящей из 6 дисков, мне нужно по крайней мере 63 хода. А как насчет головоломки в индийском храме? Для ответа на этот вопрос нужно вычислить $T(64)$, довольно страшная задача.

Равенство $T(n) = 2T(n-1) + 1$ называется *рекуррентной формулой*. Другими словами, чтобы найти $T(n)$ нужно применить ее несколько раз, на каждом этапе используя в качестве исходных данных результат предыдущего вычисления. Вы можете спросить, как доказать эту формулу? Ответ: приведенный диалог и является ее доказательством. Допускаю, что оно изложено в такой форме, которая может показаться чересчур экзотичной, чтобы называться математическим доказательством. Но этот недостаток легко лечится; все что нужно сделать — это выбрать основные моменты из диалога и переписать их на привычном математическом жаргоне.

Тот факт, что мы получили только рекуррентную формулу, не остановит наших попыток найти конечное выражение для $T(n)$, т.е. такое, из которого $T(n)$ получается простой подстановкой значения переменной n. Если Вы обратили внимание на значения $T(n)$, вычисленные выше, то можете догадаться о вероятном виде искомой формулы. Но раз конечная формула будет угадана, перед нами встает новая проблема: мы должны будем доказать, что она работает при всех значениях переменной n. Заметим, что проверяя гипотетическую формулу для какой-то таблицы чисел, можно убедиться лишь в том, что она верна именно для чисел таблицы, но не более. Для доказательства формулы мы вводим метод *математической индукции*.

Быть может, Вы думаете: «Зачем беспокоиться о поиске другой формулы? Что плохого в рекуррентной?» Резонные вопросы. В конце концов, для определения $T(n)$ при фиксированном n нам всего-то и нужно, что вычислить с ее помощью $T(0), \ldots, T(n)$. Да компьютер сделает это очень быстро! Чего же еще желать? Как раз здесь пути математики и вычислительной математики расходятся.

Несколько преувеличивая, мы могли бы сказать, что вычислительная математика предназначена для грубого, силового подхода к работе: делать настолько эффективно, насколько

получается; а математика, наоборот, нацелена на поиск пути достижения результата с минимальными вычислениями. Конечно, это, в действительности, две стороны одной медали. В реальном мире проблемы решаются органическим синтезом математики и вычислительной математики. Так что эти науки чаще сотрудничают, чем конкурируют.

§ 6.2. Математическая индукция

Слово «*индукция*» используется в математике в очень специфическом техническом значении. Иногда к нему добавляют определение *математическая*, как в названии данного параграфа. Однако это слово имеет множество значений, 12 из которых приведены в «Оксфордском словаре английского языка». Математическое значение слова «индукция» происходит из традиционного его значения, употребляемого в логике, которое близко к повседневному употреблению. Ссылаясь на «Оксфордский словарь английского языка», «индукция» в таком смысле — это

> «процесс выведения общего закона или принципа из наблюдаемых частных случаев.»

Значит, когда мы додумались до конечной формулы для $T(n)$, используя значения $T(1), \ldots, T(6)$, мы применяли индукцию. Конечно, Вы делаете подобные вещи ежедневно, на протяжении всего дня. Но в математике индукция иногда приводит к ужаснейшим ошибкам (впрочем, по мнению автора, такое случается и в повседневной жизни).

Часто цитируемый пример ошибки, к которой может привести индукция, — это утверждение Ферма, высказанное Френиклю, о том, что все числа вида

$$F(n) = 2^{2^n} + 1$$

будут простыми. Он, вероятно, проверил гипотезу для $n = 0, 1, 2, 3$ и 4, что довольно просто, а затем сделал обобщение. Следующее число

$$F(5) = 2^{2^5} + 1 = 4\,294\,967\,297$$

довольно велико, если все, что у вас есть — это ручка и бумага. Был ли Ферма напуган размерами числа, или допустил ошибку при его вычислении? Вероятно, мы этого никогда не узнаем. Но, как мы видели, $F(5)$ на самом деле составное число, и Эйлер был первым, кто в 1738 году нашел его делитель (мы будем подробно изучать эти числа в главе 10). С тех пор ошибка Ферма стала предупреждением об опасности обобщений на основе нескольких примеров.

В XVII столетии ряд математиков, и Ферма среди них, начали проявлять беспокойство по поводу отсутствия доверия к результатам, доказанным по индукции. Это побудило их развить метод, который очень подходит для доказательства фактов, продиктованных обобщением численных экспериментов, т.е. результатов, полученных ординарной индукцией. Новый метод получил известность как *математическая индукция*, или *конечная индукция*, или *рекуррентное рассуждение*. В трактате Б. Паскаля «Об арифметическом треугольнике», опубликованном в 1654 году, мы найдем объяснение метода математической индукции практически в современном виде. Конечно, «арифметический треугольник» из названия трактата сейчас известен как *треугольник Паскаля*. Паскаль, человек многих талантов, был первоклассным геометром и физиком. Именно он изобрел первую механическую машину для вычислений (калькулятор). Его «Мысли» — классика французской литературы.

Вернемся на время к Ханойским башням. Внимательно рассматривая значения $T(1), \ldots, T(6)$, вычисленные в предыдущем параграфе, мы заметим, что каждое из них удовлетворяет условию: $T(n) + 1 = 2^n$. Поэтому естественно предположить,

что *минимальное число ходов, требуемое для передвижения башни из n дисков в головоломке «Ханойские башни», равно* $2^n - 1$. Отметим, что в действительности мы имеем бесконечно много утверждений, пронумерованных натуральными числами. А наш эксперимент говорит нам только, что эти утверждения справедливы для $n = 1, \ldots, 6$ и не более.

Пример с Ханойскими башнями типичен. Изучение (конечной) таблицы с экспериментальными данными задачи часто подсказывает нам гипотетическое утверждение $S(n)$, и мы ожидаем, что оно справедливо для всех натуральных чисел. Математическая индукция предлагает систематический подход к доказательству многих таких утверждений.

Принцип математической индукции. *Предположим, что для каждого натурального числа n есть утверждение* $S(n)$*, обладающее следующими двумя свойствами:*

(1) $S(1)$ *верно;*

(2) *если* $S(k)$ *верно для натурального* k*, то* $S(k+1)$ *также верно.*

Тогда $S(n)$ *справедливо для всех натуральных n.*

Попробуем понять, почему этот принцип работает. Предположим, у нас есть утверждение, обладающее свойствами (1) и (2) принципа. Второе из них говорит нам, что если для некоторого натурального k мы можем показать справедливость $S(k)$, то $S(k+1)$ тоже будет верным. По первому свойству $S(1)$ — истинное утверждение. Поэтому, применяя (2) при $k = 1$, мы заключаем, что $S(2)$ верно. Теперь, зная об истинности $S(2)$, мы опять можем применить (2), но на этот раз при $k = 2$. Получим истинность $S(3)$. Таким образом, выбрав произвольное натуральное n, мы можем продолжать действовать подобным образом пока не достигнем $S(n)$. Поэтому утверждение должно быть справедливым для каждого положительного целого n.

Конечно, это рассуждение не доказывает принципа математической индукции. На самом деле, в некотором смысле, он вообще не может быть доказан! Анри Пуанкаре (Poincaré), один из известных математиков XIX столетия, очень хорошо объясняет этот момент:

> «Суждение, на котором основан способ рекуррентного рассуждения, может быть изложено в других формах; можно сказать, например, что в бесконечно большом множестве различных натуральных чисел всегда есть одно, которое меньше других. Можно легко переходить от одного выражения к другому и таким образом создавать иллюзию доказательства законности рассуждения путем рекурренции. Но в конце концов всегда придется остановиться; мы всегда придем к недоказуемой аксиоме, которая, в сущности, будет не что иное, как предположение, подлежащее доказательству, но только переведенное на другой язык.»

Так почему мы все же верим в строгость принципа индукции? Снова обратимся к Пуанкаре за помощью:

> «Здесь сказывается только утверждение могущества разума, который способен постичь бесконечное повторение одного и того же акта, раз этот акт оказался возможным однажды.»

По той же причине мы без труда осознаем бесконечность множества целых, хотя начинаем знакомство с ними с нескольких чисел. Цитаты Пуанкаре мы взяли из его классической книги «Наука и гипотеза» ([36], [Д.11]).

Применим принцип к проблеме «Ханойские башни». Утверждение, которое мы собираемся проверить, говорит, что минимальное число перемещений $T(n)$, требуемых для решения головоломки «Ханойские башни» из n дисков, равно $2^n - 1$. Или

короче: мы хотим доказать, что $T(n) = 2^n - 1$. Согласно принципу математической индукции, это будет верно для любого натурального n, если мы сможем доказать две вещи. Сначала мы должны проверить формулу для головоломки с 1 диском. Но мы уже убедились, что $T(1) = 1$ и $2^n - 1 = 1$. Так что первый шаг уже сделан. На математическом жаргоне он называется *базой индукции*.

Далее нужно сделать *индуктивный переход*, или *шаг индукции*, т.е. показать, что если утверждение верно для какого-то $k \geqslant 1$, то оно верно и для $k + 1$. Индуктивный переход основывается на двух моментах:

- Предположение о том, что утверждение верно для некоторого $k \geqslant 1$. В нашем примере это означает гипотетическое равенство $T(k) = 2^k - 1$ для некоторого $k \geqslant 1$. Такое предположение называется *предположением индукции*.

- Какая-нибудь связь между $T(k)$ и $T(k+1)$. В примере — это рекуррентное соотношение: $T(k + 1) = 2T(k) + 1$.

Итак, предположим, что $T(k) = 2^k - 1$ для некоторого $k \geqslant 1$. Из рекуррентной формулы следует, что

$$T(k + 1) = 2T(k) + 1 = 2(2^k - 1) + 1 = 2^{k+1} - 2 + 1 = 2^{k+1} - 1.$$

Таким образом, условие (2) принципа математической индукции в нашем случае тоже верно. Теперь принцип математической индукции дает: так как (1) и (2) справедливы, то $T(n) = 2^n - 1$ для всех натуральных n.

Зная минимальное число ходов, приводящих к решению задачи «Ханойские башни», мы можем теперь подсчитать оставшееся время до конца света. Напомним, что башня в индийском храме состоит из 64 дисков. Поэтому общее число перемещений, которые жрецу надо сделать со дня творения, равно $T(64) = 2^{64} - 1$. Мало этого, нам еще хочется знать, сколько времени займет этот процесс. Предположим, что жрецу для

перестановки одного диска потребуется, в среднем, 30 минут. Диски, разумеется, имеют разные размеры. Нам не сказано насколько велик максимальный из них, но, по-видимому, не маленький, поскольку создал его сам Господь. А так как они, к тому же, сделаны из золота, то должны быть довольно тяжелыми. Поэтому полчаса — достаточно осторожное предположение. Величина 2^{64} имеет порядок 10^{19}, и несложные вычисления показывают, что жрецу потребуется около 10^{14} лет, чтобы перенести все диски. Таким образом, считая, что от Большого Взрыва прошло около 10^{11} лет, можно прикинуть, сколько еще осталось.

Эта легенда впервые была опубликована в Париже в 1883 году, одновременно с головоломкой, неким Н. Клаусом из колледжа в Ли-Соу-Стэйна. Имя человека и название колледжа — в действительности, анаграммы имени Люка д'Амьен, преподавателя лицея Святого Людовика. Это математик Ф. Е. А. Люка (F. E. A. Lucas), придумавший как головоломку, так и рассказ о ее происхождении. Его книга «Математические развлечения» 1894 года стала классикой предмета. Люка занимался и теорией чисел. В 10 и 11 главах мы разберем два теста неприводимости, которые он открыл. Используя один из них, Люка, не прибегая к помощи компьютеров (которых тогда еще и не было), показал, что число Мерсенна

$$M(127) = 170\,141\,183\,460\,469\,231\,731\,687\,303\,715\,884\,105\,727$$

является простым.

§ 6.3. Теорема Ферма

Теорему, которую мы хотим доказать, иногда ласково называют «*малой теоремой Ферма*». Она говорит, что если p простое число и a — любое целое, то p делит $a^p - a$. Частный случай этой теоремы известен многие сотни лет, но Ферма, ка-

жется, был первым, кто доказал теорему в полной общности. Начнем с перевода формулировки теоремы на язык сравнений.

Теорема Ферма. *Пусть p — положительное простое число и a — целое; тогда*

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

Чтобы доказать теорему с помощью математической индукции, нам нужно найти утверждение $P(n)$, к которому можно применить метод. Вот оно:

$$n^p \equiv n \pmod{p} \qquad \text{для натурального } n.$$

Отметим, что утверждение $P(n)$ говорит о верности сравнения только для натуральных чисел, т.е. положительных целых n. Таким образом, мы не доказываем теоремы в полной общности. Однако, любое целое число сравнимо по модулю p с каким-то неотрицательным целым, меньшим p. Поэтому достаточно проверить теорему для $0 \leqslant a \leqslant p-1$. В частности, теорема будет доказана, если показать справедливость утверждений $P(n)$ для всех $n \geqslant 1$.

Разумеется, $P(1)$ истинное высказывание, поскольку $1^p = 1$. Таким образом, база индукции выполнена. Для индуктивного перехода от $P(n)$ к $P(n+1)$ нам нужно выявить связь между этими утверждениями. Она обеспечивается *биномиальной теоремой*. Доказательство становится намного проще, если мы выделим из него вспомогательное утверждение, в действительности являющееся вариантом бинома Ньютона для целых чисел по модулю p.

Лемма. *Пусть p — положительное простое число, a и b — целые; тогда*

$$(a+b)^p \equiv a^p + b^p \pmod{p}.$$

Доказательство леммы. Как следует из формулы бинома Ньютона,

$$(a+b)^p = a^p + b^p + \sum_{i=1}^{p-1} \binom{p}{i} a^{p-i} b^i.$$

Таким образом, для доказательства леммы достаточно показать, что

$$\sum_{i=1}^{p-1} \binom{p}{i} a^{p-i} b^i \equiv 0 \pmod{p}.$$

А это, в свою очередь, будет немедленно следовать из делимости биномиальных коэффициентов $\binom{p}{i}$ на p при $1 \leqslant i \leqslant p-1$. По определению,

$$\binom{p}{i} = \frac{p(p-1)\cdots(p-i+1)}{i!}.$$

Поскольку биномиальные коэффициенты — целые числа, знаменатель дроби должен делить числитель. Кроме того, если $1 \leqslant i \leqslant p-1$, то p не является делителем числа $i!$. Поэтому сомножитель p из числителя дроби не может сократиться ни с каким сомножителем знаменателя. Значит, знаменатель обязан делить произведение $(p-1)\cdots(p-i+1)$. Отсюда $\binom{p}{i}$ — число кратное p, что и требовалось доказать.

Теперь можно вернуться к доказательству теоремы Ферма. Предположение индукции:

$$n^p \equiv n \pmod{p} \qquad \text{для некоторого натурального } n.$$

В качестве индуктивного перехода нам нужно показать, что $(n+1)^p \equiv n+1 \pmod{p}$. По лемме

$$(n+1)^p \equiv n^p + 1^p \equiv n^p + 1 \pmod{p}.$$

По предположению индукции, n^p можно заменить на n. Проделав это, мы получаем: $(n+1)^p \equiv n^p + 1 \equiv n+1 \pmod{p}$, что мы и хотели показать.

Наиболее интересных приложений теоремы Ферма придется подождать до следующей главы. А сейчас будем довольствоваться использованием теоремы для упрощения вычислений степеней по модулю p, проблемы, с которой мы уже сталкивались. Сначала переформулируем теорему в более удобной форме.

Согласно теореме, если p простое число, а a — целое, то $a^p \equiv a \pmod{p}$. Предположим теперь, что a не делится на p. Тогда, ввиду простоты p, числа a и p взаимно просты, и из теоремы обратимости следует, что a обратимо по модулю p. Пусть a' — обратный к a элемент. Умножая на a' сравнение $a^p \equiv a \pmod{p}$, имеем

$$a'a \cdot a^{p-1} \equiv a'a \pmod{p}.$$

Но $a'a \equiv 1 \pmod{p}$, и окончательно $a^{p-1} \equiv 1 \pmod{p}$. Именно этот вариант уравнения из теоремы Ферма мы будем использовать в дальнейшем чаще всего. Для будущих ссылок сформулируем его в виде утверждения.

Теорема Ферма. *Пусть p — положительное простое число и a — целое, не делящееся на p; тогда $a^{p-1} \equiv 1 \pmod{p}$.*

Задача, к которой мы хотели бы применить теорему Ферма, звучит следующим образом. Для трех данных натуральных чисел a, k и p таких, что $k > p-1$, найти вычет a^k по модулю p.

Если p делит a, то вычет равен 0. Поэтому без ограничения общности можно предполагать, что p не делит a. Разделим k на $p-1$ с остатком: $k = (p-1)q+r$, где q и r — неотрицательные целые числа, причем $0 \leqslant r < p-1$. Следовательно,

$$a^k \equiv a^{(p-1)q+r} \equiv \left(a^{p-1}\right)^q a^r \pmod{p}.$$

По теореме Ферма $a^{p-1} \equiv 1 \pmod{p}$. Значит, $a^k \equiv a^r \pmod{p}$, и достаточно делать вычисления только для экспонент, показатель которых меньше $p-1$.

Приведем наглядный пример, демонстрирующий силу этой простой редукции. Допустим, нам нужно найти вычет числа $2^{5\,432\,675}$ по модулю 13. Рецепт предыдущей главы предписывал вычислить несколько степеней двойки по модулю 13, прежде чем что-нибудь получить. Посмотрим, что будет, если применить теорему Ферма. Сначала найдем остаток от деления 5 432 675 на $13-1=12$. Он равен 11. Затем, рассуждая, как в предыдущих абзацах, мы имеем

$$2^{5\,432\,675} \equiv 2^{11} \pmod{13}.$$

Непосредственный счет дает окончательный ответ:

$$2^{11} \equiv 7 \pmod{13}.$$

§ 6.4. Вычисление корней

Остается еще один вопрос, который нужно рассмотреть. Существует ли натуральное k, *меньшее*, чем $p-1$, при котором $a^k \equiv 1 \pmod{p}$ для всех целых чисел a, не делящихся на p?

Можно попытаться ответить на него следующим образом. Хорошо известная теорема алгебры говорит, что полиномиальное уравнение не может иметь корней больше, чем его степень. А поскольку, по предположению, сравнение $a^k \equiv 1 \pmod{p}$ справедливо для всех целых чисел, взаимно простых с p, корни полиномиального уравнения $x^k = \overline{1}$ в $\mathbb{Z}_p$ — это $\overline{1}, \ldots, \overline{p-1}$. Таким образом, уравнение имеет $p-1$ разных корня. По упомянутой теореме получаем оценку: $k \geqslant p-1$. Так что ответ на поставленный вопрос отрицателен.

Хотя эти рассуждения корректны, они прячут истинную природу проблемы в теореме, с помощью который мы решили этот вопрос. Когда мы говорим о полиномиальном уравнении, мы обычно имеем ввиду уравнение с вещественными или комплексными коэффициентами. А под «корнем» мы подразумеваем вещественный или комплексный корень. Но коэффициенты

уравнения из нашего рассуждения, как и его корни, суть элементы $\mathbb{Z}_p$. Вот где собака зарыта! Нам нужно показать, что теорема об оценке числа корней полиномиального уравнения справедлива и в этом случае. Может, это опять неоправданный педантизм? Отнюдь нет. Хотя эта теорема верна в случае простого модуля, для составного модуля она *ложна*!

Теорема. *Пусть $f(x)$ — многочлен степени k с целыми коэффициентами и старшим коэффициентом 1. Если p — простое число, то $f(x)$ имеет не более k корней в $\mathbb{Z}_p$.*

Прежде чем погрузиться в доказательство, мы должны обратить внимание на два момента. Во-первых, модуль должен быть простым, поскольку нам нужно следующее свойство: если $ab \equiv 0 \pmod{p}$, то или $a \equiv 0 \pmod{p}$ или $b \equiv 0 \pmod{p}$. Заметим, что это, фактически, фундаментальное свойство простых чисел, сформулированное на языке сравнений. Во-вторых, мы выделим один из моментов доказательства в отдельную лемму, которую докажем в конце этого параграфа.

Лемма. *Пусть $h(x)$ — многочлен степени m с целыми коэффициентами. Для любого целого числа α найдется многочлен $q(x)$ степени $m-1$, удовлетворяющий условию:*

$$h(x) = (x - \alpha)q(x) + h(\alpha).$$

Докажем теорему индукцией по n (степени многочлена) с помощью этой леммы. Если $n = 1$, то $f(x) = x + b$. И соответствующее уравнение имеет только одно решение $\overline{-b}$ в $\mathbb{Z}_p$. Итак, многочлен степени 1 имеет единственный корень в $\mathbb{Z}_p$, и теорема в этом случае верна.

Предположим теперь, что *любой* многочлен степени $k-1$ со старшим коэффициентом 1 имеет не более $k-1$ корня в $\mathbb{Z}_p$. Это предположение индукции. Мы хотим показать, что данное предположение влечет аналогичное утверждение для многочлена степени k со старшим коэффициентом 1.

Итак, пусть $f(x)$ — многочлен степени k с целыми коэффициентами, старший из которых равен 1. Если $f(x)$ не имеет корней в $\mathbb{Z}_p$, то доказывать нечего, ибо $0 \leqslant k$. Такие многочлены в действительности существуют; соответствующий пример приведен после доказательства теоремы. Следовательно, мы можем предполагать, что $f(x)$ имеет корень $\overline{\alpha} \in \mathbb{Z}_p$; иначе говоря, $f(\alpha) \equiv 0 \pmod{p}$. По лемме

$$f(x) = (x - \alpha)q(x) + f(\alpha), \tag{4.1}$$

где степень $q(x)$ равна $k - 1$. Так как старшие коэффициенты многочленов $f(x)$ и $x - \alpha$ равны 1, то же самое справедливо и для многочлена $q(x)$. Значит, мы можем применить индуктивное предположение к $q(x)$.

Редуцируя равенство (4.1) по модулю p, мы получаем

$$f(x) \equiv (x - \alpha)q(a) \pmod{p}. \tag{4.2}$$

Пусть $\overline{\beta} \neq \overline{\alpha}$ — еще один корень $f(x)$ в $\mathbb{Z}_p$. Это означает, что

$$f(\beta) \equiv 0 \pmod{p}, \qquad \text{но} \qquad \alpha - \beta \not\equiv 0 \pmod{p}.$$

Заменяя x на α в (4.2), и используя эти сравнения, мы имеем

$$0 \equiv f(\beta) \equiv (\beta - \alpha)q(\beta) \pmod{p}.$$

Так как p простое, то это влечет сравнение $q(\beta) \equiv 0 \pmod{p}$. Мы заключаем, что если $\overline{\beta}$ — корень многочлена $f(x)$, отличный от $\overline{\alpha}$, то $\overline{\beta}$ является и корнем $q(x)$ в $\mathbb{Z}_p$. Иначе говоря, у $f(x)$ только на один корень больше (в $\mathbb{Z}_p$), чем у $q(x)$. А по предположению индукции у последнего многочлена не более, чем $k-1$ различных корней в $\mathbb{Z}_p$. Следовательно, $f(x)$ не может иметь больше k разных корней, что завершает индуктивное доказательство.

Рассмотрим несколько примеров. Первый из них — многочлен $f(x) = x^2 + 3$. Он удовлетворяет всем условиям теоремы,

но не имеет корней по модулю 5. Действительно, единственно возможные вычеты квадратов целых чисел по модулю 5 — это 1 и 4, откуда немедленно следует наше заявление. Это как раз тот пример, о котором упоминалось в доказательстве теоремы.

Второй пример иллюстрирует, что происходит, когда мы пытаемся искать корни многочлена по составному модулю. Корни многочлена $x^2 - 170$ в $\mathbb{Z}_{385}$ — это $\overline{95}$, $\overline{150}$, $\overline{235}$ и $\overline{290}$, что легко проверить. Итак, мы представили полиномиальное уравнение степени 2 с четырьмя корнями. Это, конечно, не противоречит нашей теореме, поскольку 385 составное число.

Для завершения доказательства теоремы необходимо проверить лемму. Мы опять будем делать это по индукции. Однако, если попытаться применить принцип индукции в том виде, который сформулирован в § 6.2, мы убедимся, что в доказательстве откроется дырочка. Позже посмотрим, к какой проблеме может привести такое доказательство.

Мы вернемся к лемме, сформулировав принцип индукции в более тонкой форме. Индуктивная гипотеза принципа в форме из § 6.2 предполагает, что $S(k)$ верно для некоторого целого $k \geqslant 1$. Но практически, когда мы пытаемся доказывать утверждение $S(k+1)$, опираясь на $S(k)$, мы уже знаем об истинности $S(1), S(2), \ldots, S(k-1)$. Поэтому предположение о справедливости не только $S(k)$, но и $S(1), S(2), \ldots, S(k-1)$ будет небольшой и естественной модификацией метода. Зато для доказательства $S(k+1)$ мы сможем использовать более полную информацию. Формулировка принципа, включающая это дополнение, выглядит следующим образом.

Принцип математической индукции. *Предположим, что для каждого натурального числа n сформулировано утверждение $S(n)$, обладающее следующими двумя свойствами:*

(1) *$S(1)$ верно;*

(2) *если $S(1), \ldots, S(k)$ верны для натурального k, то утверждение $S(k+1)$ также верно.*

Тогда $S(n)$ справедливо для всех натуральных n.

Пользуясь новой формулировкой принципа, лемму теперь можно легко доказать индукцией по степени m многочлена $h(x)$. Если $m = 1$, то $h(x) = ax + b$ для некоторых целых a и b. Поэтому

$$h(x) = ax + b = a(x - \alpha) + a\alpha + b = a(x - \alpha) + h(\alpha).$$

Предположим теперь, что утверждение леммы выполнено для любого многочлена с целыми коэффициентами, степень которого не превосходит $m - 1$. Исходя из этого предположения, мы хотим показать, что лемма справедлива для многочлена $h(x)$ с целыми коэффициентами степени m. Пусть

$$h(x) = a_m x^m + a_{m-1} x^{m-1} + \cdots + a_1 x + a_0,$$

где $a_m \neq 0$. Обозначим через $g(x)$ разность

$$h(x) - a_m x^{m-1}(x - \alpha),$$

то есть

$$g(x) = (a_{m-1} + a_m \alpha) x^{m-1} + a_{m-2} x^{m-2} + \cdots + a_1 x + a_0.$$

Ясно, что степень многочлена $g(x)$ меньше или равна $m - 1$. Заметим, однако, что степень будет в точности равна $m - 1$, только если $a_{m-1} + a_m \alpha \neq 0$, но у нас нет способа узнать так это, или нет. Поэтому возможна ситуация, когда степень $g(x)$ *меньше*, чем $m - 1$. Это как раз тот момент, когда мы сталкиваемся с трудностями при использовании принципа индукции в форме, сформулированной в § 6.2. Заметим также, для будущих ссылок, что $g(\alpha) = h(\alpha)$.

Так как степень многочлена $g(x)$ меньше или равна $m - 1$, предположение индукции влечет

$$g(x) = j(x)(x - \alpha) + g(\alpha),$$

где $j(x)$ — многочлен с целыми коэффициентами, степень которого на единицу меньше, чем степень $g(x)$. Ввиду равенства $g(\alpha) = h(\alpha)$, получаем

$$g(x) = j(x)(x - \alpha) + h(\alpha).$$

Но $h(x) = g(x) + a_m x^{m-1}(x - \alpha)$, так что

$$h(x) = \big(j(x) + a_m x^{m-1}\big)(x - \alpha) + h(\alpha).$$

Наконец, $j(x)+a_m x^{m-1}$ имеет степень в точности $m-1$, потому что степень $j(x)$ меньше $m-1$. Это завершает доказательство леммы.

Есть более прямое доказательство этой леммы, использующее деление многочленов. Как и прежде, пусть $h(x)$ — многочлен степени m с целыми коэффициентами. Разделив $h(x)$ на $x - \alpha$, мы найдем такие многочлены $q(x)$ и $r(x)$, что

$$h(x) = q(x)(x - \alpha) + r(x), \tag{4.3}$$

причем либо $r(x) = 0$, либо его степень меньше степени $x - \alpha$. Значит, его степень должна быть равной 0, т.е. $r(x) = c$ — целое число. Заменим в уравнении (4.3) x на α:

$$h(\alpha) = q(\alpha)(\alpha - \alpha) + c = c.$$

Таким образом, мы можем переписать (4.3) в виде:

$$h(x) = q(x)(x - \alpha) + h(\alpha),$$

что завершает доказательство.

Упражнения

1. Методом математической индукции докажите, что

(1) число $n^3 + 2n$ делится на 3 при любом натуральном значении переменной n;

(2) если $n > 0$ — нечетное целое число, то $n^3 - n$ делится на 24;

(3) выпуклый n-угольник имеет ровно $n(n-3)/2$ диагоналей;

(4) для каждого целого $n \geqslant 1$ имеет место соотношение $\sum\limits_{k=1}^{n} k(k+1) = n(n+1)(n+2)/3$.

2. Числа, заданные формулой $h_n = 1 + 3n(n-1)$ $(n \in \mathbb{N})$, называются *гексагональными*. Их название обязано тому факту, что они, по определенным правилам, могут быть размещены на сторонах правильного шестиугольника.

(1) Вычислите сумму первых n гексагональных чисел для $n = 1, 2, 3, 4$ и 5. Обобщите полученные результаты до гипотетической формулы суммы первых n гексагональных чисел.

(2) Докажите найденную формулу методом математической индукции.

3. Напомним, что f_n — n-ое число Фибоначчи, если $f_0 = 1$, $f_1 = 1$ и $f_n = f_{n-1} + f_{n-2}$. Индукцией по n покажите, что

$$f_n = \frac{\alpha^n - \beta^n}{\sqrt{5}},$$

где α и β — корни квадратного уравнения $x^2 - x - 1 = 0$.

4. Рассмотрим последовательность целых положительных чисел $S_0, S_1, S_2, S_3, \ldots$, определенных рекуррентным соотношением:

$$S_0 = 4 \qquad \text{и} \qquad S_{k+1} = S_k^2 - 2.$$

Пусть $\omega = 2 + \sqrt{3}$ и $\varpi = 2 - \sqrt{3}$. Индукцией по n покажите, что $S_n = \omega^{2^n} + \varpi^{2^n}$.

5. Принцип математической индукции очень полезен, но обращаться с ним нужно крайне осторожно. Например, ниже приведено доказательство абсурдного утверждения: в конечном множестве цветных мячиков все мячи должны быть одного цвета. Найдите в нем ошибку.

> Если в множестве только один мяч, то утверждение, очевидно, справедливо. Теперь предположим, что в любом множестве из k мячей все шары одного цвета. Мы хотим показать, что все мячи в множестве из $k+1$ мяча тоже покрашены в один цвет. Обозначим мячи этого множества через $m_1, \ldots m_{k+1}$. Удалив из него m_{k+1}, мы получим множество, содержащее k мячей и, по предположению индукции, мячи $m_1, \ldots, m_k$ одного цвета. Поэтому нам осталось только показать, что цвет m_{k+1} совпадает с цветом мячей из множества $\{m_1, \ldots, m_k\}$, и доказательство будет закончено. Но $\{m_2, \ldots, m_{k+1}\}$ — множество, которое тоже насчитывает k элементов и, по индуктивному предположению, состоит из мячей одного цвета. Следовательно, m_{k+1} имеет тот же цвет, что и, скажем, m_2. Итак, мячи $m_1, \ldots, m_{k+1}$ окрашены в один цвет.

6. Допустим, что у нас есть 3^n монет, одна из которых фальшивая. Известно, что фальшивая монета легче остальных. Вам дали чашечные весы без гирь. Так что у Вас только один способ взвешивать монеты: положить несколько монет на левую чашку, еще несколько — на правую и посмотреть, какая из частей тяжелее. Покажите с помощью математической индукции, что для определения фальшивой монеты будет достаточно n взвешиваний.

7. Пусть p_n — n-ое простое число. Например, $p_1 = 2,\ p_2 = 3$ и $p_3 = 5$. Мы хотим выразить через n верхнюю оценку для p_n.

(1) Покажите, что $p_{n+1} \leqslant p_1 \cdots p_n + 1$.

(2) Используя это неравенство и математическую индукцию, докажите, что n-ое простое число удовлетворяет неравенству: $p_n \leqslant 2^{2^n}$.

8. Покажите, используя теорему Ферма, что $2^{70} + 3^{70}$ делится на 13.

9. Пусть a — натуральное число, записанное в десятичной системе счисления. Покажите, что последние знаки в записях a^5 и a совпадают.

10. С помощью теоремы Ферма докажите, что для любого целого n число $n^3 + (n+1)^3 + (n+2)^3$ делится на 9.

11. Пусть p — простое число, отличное от 2 и 5. Покажите, что p делит какое-нибудь число из множества

$$\{1;\ 11;\ 111;\ 1111;\ 11\,111; \ldots\}.$$

Указание: По теореме Ферма число $10^{p-1} - 1$ делится на p, если $p > 5$. Случай $p = 3$ нужно разобрать отдельно.

12. Покажите, что уравнение $x^{13} + 12x + 13y^6 = 1$ не имеет целочисленных решений.

Указание: Редуцируйте уравнение по модулю 13 и примените теорему Ферма.

13. Найдите остатки от деления

(1) $39^{50!}$ на 2251;

(2) 19^{39^4} на 191.

14. Цель этой задачи — показать, что для простого числа вида $p = 4n+1$ найдутся целые a и b, сумма квадратов которых $a^2 + b^2$ делится на p. Пусть x и y — два целых числа, взаимно простых с p. Положим $a = x^n$ и $b = y^n$. Тогда

$$(a^2 - b^2)(a^2 + b^2) = x^{4n} - y^{4n}.$$

(1) Используйте теорему Ферма для проверки делимости $x^{4n} - y^{4n}$ на p.
(2) Учитывая полученный результат, покажите, что p делит либо $a^2 + b^2$, либо $a^2 - b^2$.

Если p делит $a^2 + b^2$, то все доказано. Поэтому, рассуждая от противного, можно считать, что для любых целых x и y разность $x^{2n} - y^{2n} = a^2 - b^2$ делится на p. В частности, это должно быть верно, если x — произвольное целое, а $y = 1$, т.е. $x^{2n} - 1$ делится на p при любом целом x.

(3) Покажите, что сравнения $x^{2n} \equiv 1 \pmod{p}$ (для любого целого n) противоречат теореме § 6.4.
(4) Собирая вместе доказанные утверждения, завершите решение задачи.

Ферма знал более сильный факт: любое простое число вида $4n + 1$ может быть представлено в виде суммы квадратов двух целых чисел (см. [49], [Д.13]). Сравните этот результат с упражнением 13 главы 5.

15. В этом упражнении мы описываем доказательство Эйлера теоремы Ферма. В отличие от доказательства самого Ферма, оно не использует математическую индукцию. Пусть p — простое число и $\overline{a}$ — элемент из $U(p) = \mathbb{Z}_p \setminus \{\overline{0}\}$. Рассмотрим подмножество $S = \left\{\overline{a}, \overline{2a}, \ldots, \overline{(p-1)a}\right\}$.

(1) Покажите, что элементы подмножества S различны.
(2) Покажите, что множество S содержит $p-1$ элемент и выведите отсюда, что $S = U(p)$.
(3) Покажите, что (2) влечет равенство: $\overline{(p-1)!} = \overline{1} \cdot \overline{2} \cdots \overline{(p-1)}$, правая часть которого — произведение всех элементов из S.
(4) Покажите, что, с другой стороны, произведение всех элементов из S можно записать в виде $\overline{a}^{p-1} \cdot \overline{(p-1)!}$.
(5) Из (3) и (4) выведите теорему Ферма.

16. Пусть p — простое, и a — целое число, не делящееся на p. Покажите, что обратным к $\overline{a}$ в $\mathbb{Z}_p$ является элемент $\overline{a}^{p-2}$.

17. Пусть $p = 4k + 3$ — положительное простое число. Для данного целого a рассмотрим уравнение $x^2 \equiv a \pmod{p}$.

(1) Подберите a и p, для которых это уравнение не имеет решений.

(2) Покажите, что если уравнение имеет корень, то он сравним с $\pm a^{k+1}$ по модулю p.

Указание: Если уравнение имеет корень, то найдется такое целое число b, что $b^2 \equiv a \pmod{p}$. Поэтому

$$(a^{k+1})^2 \equiv b^{4(k+1)} \equiv b^{4k+2} \cdot b^2 \pmod{p}.$$

Теперь (2) немедленно следует из теоремы Ферма.

18. Следуя упражнению 16, напишите программу, которая по данным p и a вычисляет обратный к a по модулю p. Прежде всего, программа должна проверять, делится ли a на p.

19. Пусть $p = 4k + 3$ — положительное простое число. Напишите программу, которая по данному p и натуральному числу a находит решение уравнения $x^2 \equiv a \pmod{p}$. Напомним, что из упражнения 17 следует, что если это уравнение имеет решение b, то $b \equiv \pm a^{k+1} \pmod{p}$. Так что программа должна вычислять вычет a^{k+1} по модулю p и проверять, является ли этот вычет решением данного уравнения. В качестве результата программа должна выдавать решение уравнения или сообщение о том, что уравнение решений не имеет. Это частный случай упражнения 8 главы 12.

20. Некоторые задачи теории чисел приводят к простым числам, удовлетворяющим уравнению

$$a^{p-1} \equiv 1 \pmod{p^2}$$

для какого-то целого a. Напишите программу, которая по данным целым r и $a > 1$ ищет все простые p от $a + 1$ до r, удовлетворяющие этому уравнению. Сначала программа должна

с помощью решета Эратосфена определять все простые числа, не превосходящие r. Затем из найденных простых нужно будет выбрать те, которые удовлетворяют выписанному выше сравнению. Между $a+1$ и $r=10^5$, есть только 2 простых числа, удовлетворяющих сравнению, при $a=2,\ 5,\ 10$ и 14; и 5, если $a=19$.

Глава 7.
Псевдопростые числа

Здесь мы увидим, как теорема Ферма помогает выяснить, является ли данное число составным, не раскладывая его на множители. Глава заканчивается обсуждением стратегий различных систем символьных вычислений, используемых для проверки чисел на простоту или разложимость.

§ 7.1. Псевдопростые числа

Как утверждает теорема Ферма, для целого числа a, не делящегося на простое p, имеет место сравнение: $a^{p-1} \equiv 1 \pmod{p}$. Предположим, что нам нужно узнать, является ли данное *нечетное* число n простым. Допустим также, что нам как-нибудь удалось найти целое b, не делящееся на n и удовлетворяющее условию: $b^{n-1} \not\equiv 1 \pmod{n}$. В этом случае теорема Ферма говорит нам, что n не может быть простым. Такое число b будем называть *свидетелем* разложимости n. Итак, у нас есть метод тестирования числа на наличие нетривиальных делителей, при котором не нужно разлагать число на множители. Трудность применения этого теста заключается в том, что он не будет работать до тех пор, пока мы не найдем свидетеля; а это тре-

бует везения. Но как мы дальше увидим, более вероятно, что такой свидетель будет найден, чем не найден.

Заметим, что для поиска свидетеля b нам нет нужды просматривать все целые числа. Действительно, поскольку мы работаем с сравнениями по модулю n, можно ограничить поиск числами b, лежащими в интервале $0 \leqslant b \leqslant n-1$. Имеет смысл исключить еще 0 (поскольку b не должно делиться на n) и 1 (ибо $1^{n-1} \equiv 1 \pmod{n}$ для всех n). Более того, из нечетности n вытекает сравнение: $(n-1)^{n-1} \equiv 1 \pmod{n}$, т.е. сравнение выполнено и для $n-1$. Итак, можно предполагать, что искомый свидетель b удовлетворяет неравенству: $1 < b < n-1$. Прежде, чем мы будем применять этот тест, сформулируем его в виде теоремы.

Тест на разложимость. *Пусть n — нечетное натуральное число. Если найдется такое целое число b, что*

(1) $1 < b < n-1$, *и*
(2) $b^{n-1} \not\equiv 1 \pmod{n}$,

то n — разложимое, т.е. составное число.

Напомним, что повторяющиеся единицы $R(n)$ определяются формулой

$$R(n) = \frac{10^n - 1}{9}.$$

Другими словами, это целые числа, в десятичной записи которых встречаются только 1 (см. упражнение 5 главы 3). Мы уже видели, что $R(n)$ составное число для разложимых n. Однако для простого числа 229 у нас до сих пор не было возможности определить, простое ли $R(229)$ или составное. Кроме того, это число состоит из более чем 200 знаков, так что раскладывать его на множители слишком утомительно. Вместо этого применим тест на разложимость с $b = 2$. С помощью системы символьных вычислений легко найти вычет $2^{R(229)-1}$ по моду-

лю $R(229)$. Он равен

$$104516500584333397781753768885982835488612737233884898$$
$$570848288405666898406290825536552313452374268256539145$$
$$527606121567512885287283062854774198632697829520351103$$
$$663852079821692412346101479040743884170069248576365931$$
$$1045450329217$$

и не сравним с 1 по модулю $R(229)$. Так что $R(229)$ — составное число.

Поскольку нас больше интересуют простые числа, нежели составные, резонно задать вопрос: можно ли использовать теорему Ферма для доказательства простоты чисел? Более точно, предположим, что n — положительное нечетное целое число, для которого $b^{n-1} \equiv 1 \pmod{n}$ при *некотором* целом числе b, удовлетворяющем неравенству $1 < b < n-1$; обязательно ли n будет простым числом? Лейбниц, знаменитый философ и математик, полагал, что ответ утвердительный. Он использовал это соображение как тест на простоту, всегда выбирая 2 в качестве b для упрощения вычислений.

К сожалению, Лейбниц был неправ. Например, имеет место сравнение $2^{340} \equiv 1 \pmod{341}$ и, по Лейбницу, число 341 должно быть простым. Но $341 = 11 \cdot 31$ — составное. Числа, дающие ложный «положительный» результат в этом тесте, известны под именем псевдопростых. Выражаясь точнее, нечетное составное натуральное число n, удовлетворяющее сравнению $b^{n-1} \equiv 1 \pmod{n}$ для некоторого целого b из интервала $(1;\, n-1)$, называется *псевдопростым* по *основанию* b. Следовательно, 341 — псевдопростое по основанию 2.

Несомненно, тест Лейбница все-таки полезен, хотя и не абсолютно точен. Для малых целых, выбранных наугад, тест чаще дает правильный ответ, чем ошибается. Чтобы убедиться в этом, подсчитаем простые и псевдопростые числа по основанию 2, не превышающие какой-нибудь подходящей грани. Например, между 1 и 10^9 лежит 50 847 544 простых и только

5597 псевдопростых по основанию 2. Таким образом, число из этого промежутка, выдержавшее тест Лейбница, будет скорее простым, нежели псевдопростым по основанию 2.

Кроме того, мы применяли этот тест только по одному основанию, а если использовать несколько разных оснований, то количество неопределяемых составных чисел значительно уменьшится. Например, $3^{340} \equiv 56 \pmod{341}$, так что 3 свидетельствует о разложимости числа 341. На самом деле, между 1 и 10^9 находится 1272 псевдопростых по основаниям 2 и 3 и только 685 псевдопростых по основаниям 2, 3 и 5.

Так как нам нужно примерять тест только по конечному числу оснований: 2, 3, ..., $n-2$, возникает вопрос: может ли n быть составным, оставаясь псевдопростым по всем этим основаниям? Пусть $n > 2$ и предположим, что $b^{n-1} \equiv 1 \pmod{n}$ для некоторого $1 < b < n-1$. Поскольку $b^{n-1} = b \cdot b^{n-2}$, то из предположения следует обратимость b по модулю n. А теорема обратимости утверждает, что такое возможно лишь в случае НОД$(b, n) = 1$. Поэтому, если n — составное, а один из делителей b делит n, то $b^{n-1} \not\equiv 1 \pmod{n}$. В частности, любой делитель n свидетельствует о разложимости n. Итак, ответ на вышесформулированный вопрос отрицателен.

Какой можно сделать вывод из результатов этого параграфа? Напомним, что наша цель — найти эффективный способ определения, является ли данное конкретное число простым. Если число имеет небольшой делитель, его можно найти с помощью алгоритма деления методом проб из главы 3. Так что на практике тест на разложимость следует применять только в том случае, когда предварительные рассмотрения показали, что данное число не имеет малых делителей. Таким образом, вопрос, заданный выше, имеет небольшой практический интерес. Было бы быстрее найти делитель, чем пытаться проверить, является ли большое число n псевдопростым по всем основаниям между 2 и $n-2$. Однако, как мы увидим в следующем параграфе, это не конец истории.

Перед тем, как мы к нему перейдем, может быть полезно заметить, что в некоторых книгах количество $\pi(10^9)$ простых чисел, не превосходящих 10^9, считается меньшим, чем приведенное нами. Это не опечатка, поскольку все эти книги приводят одно и то же число. Ошибочная информация о количестве простых обязана своим появлением датскому математику Бертельсену (Bertelsen), который в 1893 году насчитал на 56 простых чисел меньше, чем в действительности. По иронии судьбы, предполагалось, что его вычисления направлены на исправления неточностей в неких таблицах. Вместо этого он допустил ошибку, которую можно найти в книгах, опубликованных до 1993 года.

§ 7.2. Числа Кармайкла

Как мы показали в конце последнего параграфа, составное число n не является псевдопростым по основанию b, если числа n и b имеют общий нетривиальный делитель. К сожалению, эта информация не очень полезна. Практически, чтобы ограничить объем вычислений разумными рамками, мы выделяем несколько оснований среди небольших простых чисел. И если наименьший делитель числа n очень большой, то все выбранные основания будут взаимно просты с n. Поэтому вопрос, который нам стоило бы задать, нужно нацелить на усовершенствование метода, приведенного в конце предыдущего параграфа. А именно, может ли составное нечетное число n быть псевдопростым по всем *взаимно-простым с* ним основаниям b? Забегая вперед, скажем, что ответ: «да».

Заметим, что если число b взаимно просто с n, то сравнение $b^n \equiv b \pmod{n}$ равносильно $b^{n-1} \equiv 1 \pmod{n}$. Это позволяет поставить вопрос чуть более строго: существует ли такое нечетное натуральное n, которое, будучи *составным*, удовлетворяет сравнениям $b^n \equiv b \pmod{n}$ для всех целых b? Одно из преимуществ такой формулировки вопроса заключа-

ется в отсутствии необходимости каких-либо предварительных предположений относительно b. Первым привел пример таких чисел n математик Р. Д. Кармайкл (Carmichael) в своей работе, опубликованной в 1912 году ([10]). Поэтому они и называются числами Кармайкла.

Поскольку эти числа играют важную роль во многом из того, о чем нам еще предстоит говорить, хорошо бы дать их формальное определение. Нечетное натуральное n называется *числом Кармайкла*, если оно составное и $b^n \equiv b \pmod{n}$ для всех целых b. Конечно, достаточно проверить это сравнение только для чисел, удовлетворяющих неравенству $1 < b < n-1$, поскольку мы работаем по модулю n.

Как показал сам Кармайкл, наименьшее из чисел, открытых им, равно 561. В принципе, мы можем проверить этот факт, исходя из определения. Однако даже для относительно небольших чисел такая процедура очень длинна и скучна. Действительно, чтобы доказать прямо из определения, что число 561 — число Кармайкла, нам потребуется проверять истинность сравнения $b^{561} \equiv b \pmod{561}$ для $b = 2, 3, 4, \ldots, 559$, т.е. всего — 557 раз. Это может показаться не такой уж тяжелой работой, если у Вас есть компьютер, а что делать, когда у Вас такой кандидат на число Кармайкла:

$$349\,407\,515\,342\,287\,435\,050\,603\,204\,719\,587\,201?$$

Самое подходящее время вернуться к доске и мелу (т.е. теоретическим изысканиям).

Попытаемся найти обходной путь для доказательства того, что 561 — число Кармайкла. Прежде всего заметим, что оно довольно легко раскладывается на простые множители:

$$561 = 3 \cdot 11 \cdot 17.$$

Теперь мы хотим проверить сравнение

$$b^{561} \equiv b \pmod{561} \tag{2.1}$$

для некоторого целого b. Наша стратегия состоит в демонстрации делимости разности $b^{561} - b$ на 3, 11 и 17. Так как это *различные* простые числа, то лемма из главы 3 говорит нам, что тогда их произведение тоже должно делить $b^{561} - b$. Но упомянутое произведение равно 561, так что (2.1) этим будет доказано.

Чтобы сделать нашу стратегию рабочей, нужно исхитриться доказать, что $b^{561} - b$ делится на каждый простой множитель числа 561. В этом нам поможет теорема Ферма. Мы приведем подробное доказательство делимости разности на 17, оставив случаи делимости на 3 и на 11 в качестве полезных и необременительных упражнений. Итак, мы хотим доказать, что 17 делит разность $b^{561} - b$, или, на языке сравнений:

$$b^{561} \equiv b \pmod{17}. \tag{2.2}$$

Необходимо рассмотреть два случая.

1) Число 17 делит b. В этой ситуации обе части уравнения (2.2) сравнимы с 0 по модулю 17, т.е. сравнение справедливо.

2) Число 17 не делит b. Тогда теорема Ферма влечет равенство: $b^{16} \equiv 1 \pmod{17}$. Прежде чем применить это наблюдение к сравнению (2.2), нам нужно найти остаток от деления 561 на 16. Но $561 = 35 \cdot 16 + 1$, поэтому

$$b^{561} \equiv (b^{16})^{35} \cdot b \equiv b \pmod{17}.$$

Заметим, теорема Ферма так сильно сократила наши вычисления благодаря тому, что остаток от деления 561 на 16 оказался равным 1. Удачно, что остатки от деления 561 на $2(= 3 - 1)$ и на $10(= 11 - 1)$ тоже равны 1. Так что вычисления, которые Вам предстоит сделать для 3 и 11 — дословное повторение проведенных.

Успех нашей стратегии обязан двум свойствам числа 561. Первое: деление 561 на разность каждого из его делителей и единицы дает в остатке 1. Второе: каждый простой делитель

числа 561 входит в его разложение на простые множители с кратностью 1. Что же это: нам очень повезло с выбором примера, или числа Кармайкла встречаются крайне редко? Реальность оказывается еще более удивительной. Существует бесконечно много чисел Кармайкла, и все они обладают свойствами, столь облегчившими наши вычисления с 561. Характеристика чисел Кармайкла, вытекающая из этого наблюдения, впервые была дана А. Корселтом (A. Korselt) за пятнадцать лет до публикации работы Кармайкла на эту тему. Однако, Корселт не привел никаких новых примеров чисел, которые удовлетворяли бы описанным им свойствам.

Теорема Корселта. *Нечетное натуральное число n является числом Кармайкла, если и только если для каждого его простого делителя p выполнены следующие два условия:*

(1) p^2 *не делит* n*;*

(2) $p-1$ *делит* $n-1$.

Покажем для начала, что если число n удовлетворяет условиям (1) и (2) теоремы, то оно является числом Кармайкла. Для этого мы применим стратегию, разработанную в предыдущих вычислениях с числом 561. Предположим, что p — простой делитель числа n. Покажем, что

$$b^n \equiv b \pmod{p}. \tag{2.3}$$

Если b делится на p, то обе части в (2.3) сравнимы с нулем по модулю p, и доказывать нечего. Предположим теперь, что p не делит b. Как следует из теоремы Ферма, $b^{p-1} \equiv 1 \pmod{p}$. Прежде чем применить это соображение к (2.3), мы должны найти остаток от деления n на $p-1$. Но $p-1$ делит $n-1$ согласно условию (2) теоремы, т.е. $n-1 = (p-1)q$ для некоторого целого q и

$$n = (n-1)+1 = (p-1)q+1.$$

Значит

$$b^n = (b^{p-1})^q \cdot b \equiv b \pmod{p},$$

где второе сравнение следует из теоремы Ферма. Суммируя все вышесказанное, получаем, что если p простой делитель числа n, то $b^n \equiv b \pmod{p}$ для любого целого b.

Ввиду условия (1) теоремы, $n = p_1 \cdots p_k$, где $p_1, \ldots, p_k$ — попарно различные простые числа. Мы уже видели, что $b^n - b$ делится на каждое из этих чисел, а поскольку все они различны, то, учитывая лемму § 3.6, можно сделать вывод: $b^n - b$ делится на их произведение $p_1 \cdots p_k = n$. Иначе говоря, $b^n \equiv$ $\equiv b \pmod{n}$. Поскольку вычисления справедливы для любого целого b, то n является числом Кармайкла.

Теперь следует показать, что всякое число Кармайкла удовлетворяет условиям (1) и (2) теоремы. Сделаем это методом «от противного». Сначала, предположив, что n число Кармайкла, получим, что делимость n на p^2 приводит к противоречию. Этим мы докажем, что числа Кармайкла удовлетворяют условию (1) теоремы.

Так как n — число Кармайкла, противоречие получится, если мы найдем целое b, для которого $b^n \not\equiv b \pmod{n}$. Положим $b = p$. Тогда

$$p^n - p = p(p^{n-1} - 1).$$

Но p не делит $p^{n-1}-1$, поэтому p^2 не может делить p^n-p. Иначе говоря, $p^n \not\equiv p \pmod{n}$. Это противоречит предположению о том, что n — число Кармайкла.

Для завершения доказательства нам осталось показать, что числа Кармайкла удовлетворяют условию (2) теоремы. Однако для этого необходима *теорема о примитивных корнях*, которая будет доказана только в § 11.3.

К сожалению, для проверки данного натурального n на принадлежность к числам Кармайкла с использованием теоремы Корселта, нам нужно разложить число на простые множители, что довольно сложно в случае больших чисел. Тем не менее, довольно часто бывает так, что очень большие числа Кармайкла имеют много небольших делителей. Например, число с 36 знаками, приведенное в начале этого параграфа, являет-

ся наименьшим числом Кармайкла с 20 простыми делителями. Его разложение на множители выглядит следующим образом:

$$11{\cdot}13{\cdot}17{\cdot}19{\cdot}29{\cdot}31{\cdot}37{\cdot}41{\cdot}43{\cdot}61{\cdot}71{\cdot}73{\cdot}97{\cdot}101{\cdot}109{\cdot}113{\cdot}151{\cdot}181{\cdot}193{\cdot}641.$$

Используя одну из программ символьных вычислений, теперь можно быстро проверить, что это действительно число Кармайкла. Другие примеры чисел Кармайкла можно найти в упражнении 3, где приведено семейство целых, содержащее несколько таких чисел.

В своей работе 1912 года Кармайкл выписал 15 чисел, названных его именем, а затем добавил: «этот список можно продолжать бесконечно». То есть он имел в виду, что существует бесконечно много чисел Кармайкла. Однако скоро стало ясно, что доказательство этого утверждения — очень трудная проблема. Причина высокой сложности задачи в том, что такие числа достаточно редко встречаются. Например, между 1 и 10^9 лежит только 646 чисел Кармайкла, в то время как простых — 50 847 534. Проблема, наконец, была решена Альфордом (Alford), Гранвиллем (Granville) и Померанцем (Pomerance) в 1994 году. Они показали, что на самом деле существует бесконечно много чисел Кармайкла. Побочный продукт этого результата имеет отношение к тестированию простоты и будет обсуждаться в § 7.4.

§ 7.3. Тест Миллера

В § 7.1 мы видели, что теорема Ферма предлагает способ проверки разложимости данного числа, не требуя поиска его делителей. Однако этот подход не всегда работает и, если очень не повезет, может потерпеть неудачу. Далее (§ 7.2) мы пытались объяснить, что означает «невезение» в данном контексте. Было обнаружено, что числа Кармайкла ведут себя как простые, в том смысле, что мы не можем установить

их разложимость методом, развитым в конце параграфа. Но этот тест можно так усовершенствовать, что его не смогут обмануть даже числа Кармайкла. Новый тест ввел Миллер (G. L. Millller) в 1976 году.

Пусть $n > 0$ — нечетное целое. Выберем целое число b, удовлетворяющее неравенству $1 < b < n-1$, и назовем его, как и раньше, основанием. Поскольку n — нечетное, $n-1$ должно быть четным числом. Первый шаг теста Миллера состоит в определении такого показателя $k \geqslant 1$, для которого $n-1 = 2^k q$, где q не делится на 2. Иначе говоря, мы должны найти наибольшую степень двойки, делящую $n-1$, и соответствующее частное q.

Далее тест вычисляет вычеты по модулю n у следующей последовательности степеней:

$$b^q,\ b^{2q},\ \ldots,\ b^{2^{k-1}q},\ b^{2^k q}.$$

Разберемся, какими свойствами обладает эта последовательность в случае простого n. Итак, пока не будет оговорено особо, n — *простое* число. Теорема Ферма говорит нам, что

$$b^{2^k q} \equiv b^{n-1} \equiv 1 \pmod{n}.$$

Значит, если n — простое, то последний вычет в последовательности всегда равен 1. Конечно, единица среди вычетов может встретиться и раньше. Пусть j — *наименьший* показатель, для которого $b^{2^j q} \equiv 1 \pmod{n}$. Если $j \geqslant 1$, то

$$b^{2^j q} - 1 = \left(b^{2^{j-1}q} - 1\right)\left(b^{2^{j-1}q} + 1\right).$$

По предположению, число n — простое и (так как оно делит разность квадратов $b^{2^j q}-1$) делит либо $b^{2^{j-1}q}-1$, либо $b^{2^{j-1}q}+1$. С другой стороны, в силу выбора показателя j, число n не может делить $b^{2^{j-1}q} - 1$. Остается только одна возможность: n — делитель числа $b^{2^{j-1}q} + 1$, т.е. $b^{2^{j-1}q} \equiv -1 \pmod{n}$.

Эти рассуждения показывают, что в случае *простого* n среди последовательности степеней:

$$b^q,\ b^{2q},\ \ldots,\ b^{2^{k-1}q}$$

найдется по крайней мере одна, сравнимая с -1 по модулю n. Хорошо, но не слишком. Дело в том, что наше рассуждение основывалось на предположении: $j > 0$. Если $j = 0$, то $b^q \equiv \equiv 1 \pmod{n}$. А у нас нет простого способа разложения числа $b^q - 1$ на множители, поскольку q нечетно. Значит, если n — *простое*, то с последовательностью вычетов по модулю n, о которой мы говорили, должно произойти одно из двух: либо первый же вычет равен 1, либо среди них появится $n-1$. В противном случае (не будет ни того, ни другого) число n должно быть составным.

Последовательность вычетов, используемая в тесте Миллера, довольно легко вычисляется, потому что каждый вычет (за исключением первого) — квадрат предыдущего. В самом деле, $b^{2^j q} = \left(b^{2^{j-1}q}\right)^2$ при $j \geqslant 1$. Отсюда вытекает, что как только в последовательности вычетов по модулю n встретится $n-1$, все остальные вычеты будут равны 1.

У нас появился еще один тест, который позволяет нам показать разложимость данного числа, но работающий только при удачном стечении обстоятельств. Тем не менее, тест Миллера более эффективен, нежели предложенный в § 7.1. Чтобы понять почему, заметим, что последовательность степеней, выписанная для псевдопростого n по основанию b, должна иметь член, сравнимый с 1 по модулю n. А так как n не простое, то существует хороший шанс, что этой степени не будет предшествовать другая, сравнимая с $n-1$. В этом случае тест Миллера обнаружит разложимость числа. Приведем алгоритм, реализующий тест Миллера.

Тест Миллера

Ввод: нечетное натуральное n и основание b, где $1 < b < n-1$.

Вывод: одно из двух сообщений: «n составное» или «ничего определенного сказать нельзя».

Шаг 1. Последовательно делим $n-1$ на 2 пока не получим нечетного частного. В результате найдем положительное целое k и нечетное q, для которых $n-1=2^k q$.
Шаг 2. Присвоим i нулевое значение, а r значение вычета b^q по модулю n.
Шаг 3. Если $i=0$ и $r=1$, или $i>0$, а $r=n-1$, то вывести сообщение: «ничего определенного сказать нельзя»; в противном случае переходим к шагу 4.
Шаг 4. Увеличиваем i на 1 и заменяем r на r^2 по модулю n; переходим к шагу 5.
Шаг 5. Если $i<k$, то возвращаемся к шагу 3; в противном случае выдаем сообщение: «n составное».

В случае неопределенного сообщения возможны две ситуации: либо n простое, либо составное. К сожалению, второй случай реально встречается. Рассмотрим несколько примеров, причем начнем с хороших новостей. Мы видели в § 7.1, что 341 — псевдопростое по основанию 2, так что это хороший объект для теста Миллера. Прежде всего, $340=2^2\cdot 85$. Теперь нам нужно найти вычеты по модулю 341 у степеней 2 с показателями 85 и 170:

$$2^{85}\equiv 32 \pmod{341}, \qquad 2^{170}\equiv 32^2\equiv 1 \pmod{341}.$$

Это позволяет тесту дать заключение: *число составное*.

Более наглядный пример дает нам число Кармайкла 561. Подвергнем его тесту Миллера по основанию 2. Простые вычисления показывают, что $560=2^4\cdot 35$. Выпишем последовательность вычетов степеней 2 по модулю 561:

Степени	35	$2\cdot 35$	$2^2\cdot 35$	$2^3\cdot 35$
Вычеты	263	166	67	1

И в этом случае тест сообщит нам о разложимости числа.

Хотя 561 — число Кармайкла, мы обнаружили, что оно составное, применив наименьшее из возможных оснований.

Теперь плохие новости. Применим тест Миллера по основанию 7 к 25. Поскольку $24 = 2^3 \cdot 3$, последовательность степеней и их остатков имеет вид:

Степени	3	$2 \cdot 3$	$2^2 \cdot 3$
Вычеты	18	24	1

Так что тест здесь не может сказать ничего определенного, несмотря на то, что разложимость числа 25 видна невооруженным взглядом. Конечно, основание 7 мы выбрали не случайно, если бы основанием было 2, то тест бы узнал в двадцати пяти составное число.

Пусть $n > 0$ — нечетное целое число и $1 < b < n - 1$. Если n составное, а тест Миллера его не распознал, то оно называется *строго псевдопростым* по основанию b. Вышеприведенный пример показывает, что 25 — строго псевдопростое по основанию 7. Легко увидеть, что строго псевдопростое число по основанию b является псевдопростым по этому же основанию (см. упражнение 7).

Как мы уже отметили, число 25 не является строго псевдопростым по основанию 2. Наименьшее строго псевдопростое число по основанию 2 — это 2047. Более того, существует только 1282 строго псевдопростых чисел по основанию 2, лежащих между 1 и 10^9, что дает хорошее представление об эффективности данного теста. Конечно, можно применять тест Миллера по нескольким основаниям, что существенно увеличит его эффективность. Например, наименьшее строго псевдопростое одновременно по основаниям 2, 3 и 5 — это 25 326 001.

Более того, не существует «строго Кармайкловских чисел». Это следует из результата М. О. Рабина (Rabin).

Теорема Рабина. *Пусть $n > 0$ — нечетное натуральное число. Если тест Миллера, примененный к n для более, чем*

$n/4$ оснований, лежащих между 1 и $n-1$, не распознает в n составного числа, то n — простое.

За подробностями можно обратиться к [39], или к [28]. Не стоит и говорить, что в случае больших n применение теста Миллера по $n/4$ основаниям займет очень много времени. Вопреки этому замечанию, наиболее практичные тесты на простоту, как мы увидим в следующем параграфе, основываются на теореме Рабина.

§ 7.4. Тестирование простоты и системы символьных вычислений

Многие системы символьных вычислений имеют простую команду для проверки, является ли данное число простым. Самое удивительное при этом, что ответ выдается почти мгновенно, даже если проверяемое число было довольно большим. Так происходит из-за того, что большинство программ основаны на тесте Миллера, применяемом по большому числу оснований.

Рациональное зерно в таком использовании теста Миллера прорастает из теоремы Рабина. Пусть n — нечетное *составное* число. Основание b между 1 и $n-1$ выберем случайным образом. Из теоремы Рабина следует, что вероятность неопределенного результата при применении теста Миллера к n и b не превосходит

$$\frac{n/4}{n} = \frac{1}{4}.$$

Так что правдоподобно предположение: число n в этом случае будет составным с вероятностью $1/4$. Если теперь выбрать k различный оснований, то вероятность станет равной $1/4^k$. Поэтому, беря все больше и больше оснований, мы можем сделать вероятность сколь угодно малой.

Эти рассуждения приводят к *вероятностному тесту Рабина на простоту*. Поскольку он вероятностный, нам нужно решить, насколько малой должна быть вероятность ошибки. Предположим, нам нужно, чтобы вероятность ошибки не превосходила ε, и пусть k — такое натуральное число, для которого $1/4^k < \varepsilon$. Тогда тест Рабина выбирает k оснований и применяет тест Миллера по каждому из них. Из вышеприведенных рассуждений вытекает, что в случае неопределенного ответа на тест Миллера по всем этим основаниям вероятность того, что проверяемое число все же окажется составным, будет меньше или равна $1/4^k$, т.е. меньше требуемого ε. Как же выбираются основания? Конечно, удобнее было бы брать основания поменьше, иначе вычисления, необходимые для теста Миллера, будут занимать слишком много времени. Как правило, выбирают первые k положительных простых чисел.

Естественно, мы хотим быть абсолютно уверены в том, что числа, прошедшие тест, будут простыми. Для этого необходимо выбрать ε очень маленьким. Например, пусть $\varepsilon = 10^{-20}$. Так как $1/4^{40}$ — величина порядка 10^{-24}, нам нужно выбрать 40 оснований, чтобы вероятность ошибки была меньше 10^{-20}. Допустим, что в качестве оснований выбрали 40 первых простых чисел. К сожалению, есть число Кармайкла (из 397 знаков), для которого тест Миллера дает неопределенный ответ, если в качестве оснований берутся простые числа, не превосходящие 300 (см. [5]). А так как существует ровно 62 таких простых числа, тест, проверяя упомянутое число Кармайкла по всем основаниям, которые мы выберем, будет выдавать неопределенный ответ!

Посмотрим как выполняется тест Рабина в одной широко известной системе символьных вычислений. Конечно, каждая программа использует ей одной свойственную стратегию. Например, *Maple* V.2[1] проверяет простоту в три этапа. Сначала

[1] *Maple*™ — пакет компьютерных программ по символьным вычислениям, разработанный Waterloo Maple Software, Inc.

подбираются простые делители тестируемого числа, не превосходящие 10^3. Если таких делителей не обнаружено, то программа применяет тест Миллера по основаниям $2, 3, 5, 7$ и 11. А в последнюю очередь проверяется, что данное число не относится ни к одному из следующих семейств:

$$(u+1)(k\frac{u}{2}+1) \qquad \text{для} \qquad 3 \leqslant k \leqslant 9,$$

или

$$(u+1)(ku+1) \qquad \text{для} \qquad 5 \leqslant k \leqslant 20.$$

Причиной необходимости последнего этапа служит информация о том, что в этих семействах найдено довольно много строго псевдопростых чисел по тем основаниям, которые используются программой (см. [38]). Тем не менее, составное число

$$12\,530\,759\,607\,784\,496\,010\,584\,573\,923$$

определяется Maple'ом как простое. Его наименьший простой делитель равен $286\,472\,803$. В более поздних версиях Maple'а тест на простоту был модифицирован, так что теперь это число правильно идентифицируется как составное.

Axiom 1.1[2] применяет другую стратегию: подбирает количество оснований в зависимости от числа, которое собирается тестировать. Было показано, что тест, используемый *Axiom 1.1*, правильно обнаруживает простоту чисел, если они меньше, чем $341\,550\,071\,728\,321$ (см. [26]). Для чисел, выходящих за эту границу, *Axiom 1.1* в качестве оснований для теста Миллера использует 10 наименьших простых чисел. Подобно *Maple*, эта программа делает дополнительные проверки для чисел, считающихся особенно неприятными. И этот тест не совершенен; он дает сбой на составном числе из 56 знаков.

Сначала можно подумать, что стратегии, выбранные для этих программ, дадут «совершенные» тесты, стоит только подобрать подходящие основания. Но печальная истина в том,

[2] *Axiom* — зарегистрированная торговая марка NAG (Numarical Algorithms Group), Ltd.

что это невозможно сделать даже теоретически. Одно из следствий, вытекающих из работы Альфорда, Гранвилля и Померанца над числами Кармайкла, состоит в следующем:

> «Для любого конечного набора оснований найдется бесконечно много чисел Кармайкла, строго псевдопростых по всем этим основаниям.»

Итак, следует остерегаться категоричных утверждений о простоте числа, руководствуясь тестом Миллера, применяемым по фиксированному числу оснований. Возможный выход из создавшейся ситуации был реализован в *Axiom 2.2*. Теперь программа, учитывая размер тестируемого числа, увеличивает число оснований. Для числа из $2k$ десятичных знаков программа берет примерно k оснований, повышая тем самым точность теста.

Подробности о тестировании простых чисел этими программами, а также примеры, на которых они дают сбой, можно найти в [5]. В главе 11 мы будем изучать тесты, которые позволят нам с уверенностью определять простоту чисел. Но если быть до конца откровенным, они не такие простые и эффективные, как тест Миллера.

Упражнения

1. Какие из чисел: 645, 567 и 701 являются псевдопростыми по основанию 2? Какие из них псевдопросты по основанию 3? Какие просты?

2. Покажите, что если n — псевдопростое число одновременно по основаниям a и ab, то оно псевдопростое и по основанию b.

3. Пусть n — натуральное число. Положим $p_1 = 6n + 1$, $p_2 = 12n + 1$ и $p_3 = 18n + 1$. Покажите, что если p_1, p_2 и

p_3 — простые числа, то произведение $p_1p_2p_3$ является числом Кармайкла. Покажите также, что эти условия выполнены для $n = 1,\ 6$ и 35. Какие числа Кармайкла получаются при таких значениях n?

4. Разложите число 29 341 на множители и покажите, что оно — число Кармайкла.

5. Пусть $p_1 < p_2$ — нечетные простые числа. Положим $n = p_1p_2$ и допустим, что как $p_1 - 1$, так и $p_2 - 1$ делит $n - 1$. Покажите, что в этом случае $n - 1 \equiv p_1 - 1 \pmod{p_2 - 1}$. Воспользуйтесь этим фактом для получения противоречия. Сделайте отсюда вывод, что числа Кармайкла должны иметь более двух простых сомножителей.

6. Какие из чисел: 645, 2047 и 2309 строго псевдопростые по основанию 2? Какие из них строго псевдопростые по основанию 3? Какие простые?

7. Покажите, что если нечетное натуральное число n *строго* псевдопростое по основанию b, то оно и псевдопростое по этому основанию.

8. Напишите программу, находящую все псевдопростые числа по основаниям 2 и 3, не превосходящие 10^6. Напомним, что n — псевдопростое по основаниям 2 и 3, если оно нечетное, составное и удовлетворяет соотношениям:

$$2^{n-1} \equiv 1 \pmod{n} \qquad \text{и} \qquad 3^{n-1} \equiv 1 \pmod{n}.$$

Поэтому программа будет проверять только нечетные составные числа. Один из способов нахождения таких чисел состоит в применении решета Эратосфена к нечетным целым, не превышающим 10^6. Только отбрасывать нужно простые числа, оставляя составные. Сколько из найденных Вами псевдопростых чисел являются числами Кармайкла?

9. Напишите программу, которая находила бы все числа Кармайкла, являющиеся произведением d простых множителей, каждый из которых не превосходит 10^3. Основная проблема при этом (даже для относительно небольших значений d) заключается в том, что числа Кармайкла будут довольно большими. Поэтому для простого делителя p числа n, проверяя делимость $n-1$ на $p-1$, необходимо использовать сравнения. Более того, если программа сгенерировала число $n = p_1 p_2 \cdots p_d$, то вычет n по модулю $p_i - 1$ вычисляется как произведение вычетов множителей n и редукцией этого произведения по модулю $p_i - 1$. Поскольку все простые делители числа n меньше, чем 10^3, такой подход позволяет производить операции над числами, не прибегая к специальному программному обеспечению. Примените написанную программу для вычисления всех чисел Кармайкла с d сомножителями, не превосходящими 10^3, для $3 \leqslant d \leqslant 8$. Для этого не нужно перемножать найденные множители. Достаточно выписать их набор для каждого числа, найденного программой.

10. Напишите программу для определения наименьшего псевдопростого числа по данному основанию. Ее исходными данными будет целое число $b \geqslant 2$. В программе следует применять тест Миллера (по основанию b) ко всем составным нечетным числам до тех пор, пока Вы не получите сообщения о неопределенном ответе. Это будет наименьшее строго псевдопростое число по основанию b. Разумеется, область поиска будет ограничена наибольшим целым числом K, поддерживаемым языком программирования, который Вы выбрали. Ваша программа должна выдавать два возможных результата: наименьшее строго псевдопростое число по основанию b или сообщение: «строго псевдопростых чисел по основанию b, меньших K, не существует». Чтобы найти все нечетные составные числа, меньшие K, можете использовать решето Эратосфена. Примените программу для вычисления наименьшего псевдопростого числа по основаниям 2, 3, 5 и 7.

11. Напишите программу для определения псевдопростых чисел по основанию 2, равных квадрату простого числа p, при условии, что $p < r = 5 \cdot 10^4$. Программа будет использовать решето Эратосфена для поиска всех простых $p \leqslant r$, а затем подставлять каждое простое число в сравнение $2^{p^2} \equiv 2 \pmod{p^2}$. Существует только два примера квадратично псевдопростых по основанию 2, удовлетворяющих наложенным ограничениям.

Глава 8.

Системы сравнений

В этой главе мы изучаем метод решения систем линейных сравнений, называемый *китайским алгоритмом остатков*. В последнем параграфе мы увидим, как этот алгоритм применяется для передачи ключа к шифру нескольким людям.

§ 8.1. Линейные уравнения

Начнем со случая одного линейного уравнения:

$$ax \equiv b \pmod{n}, \tag{1.1}$$

в котором n — натуральное число. В § 5.7 мы видели, что это уравнение легко решается, если НОД$(a, n) = 1$. По теореме обратимости, взаимная простота чисел a и n влечет обратимость $\overline{a}$ в $\mathbb{Z}_n$. Пусть $\overline{\alpha}$ — соответствующий обратный элемент. Умножая на него сравнение (1.1), мы получаем:

$$\alpha(ax) \equiv \alpha b \pmod{n}.$$

Так как $\alpha a \equiv 1 \pmod{n}$, отсюда следует сравнение

$$x \equiv \alpha b \pmod{n},$$

что дает решение уравнения. В частности, если n — простое и $a \not\equiv 0 \pmod{n}$, уравнение (1.1) всегда имеет решение.

Предположим теперь, что $\overline{a}$ *не* обратим в $\mathbb{Z}_n$. Напомним, что это равносильно условию НОД$(a, n) \neq 1$. Наличие решения у (1.1) означает, что найдутся элементы $x, y \in \mathbb{Z}$, для которых

$$ax - ny = b, \tag{1.2}$$

а это возможно только если НОД(a, n) делит b. Итак, если уравнение (1.1) имеет решение, то b делится на НОД(a, n). Разумеется, в случае обратимости класса $\overline{a}$ в $\mathbb{Z}_n$, необходимое условие выполнено, поскольку тогда НОД$(a, n) = 1$.

Проверим, что обратное тоже верно. Пусть $d =$ НОД(a, n) делит b. Тогда $a = da'$, $b = db'$ и $n = dn'$ для некоторых натуральных a', b' и n'. Сокращение равенства (1.2) на d дает:

$$a'x - n'y = b',$$

что равносильно сравнению $a'x \equiv b' \pmod{n'}$. Отметим, что сравнение берется по модулю n' — делителю исходного модуля n. Более того, НОД$(a', n') = 1$, так что новое сравнение должно иметь решение. Итак, мы доказали, что если НОД(a, n) делит b, то множество решений уравнения (1.1) непусто.

Суммируя вышесказанное, заключаем: уравнение (1.1) имеет решение тогда и только тогда, когда НОД(a, n) делит b (см. упражнение 7 главы 2). Кроме того, приведенный метод решения линейных сравнений легко применим, поскольку использует только расширенный алгоритм Эвклида. Однако, когда решения будут получены, у нас могут возникнуть несколько поводов для удивления.

Решим сравнение $6x \equiv 4 \pmod{8}$. Поскольку НОД$(6, 8) = 2 \neq 1$, то $\overline{6}$ не имеет обратного в $\mathbb{Z}_8$. Если данное сравнение имеет решение, то найдутся такие целые x и y, для которых $6x - 8y = 4$. Разделим это равенство на 2: $3x - 4y = 2$, что равносильно сравнению $3x \equiv 2 \pmod{4}$. Но $\overline{3}$ сам себе обратен в $\mathbb{Z}_4$. Умножая последнее сравнение на 3, мы приходим к

решению:

$$x \equiv 2 \pmod 4. \tag{1.3}$$

Это не совсем то, что нужно. Действительно, мы начинали со сравнения по модулю 8, и решение тоже хотели бы найти по модулю 8, а не по модулю 4, как в (1.3). Беда легко поправима. Как следует из (1.3), решение x сравнения $6x \equiv 4 \pmod 8$ записывается в виде $x = 2 + 4k$ для некоторого $k \in \mathbb{Z}$. Если k четно, то $x \equiv 2 \pmod 8$ — одно из решений. С другой стороны, если k нечетно, то $k = 2m + 1$ и $x = 6 + 8m$. Так что $x \equiv 6 \pmod 8$ — другое решение. Более того, поскольку k может быть либо четным, либо нечетным, существуют только эти возможности. Следовательно, уравнение $\overline{6} \cdot \overline{x} = \overline{4}$ имеет ровно два разных решения в $\mathbb{Z}_8$, а именно $\overline{2}$ и $\overline{6}$. Это пример линейного сравнения с *двумя* решениями. Как мы видели в §6.4, такое произошло потому, что модуль сравнения был составным.

§8.2. Астрономический пример

В этом параграфе мы описываем один из методов решения систем линейных сравнений. Это очень древний алгоритм. Он применялся еще в античности для решения проблем астрономии. Мы начнем с задачи, сформулированной на современном языке, которая могла бы рассматриваться древними астрономами.

> Три спутника пересекут меридиан города Лидса сегодня ночью: первый — в 1 ночи, второй — в 4 утра, а третий — в 8 утра. У каждого спутника свой период обращения. Первому на полный оборот вокруг Земли требуется 13 часов, второму — 15, а третьему — 19 часов. Сколько часов пройдет (от полуночи) до того момента, когда спутники одновременно пересекут меридиан Лидса?

Посмотрим, как эта задача переводится на язык сравнений. Пусть x — количество часов, которые пройдут с 12 часов ночи до момента одновременного прохождения спутниками над меридианом Лидса. Первый спутник пересекает этот меридиан каждые 13 часов, начиная с часу ночи. Это можно записать как $x = 1 + 13t$ для некоторого целого t. Другими словами, $x \equiv 1 \pmod{13}$. Соответствующие уравнения для остальных спутников имеют вид:

$$x \equiv 4 \pmod{15} \qquad \text{и} \qquad x \equiv 8 \pmod{19}.$$

Таким образом, три спутника одновременно пересекут меридиан Лидса через x часов, если x удовлетворяет эти трем уравнениям. Следовательно, для ответа на поставленный вопрос достаточно решить систему сравнений:

$$\begin{cases} x \equiv 1 \pmod{13}, \\ x \equiv 4 \pmod{15}, \\ x \equiv 8 \pmod{19}. \end{cases} \tag{2.1}$$

Заметим, что мы не можем складывать или вычитать уравнения системы, поскольку модули сравнений в них разные. Будем решать эту задачу, переходя от сравнений к уравнениям в целых числах. Так, сравнение $x \equiv 1 \pmod{13}$ соответствует диофантову уравнению: $x = 1 + 13t$. Заменяя x во втором сравнении системы на $1 + 13t$, получаем:

$$1 + 13t \equiv 4 \pmod{15}, \qquad \text{т.е.} \qquad 13t \equiv 3 \pmod{15}.$$

Но 13 обратимо по модулю 15, обратный к нему элемент — это 7. Умножая последнее сравнение на 7 и переходя в нем к вычетам по модулю 15, имеем:

$$t \equiv 6 \pmod{15}.$$

Значит, t может быть записан в виде: $t = 6 + 15u$ для какого-то целого u. Следовательно,

$$x = 1 + 13t = 1 + 13(6 + 15u) = 79 + 195u.$$

Заметим, что все числа вида $79 + 195u$ являются целыми решениями первых двух сравнений системы (2.1). Наконец, подставим в третье сравнение вместо x выражение $79 + 195u$:

$$79 + 195u \equiv 8 \pmod{19}, \qquad \text{так что} \qquad 5u \equiv 5 \pmod{19}.$$

Ввиду обратимости остатка 5 по модулю 19, на него можно сократить и увидеть, что $u \equiv 1 \pmod{19}$. Переписывая это сравнение как диофантово уравнение, мы получим $u = 1 + 19v$ для некоторого целого v. Итак,

$$x = 79 + 195u = 79 + 195(1 + 19v) = 274 + 3705v.$$

Какой отсюда можно сделать вывод относительно спутников? Напомним, что x — количество часов, которые пройдут от полуночи до момента одновременного прохождения спутников над меридианом Лидса. Поэтому нам нужно было найти наименьшее натуральное значение переменной x, удовлетворяющее системе (2.1). Мы это сделали. Поскольку решение системы: $x = 274 + 3705v$, то ответ: 274. Итак, спутники одновременно пройдут над меридианом Лидса через 274 часа после 0 часов сегодняшней ночи, что соответствует 11 дням и 10 часам. Но общее решение системы дает больше информации. Прибавляя к 274 любое кратное 3705, мы получаем другое решение системы. Иначе говоря, спутники одновременно пересекают означенный меридиан каждые 3705 часов после первого такого момента, что соответствует 154 дням и 9 часам.

В следующем параграфе мы проведем детальный анализ примененного метода решения системы линейных сравнений. Заметим, что мы решали эту систему трех сравнений, рассматривая по два сравнения за раз. Действительно, сначала мы получили решение первых двух сравнений: $x = 79 + 195u$, что равносильно $x \equiv 79 \pmod{195}$. Для поиска решений третьего сравнения мы решаем другую систему двух уравнений, а именно

$$\begin{cases} x \equiv 79 \pmod{195}, \\ x \equiv 8 \pmod{19}. \end{cases}$$

В общей ситуации нам предстоит решать несколько систем двух сравнений. Поэтому в следующем параграфе мы детально проанализируем алгоритм решения систем только двух сравнений.

§ 8.3. Китайский алгоритм остатков: взаимно простые модули

Китайский алгоритм остатков так назван потому, что впервые был найден в «Учебнике математики мастера Сана», написанном между 287 и 473 годами нашей эры. В своей книге мастер Сан решал численные примеры, а потом выводил из них общие правила решения аналогичных задач. Более общий анализ той же проблемы с несколькими примерами можно найти в книге, написанной Цзинь Цзю-шао (Qin Jiushaq) в 1247 году. Аналогичные задачи рассматривались многими другими математиками, включая индийца Бхаскару (Bhaskara, VI век нашей эры) и Никомаха из Герасы. Историческую справку о теореме см. в [27].

Китайский алгоритм остатков — это простое обобщение метода, использованного при решении системы из § 8.2. Подробному изучению алгоритма и посвящен настоящий параграф.

Рассмотрим систему

$$\begin{cases} x \equiv a \pmod{m}, \\ x \equiv b \pmod{n}. \end{cases} \tag{3.1}$$

Как и в § 8.2, из первого сравнения следует, что $x = a + my$, где y — целое число. Подставляя вместо x во втором сравнении $a + my$, мы получаем: $a + my \equiv b \pmod{n}$. Другими словами,

$$my \equiv (b - a) \pmod{n}. \tag{3.2}$$

Но благодаря § 8.1 мы знаем, что это сравнение имеет решение тогда и только тогда, когда наибольший общий делитель

m и n делит $b-a$. Для уверенности в том, что это условие выполнено, достаточно предположить, что НОД$(n, m) = 1$. Предположение равносильно тому, что $\overline{m}$ имеет обратный элемент в $\mathbb{Z}_n$; скажем, $\overline{\alpha}$.

Теперь сравнение (3.2) легко решается. Умножая обе его части на α, получаем: $y \equiv \alpha(b-a) \pmod{n}$. Следовательно, $y = \alpha(b-a) + nz$, где z — целое. Так как $x = a + my$, то

$$x = a + m\alpha(b-a) + mnz.$$

Но $\overline{\alpha m} = \overline{1}$ в $\mathbb{Z}_n$. Значит, найдется такое целое β, при котором $1 - \alpha m = \beta n$. Итак,

$$x = a(1 - m\alpha) + m\alpha b + mnz = \alpha\beta n + m\alpha b + mnz.$$

Преимущество такой записи решения в том, что α и β легко вычисляются. Действительно, $1 = \alpha m + \beta n$, так что α и β находятся расширенным алгоритмом Эвклида, примененным к m и n. В итоге, если НОД$(m, n) = 1$, то при любом целом k число $a\beta n + b\alpha m + kmn$ является решением системы (3.1).

А сколько решений имеет такая система сравнений? Бесконечно много, если мы имеем в виду целочисленные решения. Ведь при каждом конкретном выборе z мы получаем новое решение, согласно формуле, выписанной выше. Рассмотрим этот момент более подробно. Предположим, что целые x и y удовлетворяют системе (3.1). Тогда $x \equiv a \pmod{m}$ и $y \equiv a \pmod{m}$. Разность этих сравнений приводит к соотношению: $x - y \equiv 0 \pmod{m}$, т.е. $x - y$ делится на m. Проделывая то же самое со вторым сравнением, получаем, что $x - y$ делится на n. А ввиду взаимной простоты m и n и леммы из § 3.4, это доказывает делимость $x - y$ на mn. Значит, если x и y — целые решения системы (3.1), то $x \equiv y \pmod{mn}$. Поэтому, хотя система и имеет бесконечно много решений, все они сравнимы друг с другом по модулю mn. Другими словами, система имеет только одно решение в $\mathbb{Z}_{mn}$. Но нельзя забывать, что все

наши выводы справедливы только благодаря предположению: НОД$(m, n) = 1$. Сведем все полученные факты в следующую теорему.

Китайская теорема об остатках. *Пусть m и n — взаимно простые натуральные числа. Система*

$$\begin{cases} x \equiv a \pmod{m}, \\ x \equiv b \pmod{n} \end{cases}$$

имеет одно и только одно решение в $\mathbb{Z}_{mn}$.

Хороший способ разобраться, действительно ли Вы поняли эту теорему, — рассмотреть ее геометрическую интерпретацию. Предположим, что у Вас есть таблица с mn клетками. Столбцы таблицы пронумерованы элементами $\mathbb{Z}_m$, а строки — элементами $\mathbb{Z}_n$. В клетку таблицы, стоящую на пересечении столбца, соответствующего $\overline{a} \in \mathbb{Z}_m$, и строки, соответствующей $\overline{b} \in \mathbb{Z}_n$, поместим целое число x, удовлетворяющее следующим условиям:

- $0 \leqslant x \leqslant mn - 1$,
- $x \equiv a \pmod{m}$,
- $x \equiv b \pmod{n}$.

Скажем, что соответствующая клетка таблицы имеет координаты $(\overline{a}, \overline{b})$. Поскольку $0 \leqslant x \leqslant mn - 1$, мы можем считать число x представителем класса сопряженности по модулю mn, т.е. реально x представляет некоторый класс $\overline{x} \in \mathbb{Z}_{mn}$.

Что говорит об этой таблице китайская теорема об остатках? Предположим, что НОД$(m, n) = 1$. Из теоремы следует, что каждая клетка таблицы соответствует в точности одному целому, заключенному между 0 и $mn - 1$, т.е. одному классу из $\mathbb{Z}_{mn}$. Таким образом, в разных клетках стоят разные числа и наоборот. Еще раз напомним, что модули у нас взаимно

просты. Приведем таблицу для $m = 4$ и $n = 5$.

	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{5}$	$\overline{10}$	$\overline{15}$
$\overline{1}$	$\overline{16}$	$\overline{1}$	$\overline{6}$	$\overline{11}$
$\overline{2}$	$\overline{12}$	$\overline{17}$	$\overline{2}$	$\overline{7}$
$\overline{3}$	$\overline{8}$	$\overline{13}$	$\overline{18}$	$\overline{3}$
$\overline{4}$	$\overline{4}$	$\overline{9}$	$\overline{14}$	$\overline{19}$

Заметим, что таблица соответствует прямому произведению $\mathbb{Z}_4 \times \mathbb{Z}_5$. С первого взгляда может показаться, что для заполнения таблицы необходимо решить 20 систем линейных сравнений. Но китайская теорема об остатках подсказывает заполнять таблицу «в обратном порядке». Для целого числа x между 0 и $mn - 1$ найдем место в таблице, вычисляя его вычеты по модулю m и n. Например, в нашем случае вычет 14 по модулю 4 равен 2, а по модулю 5 — 4. Поэтому ему соответствует клетка с координатами $(\overline{2}, \overline{4})$.

Но это не последнее слово в науке; можно сделать еще проще. Реально, существует возможность заполнить всю таблицу целиком, не производя вычислений для отдельных чисел! Чтобы понять как, вспомним, что у нас есть геометрическая интерпретация множества $\mathbb{Z}_4$: четыре точки, расположенные на равных расстояниях друг от друга вдоль окружности, каждая из которых представляет один класс в $\mathbb{Z}_4$. Похожая картинка есть и для $\mathbb{Z}_5$.

В действительности нашу таблицу можно интерпретировать как плоскую карту, или план, представляющий некую расположенную в пространстве двумерную поверхность. Чтобы найти эту поверхность, мы поступаем следующим образом. Поскольку классы $\mathbb{Z}_4$ (горизонтальные координаты) можно считать расположенными вдоль окружности, мы склеим правую сторону таблицы с левой. Получится цилиндр. Но классы $\mathbb{Z}_5$ (вертикальные координаты) также удобно представлять

точками на окружности. Поэтому верх таблицы нужно склеить с низом. В результате получится поверхность, которая называется *тором*, напоминающая по форме бублик или баранку.

Вернемся к задаче о заполнении таблицы. Поскольку числа 0, 1, 2 и 3 меньше и четырех, и пяти, они совпадают со своими вычетами по обоим этим модулям. Поэтому нам не нужно делать каких-либо вычислений для определения их координат, и мы сразу можем поместить их в таблицу:

	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$			
$\overline{1}$		$\overline{1}$		
$\overline{2}$			$\overline{2}$	
$\overline{3}$				$\overline{3}$
$\overline{4}$				

Заметим, что расставляя по очереди эти классы по клеткам таблицы, мы, начав с левого верхнего угла таблицы, переходили каждый раз на одну клетку вправо и одну вниз. А проставив четыре первых класса, уперлись в правую границу таблицы. Если бы в таблице был еще один столбец, то, придерживаясь сформулированного правила, мы поместили бы в нем 4, но на одну строчку ниже 3, т.е. в последней строке. Однако у нас нет этого лишнего столбца, не так ли? На помощь приходит геометрическая интерпретация таблицы. Склеив ее вертикальные границы, мы видим, что первый левый столбец таблицы можно считать идущим сразу за ее последним правым столбцом. В терминах несклеенной таблицы это означает, что мы должны «перепрыгнуть» с последнего столбца на первый, одновременно спустившись на одну строчку вниз. Следовательно, 4 нужно поставить на пересечение первого столбца

и последней строки таблицы:

	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$			
$\overline{1}$		$\overline{1}$		
$\overline{2}$			$\overline{2}$	
$\overline{3}$				$\overline{3}$
$\overline{4}$	$\overline{4}$			

Кажется, у нас появилась новая проблема: мы дошли до нижней границы и опять не можем двигаться дальше. Но в силу геометрической картинки, нижняя и верхняя границы таблицы склеены, и мы можем перейти к первой строке таблицы. Только теперь нужно сместиться на один столбец вправо относительно предыдущего положения. Сделав это в нашем случае, мы получим:

	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$	$\overline{5}$		
$\overline{1}$		$\overline{1}$		
$\overline{2}$			$\overline{2}$	
$\overline{3}$				$\overline{3}$
$\overline{4}$	$\overline{4}$			

Мы можем повторять этот процесс, пока не заполним все клетки таблицы.

§ 8.4. Китайский алгоритм остатков: общий случай

Мы в деталях проанализировали решение системы линейных сравнений для взаимно простых модулей потому, что именно этот случай встретится нам в следующих главах. Однако

китайский алгоритм остатков можно так же использовать и при решении систем, у которых модули не взаимно простые. Но в этом случае при решении линейных сравнений требуется предельная внимательность, возрастающая с каждым шагом алгоритма. Достаточно продемонстрировать один пример. Рассмотрим систему

$$\begin{cases} x \equiv 3 \pmod{12}, \\ x \equiv 19 \pmod{8}. \end{cases}$$

Из первого сравнения мы получаем: $x = 3 + 12y$ для некоторого целого y. Подставляя найденное выражение для x во второе сравнение системы, имеем: $12y \equiv 16 \pmod 8$. Так как НОД$(12, 8) = 4$ делит 16, последнее сравнение должно иметь решение. Действительно, его целочисленный эквивалент имеет вид: $12y - 8z = 16$. Разделим это равенство на 4: $3y - 2z = 4$, т.е. $3y \equiv 4 \pmod 2$. Но $3 \equiv 1 \pmod 2$, а $4 \equiv 0 \pmod 2$, так что $y \equiv 0 \pmod 2$. Следовательно, $y = 2k$ для некоторого целого k. Наконец, подставляя $2k$ вместо y в равенство $x = 3 + 12y$, находим $x = 3 + 24k$. Итак, данная система имеет единственное решение по модулю 24. Однако $8 \cdot 12 = 96$. Так какое же отношение число 24 имеет к модулям 8 и 12? Ответ на этот вопрос найдете в упражнении 5.

Для любой пары не взаимно простых модулей всегда можно написать систему сравнений, не имеющую ни одного решения. С точки зрения геометрической интерпретации § 8.3, это означает, что если модули обладают общим нетривиальным делителем, то в соответствующей таблице всегда останутся незаполненные клетки.

Еще раз повторим, нет никакой необходимости делать какие-либо вычисления для заполнения таблицы. Просто мы должны проставлять числа $0, 1, \ldots$, начиная с левой верхней клетки, сдвигаясь каждый раз на один столбец вправо и одну строку вниз, не забывая «перепрыгивать» справа налево и снизу вверх при приближении к соответствующей границе таблицы.

Заполняя таким образом таблицу для не взаимно простых модулей, мы вернемся в клетку с координатами $(\overline{0}, \overline{0})$, не перебрав всех $mn - 1$ чисел. Это объясняет, почему некоторые клетки таблицы останутся пустыми. Для $m = 4$ и $n = 6$ таблица выглядит следующим образом:

	$\overline{0}$	$\overline{1}$	$\overline{2}$	$\overline{3}$
$\overline{0}$	$\overline{0}$		$\overline{6}$	
$\overline{1}$		$\overline{1}$		$\overline{7}$
$\overline{2}$	$\overline{8}$		$\overline{2}$	
$\overline{3}$		$\overline{9}$		$\overline{3}$
$\overline{4}$	$\overline{4}$		$\overline{10}$	
$\overline{5}$		$\overline{5}$		$\overline{11}$

§ 8.5. Снова степени

Для систем сравнений, число уравнений в которых больше двух, есть свой вариант китайской теоремы об остатках. Мы ее сформулируем без доказательства, поскольку это просто еще одно приложение китайского алгоритма остатков. Сначала определение: натуральные числа $n_1, \ldots, n_k$ называются *попарно взаимно простыми*, если НОД$(n_i, n_j) = 1$ для любой пары $i \neq j$. Например, три числа n_1, n_2, n_3 — попарно взаимно простые, если НОД$(n_1, n_2) = 1$, НОД$(n_1, n_3) = 1$, НОД$(n_2, n_3) = 1$.

Китайская теорема об остатках. *Пусть $n_1, \ldots, n_k$ — попарно взаимно простые натуральные числа. Тогда всякая система*

$$\begin{cases} x \equiv a_1 \pmod{n_1}, \\ x \equiv a_2 \pmod{n_2}, \\ \ldots\ldots\ldots\ldots\ldots \\ x \equiv a_k \pmod{n_k} \end{cases}$$

имеет одно и только одно решение в $\mathbb{Z}_{n_1 \cdots n_k}$.

Применив эту версию теоремы, мы можем упростить вычисление вычетов степеней по модулю n, если известно его разложение на простые множители. Мы будем также предполагать, что каждый простой множитель входит в это разложение с кратностью 1, потому что именно в этом случае метод наиболее эффективен.

Допустим, что разложение имеет вид: $n = p_1 \cdots p_k$, где $0 < p_1 < \cdots < p_k$ — простые числа. Для целых a и m мы сначала находим вычет a^m по каждому модулю p_i. Если простые множители не слишком велики, то вычисления будут очень быстрыми даже для больших m и a, поскольку нам помогает это делать теорема Ферма. Предположим, что мы уже сделали эти вычисления, причем

$$\begin{array}{llll} a^m \equiv r_1 \pmod{p_1} & \text{и} & 0 \leqslant r_1 < p_1, \\ a^m \equiv r_2 \pmod{p_2} & \text{и} & 0 \leqslant r_2 < p_2, \\ \ldots\ldots\ldots\ldots\ldots & \ldots & \ldots\ldots\ldots\ldots \\ a^m \equiv r_k \pmod{p_k} & \text{и} & 0 \leqslant r_k < p_k. \end{array}$$

Поэтому для определения вычета a^m по модулю n нам нужно только решить систему сравнений:

$$\begin{cases} x \equiv r_1 \pmod{p_1}, \\ x \equiv r_2 \pmod{p_2}, \\ \ldots\ldots\ldots\ldots\ldots \\ x \equiv r_k \pmod{p_k}. \end{cases}$$

Заметим, что модули системы — различные простые числа, поэтому они попарно взаимно просты. Значит, по китайской теореме об остатках, система всегда имеет решение, скажем, r, $0 \leqslant r \leqslant n-1$. Более того, любые два таких решения сравнимы по модулю $p_1 \cdots p_k = n$. Так как a^m — тоже решение системы, имеем $a^m \equiv r \pmod{n}$. Следовательно, r — вычет a^m по модулю n.

Приведем пример. Допустим, нам нужно найти вычет числа 2^{6754} по модулю 1155. Раскладывая 1155 на множители,

найдем: $1155 = 3 \cdot 5 \cdot 7 \cdot 11$. Применив теорему Ферма к каждому из этих простых чисел, получим:

$$\begin{aligned} 2^{6754} &\equiv 1 \pmod 3, \\ 2^{6754} &\equiv 4 \pmod 5, \\ 2^{6754} &\equiv 2 \pmod 7, \\ 2^{6754} &\equiv 5 \pmod{11}. \end{aligned}$$

Таким образом, нам осталось решить систему:

$$\begin{cases} x \equiv 1 \pmod 3, \\ x \equiv 4 \pmod 5, \\ x \equiv 2 \pmod 7, \\ x \equiv 5 \pmod{11} \end{cases}$$

с помощью китайского алгоритма остатков. По первому сравнению системы $x = 1 + 3y$, поэтому второе сравнение дает:

$$1 + 3y \equiv 4 \pmod 5, \qquad \text{т.е.} \qquad y \equiv 1 \pmod 5,$$

потому что тройка обратима по модулю 5 и на нее можно сократить обе части сравнения. Итак, $x = 4 + 15z$. Подставляя вместо x его выражение через z в третье сравнение и решая его, получаем: $x = 79 + 105t$. Наконец, решая четвертое сравнение относительно t, мы имеем $t \equiv 6 \pmod{11}$. Значит $x = 709 + 1155u$ и 709 — искомый вычет числа 2^{6754} по модулю 1155.

§ 8.6. Посвящение в тайну

Бенджамен Франклин (Franklin) однажды сказал: «Трое могут хранить тайну, если двое из них мертвы.» В этом параграфе мы изучаем безопасную систему допуска живых к секретным сведениям, основанную на китайской теореме об остатках. Представьте себе следующую ситуацию. Подвал банка

должен открываться каждый день. В банке служат пять старших кассиров, имеющих доступ к подвалу. По причинам безопасности руководство банка предпочитает систему, требующую присутствия хотя бы двух из этой пятерки для возможности открыть подвал. Проблема в том, чтобы подвал могли открыть *любые* два старших кассира.

Рассмотрим эту проблему в более общем виде. Для того, чтобы открыть подвал банка, необходимо знать код, который можно считать натуральным числом s. Мы хотим распределить этот код между n старшими кассирами так, чтобы каждый из них знал что-то об s. Назовем такую частичную информацию *фрагментом* кода. Более того, открыть подвал должно быть невозможно, если в банке присутствуют менее k старших кассиров, где $k \geqslant 2$ — натуральное число, меньшее n. Мы добьемся этого условия, распределив информацию о коде таким образом, что

- число s *легко* определяется, если известно k или более фрагментов;
- число s *трудно* определимо, если известно менее k фрагментов.

Фрагменты кода, сообщаемые каждому из старших кассиров, — это, в действительности, элементы множества $\mathbb{S}$, состоящего из n упорядоченных пар натуральных чисел. Чтобы построить $\mathbb{S}$, выберем сначала множество $\mathfrak{L}$ из n попарно взаимно простых чисел. Пусть N — произведение наименьших k из них, а M — произведение $k-1$ наибольших. Будем говорить, что k является *порогом для* $\mathfrak{L}$, если $M < N$. Из этого условия следует, что произведение любых k (или более) элементов из $\mathfrak{L}$ всегда больше, чем N, а произведение $k-1$ (или менее) его элементов — всегда меньше M.

Предположим, код s выбран так, что $M < s < N$, а множество $\mathbb{S}$ состоит из пар (m, s_m), где $m \in \mathfrak{L}$, а s_m — вычет числа s по модулю m. Эти пары и являются теми *фрагментами кода*,

которые сообщаются старшим кассирам. Тот факт, что множество $\mathfrak{L}$ имеет порог $k \geqslant 2$, обеспечивает неравенство $s > m$ для каждого $m \in \mathfrak{L}$. В частности, $s_m < s$ для любого $m \in \mathfrak{L}$.

Что произойдет, если k или более старших кассиров находятся в банке? В этом случае известны t ($\geqslant k$) пар из множества $\mathbb{S}$. Обозначив эти пары через $(m_1, s_1), \ldots, (m_t, s_t)$, рассмотрим систему сравнений:

$$\begin{cases} x \equiv s_1 \pmod{m_1}, \\ x \equiv s_2 \pmod{m_2}, \\ \ldots\ldots\ldots\ldots\ldots \\ x \equiv s_t \pmod{m_t}. \end{cases} \tag{6.1}$$

Элементы множества $\mathfrak{L}$ попарно взаимно просты. Значит, по китайской теореме об остатках, эта система имеет решение $0 \leqslant x_0 < m_1 \cdots m_t$. Но совпадает ли x_0 с s? Это как раз та причина, по которой мы накладывали требование: $\mathfrak{L}$ имеет порог k. Поскольку $t \geqslant k$, то наше требование влечет:

$$m_1 \cdots m_t \geqslant N > s.$$

Но s тоже удовлетворяет системе (6.1), и по китайской теореме об остатках

$$x_0 \equiv s \pmod{m_1 \cdots m_t}.$$

А так как s и x_0 — натуральные числа, меньшие $m_1 \cdots m_t$, то $s = x_0$.

Предположим теперь, что в банке находится менее k старших кассиров. Несмотря на то, что t теперь меньше k, мы все равно сможем решить систему (6.1). Пусть x_0 — наименьшее неотрицательное решение, тогда $0 \leqslant x_0 < m_1 \cdots m_t$. Но произведение меньшего, чем k количества элементов из $\mathfrak{L}$ всегда меньше M; так что $x_0 < M < s$. Следовательно, решения системы не достаточно для восстановления кода s. Однако как x_0, так и s — решения системы (6.1), поэтому

$$s = x_0 + y \cdot (m_1 \cdots m_t),$$

где y — некоторое натуральное число. Неравенство

$$N > s > M > x_0$$

влечет

$$\frac{M - x_0}{m_1 \cdots m_t} \leqslant y \leqslant \frac{s - x_0}{m_1 \cdots m_t} \leqslant \frac{N - x_0}{m_1 \cdots m_t}.$$

Приходим к выводу: если $t < k$, то для восстановления кода s нам предстоит отыскивать недостающий множитель y среди более чем

$$d = \left[\frac{N - M}{M}\right]$$

целых чисел. Выбрав модули так, чтобы d оказалось очень большим, мы сделаем задачу поиска y практически нерешаемой.

Для завершения разбора задачи осталось осветить один вопрос: можно ли найти множество $\mathfrak{L}$, удовлетворяющее всем необходимым требованиям? Ответ на него положителен, но нуждается в результатах о распределении простых чисел, которые выходят за рамки данной книги. Этот вопрос детально обсуждается в [32].

Сделаем обзор рассмотренной конструкции. Для нее требуются начальные данные: число n старших кассиров, имеющих доступ в подвал банка, и наименьшее число k из них, присутствие которых в банке достаточно для открытия подвала. Первое число определяет размер множества $\mathfrak{L}$, а второе — его порог k. Далее нам нужно подобрать множество $\mathfrak{L}$ из n элементов с порогом k (эту часть конструкции мы подробно не обсуждали), и вычислить M и N, определенные выше. Напомним, что $\mathfrak{L}$ нужно выбирать с таким расчетом, чтобы число d, о котором мы говорили, было как можно больше; в противном случае код может быть разгадан простым перебором. Код s — натуральное число, которое выбирается лежащим между M и N. Теперь можно вычислить элементы множества $\mathbb{S}$ и сообщить их сотрудникам. Конечно, безопасность этой схемы зависит от того, насколько велико k, уменьшающее вероятность, что одновременно k кассиров из одного банка окажутся

нечестными. Если это все-таки произойдет, то нам придется утешать себя мыслью, что не существует систем безопасности 100-процентной надежности.

Рассмотрим пример. Допустим, что в банке работают 5 старших кассиров и из соображений безопасности по крайней мере двое из них должны присутствовать при открытии подвала. Значит, $\mathfrak{L}$ должно состоять из пяти элементов, а его порог равен 2. Выбрав элементы $\mathfrak{L}$ среди малых простых чисел, получим:

$$\mathfrak{L} = \{11,\ 13,\ 17,\ 19,\ 23\}.$$

Произведение двух наименьших чисел этого множества равно $N = 11 \cdot 13 = 143$. С другой стороны, поскольку $k = 2$, произведение $k - 1$ наибольших простых из $\mathfrak{L}$ в действительности равно его максимальному элементу. Таким образом, $M = 23$ и $\mathfrak{L}$ имеет порог 2. Код s может быть любым целым числом, лежащим между 23 и 143. Пусть $s = 30$. Тогда

$$\mathbb{S} = \{(11, 19), (13, 17), (17, 13), (19, 11), (23, 7)\}.$$

Наконец, что будет, если в банке присутствуют старшие кассиры с фрагментами (17,13) и (23,7)? Код из их фрагментов получается как наименьшее число, удовлетворяющее системе:

$$\begin{cases} x \equiv 13 \pmod{17}, \\ x \equiv 7 \pmod{23}. \end{cases}$$

Легко увидеть, что таким числом будет 30. Этот код корректен, он позволяет открыть подвал.

Упражнения

1. Решите систему сравнений, упоминавшуюся еще в китайской книге 717 года н. э.:

$$\begin{cases} x \equiv 1 \pmod{2}, \\ x \equiv 2 \pmod{5}, \\ x \equiv 5 \pmod{12}. \end{cases}$$

2. Задача из «Учебника математики мастера Сана»: «Есть несколько предметов. Если их число разделить на 3, то в остатке получится 2; при делении числа предметов на 5, в остатке остается 3, а на 7 — 2. Каково число предметов?»

3. Задача из «Ариабхатиамы», индийского трактата по арифметике VI века: «Найдите наименьшее натуральное число, дающее остаток 5 при делении на 8; остаток 4 при делении на 9, а при делении на 7 — остаток 1.»

4. В древней индийской астрономии использовался период Кальпа длительностью в 4320 миллионов лет; предполагалось, что в его начале и конце все фундаментальные астрономические константы планет были нулевыми. Допустим, что в некоторый момент времени T от начала Кальпы солнце, луна и т.д. пропутешествовали следующие количества дней после завершения их полных оборотов:

Солнце	Луна	Марс	Меркурий	Юпитер	Сатурн
1000	41	315	1000	1000	1000

Зная, что солнце совершает три оборота за 1096 дней, луна — один оборот за 185 дней, Меркурий — тринадцать оборотов за 1096 дней, Юпитер — три оборота за 10 960 дней и Сатурн — один оборот за 10 960, найдите число дней от начала Кальпы до момента T.

5. Покажите, что система

$$\begin{cases} x \equiv a \pmod{m}, \\ x \equiv b \pmod{n} \end{cases}$$

не может иметь более одного решения по модулю наименьшего общего кратного m и n. Заметим, что мы не предполагаем взаимной простоты чисел m и n.

6. Используя китайскую теорему об остатках, вычислите остатки от деления $2^{45\,632}$ и $3^{54\,632}$ на 12 155.

7. Решите сравнение: $x^2 + 42x + 21 \equiv 0 \pmod{105}$.

Указание: Разложите 105 на простые множители и решите сравнение по модулю каждого простого делителя числа 105. Затем примените китайский алгоритм остатков.

8. Найдите последовательность наибольшей длины, состоящую из подряд идущих простых чисел, начиная с 11, с порогом 3. Сделайте то же самое для порога 4.

9. Пусть p и q — различные простые числа и $n = pq$. Предположим, что мы знаем решения уравнений: $x^2 \equiv a \pmod{p}$ и $x^2 \equiv a \pmod{q}$. Покажите, как китайский алгоритм остатков можно использовать для решения уравнения $x^2 \equiv a \pmod{n}$. Сравните Ваше решение с методом из § 8.5 и упражнением 7.

10. Пусть p и q — различные простые числа и $n = pq$. Предположим, что оба простых числа имеют остаток 3 при делении на 4. Напишите программу, которая по данным p, q и a находит решения уравнения $x^2 \equiv a \pmod{n}$. Наше предположение о простоте чисел облегчает решение уравнений $x^2 \equiv a \pmod{p}$ и $x^2 \equiv a \pmod{q}$ (см. главу 6, упражнение 17). Это третья задача из серии, приведенной в конце упражнения 8 главы 12.