Арифметика остатков

1. Чётные и нечётные числа

Все знают, что числа бывают чётные и нечётные. Чётные делятся на 2 без остатка, а нечётные дают остаток 1. Другими словами, чётные числа имеют вид $2k$ для целых k, а нечётные $2k+1$, тоже при целых k. Скажем, число 0 чётное ($0 = 2 \cdot 0$), а число -3 нечётное ($-3 = 2 \cdot (-2) + 1$).

Несложно проверить, что сумма двух чётных или двух нечётных чисел чётна, а сумма чётного и нечётного числа нечётна:

+	Ч	Н
Ч	Ч	Н
Н	Н	Ч

×	Ч	Н
Ч	Ч	Ч
Н	Ч	Н

Например, если мы складываем чётное и нечётное число, то получаем $2k + (2l+1) = 2(k+l)+1$, то есть нечётное число. А если мы умножаем два нечётных числа, то получаем

$$(2k+1)(2l+1) = 4kl + 2k + 2l + 1 = 2(2kl + k + l) + 1,$$

то есть нечётное число.

► Папа с сыном играют в такую игру: каждый пишет на бумажке число, не говоря его другому, потом они открывают эти числа, и если сумма чётная, выигрывает папа, а если нечётная, то сын. Честная ли это игра? Тот же вопрос, если вместо суммы чисел берётся их произведение. ◄

Среди целых чисел чётные и нечётные встречаются одинаково часто: если мы возьмём большой интервал подряд идущих чисел, то количество чётных и нечётных в этом интервале будет почти одинаково (отличаться максимум на 1)

► Почему? ◄

Если заменить в таблицах буквы Ч и Н на 0 и 1, взяв нуль и единицу в качестве представителей классов чётных и нечётных чисел, то получатся почти что обычные таблицы сложения и умножения:

+	0	1
0	0	1
1	1	0

×	0	1
0	0	0
1	0	1

Разница с обычными только в одном месте: $1 + 1 = 0$.

2. Деление на 3 и остатки

Попробуем составить аналогичные таблицы сложения и умножения для чисел, делящихся и не делящихся на 3. Тут сразу же возникает проблема: мы не знаем, что сказать про сумму двух чисел, не делящихся на 3. Она может делиться на 3 (например, $1+5=6$), а может и не делиться (например, $2+5=7$). Дело в том, что не делящиеся на 3 числа могут быть двух видов: одни дают остаток 1 (имеют вид $3k+1$ при целом k), а другие дают остаток 2 (имеют вид $3k+2$).

► К какому типу относится число 1000? А число -1? Найдите соответствующие значения k. ◄

► Покажите, что все три типа чисел (делящиеся на 3, дающие остаток 1 и дающие остаток 2) встречаются одинаково часто: в большом отрезке подряд идущих чисел их количества отличаются максимум на 1. ◄

Теперь можно составить аналогичную таблицу сложения и умножения остатков по модулю 3:

$+$	0	1	2
0	0	1	2
1	1	2	0
2	2	0	1

$\times$	0	1	2
0	0	0	0
1	0	1	2
2	0	2	1

Каждую клеточку в этой таблице несложно проверить. Например,

$$(3k+1)+(3l+2)=3(k+l)+3=3(k+l+1)+0$$

и

$$(3k+2)(3l+2)=9kl+6k+6l+4=3(3kl+2k+2l+1)+1.$$

► Папа с сыном играют в такую игру: каждый пишет на бумажке число, не говоря его другому, потом они открывают эти числа, и если сумма не делится на 3, выигрывает папа, а если делится, то сын. Честная ли это игра? ◄

Аналогичные таблицы можно составить не только для деления на 2 и 3, но и для любого числа. Но сначала мы дадим более формальные определения.

3. Деление с остатком

Говорят, что целое число a делится на целое число b, если $a=bk$ для некоторого целого числа k. В этом случае говорят также «a кратно b», и «b является делителем числа a».

В этом определении можно было бы сказать: «если частное a/b целое», но этим бы исключался случай $b=0$, который формально допустим по нашему определению. Правда, особого смысла в нём всё равно нет: единственное число, которое делится на 0, это число 0. Определение допускает также отрицательные a и b: скажем, число -6 делится на -2 (а также и на 2), всего у него 8 делителей, если считать и положительные, и отрицательные.

► Что это за делители? ◄

Впрочем, обычно, говоря о количестве делителей у положительного целого числа, имеют в виду только положительные делители (считая единицу и само число).

► Сколько (положительных) делителей у числа $30=2\cdot3\cdot5$? у числа $210=2\cdot3\cdot5\cdot7$? ◄

► Придумайте (положительное целое) число, у которого было бы ровно 6 (положительных) делителей. Тот же вопрос для 7 делителей. ◄

► Говорить о кратных можно не только для целых чисел, но и для отрезков: один отрезок кратен другому, если второй укладывается в первом целое число раз, то есть если отношение (длина первого)/(длина второго) целое. Докажите, что два отрезка имеют *общую меру* (отрезок, которому они оба кратны) тогда и только тогда, когда они имеют общее кратное. ◄

► Найдите наименьшее общее кратное отрезков длиной $15/6$ и $21/10$. ◄

► Могут ли два отрезка не иметь общего кратного? ◄

Если два числа a и b делятся на третье число c, то и их сумма $a+b$ и разность $a-b$ делятся на это c. В самом деле, если $a=kc$ и $b=lc$, то $a+b=(k+l)c$ и $a-b=(k-l)c$.

Для делимости произведения достаточно делимости одного из сомножителей: если a делится на c, то и ab делится на c, каково бы ни было (целое) b. В самом деле, если $a=kc$, то $ab=k(bc)=(kb)c$.

Множество целых чисел называют *идеалом*, если вместе с любыми двумя числами оно содержит их сумму и разность, и вместе с любым числом оно содержит все его кратные. Используя эту терминологию, можно сказать, что для любого c множество всех кратных числа c является идеалом.

► Докажите, что произведение любых трёх последовательных целых чисел делится на 3. ◄

► Докажите, что число $a^3 - a$ делится на 3 при любом целом a. ◄

► Докажите, что сумма $84 + 85 + 86 + 87 + 88 + 89 + 90$ делится на 7 и на 87. ◄

► Найдите трёхзначный и семизначный делители числа 103103103. ◄

► Какие из следующих утверждений верны: (1) если a делится на c, а b не делится на c, то $a + b$ не делится на c; (2) если a не делится на c и b не делится на c, то $a + b$ не делится на c; (3) если a не делится на c и b не делится на c, то ab не делится на c? Докажите верные и приведите контрпримеры к неверным. ◄

► Известно, что $a, b, , c, d$ — положительные целые числа, и $ab = cd$. Докажите, что если a делится на c, то d делится на b. ◄

► Числа a и b целые, причём $2a + 3b$ делится на 7. Докажите, что $a + 5b$ также делится на 7. ◄

► Положительное целое число a чётно, но не делится на 4. Покажите, что количество (положительных) чётных делителей a равно количеству (положительных) нечётных делителей a. ◄

► Докажите, что произведение любых k подряд идущих целых чисел делится на $k! = 1{\cdot}2{\cdot}3{\cdot}\ldots{\cdot}k$.◄

Теперь определим деление с остатком. Пусть b — целое положительное число. Деля на b с остатком, мы связываем предметы в пачки по b в каждой, пока это возможно: количество полных пачек называется *частным* (говорят ещё «неполное частное», чтобы отличать от частного как дроби), и сколько-то предметов останется, их количество и называют *остатком*

Формально: разделить целое a на целое положительное b означает найти такое целое q (*частное*) и такое целое r (остаток), что

$$a = b \cdot q + r; \quad 0 \leqslant r < b.$$

Ограничения на r понятны: остаётся сколько-то предметов (возможно, ни одного), но меньше b, иначе возникла бы ещё одна целая пачка.

Теперь можно сформулировать теорему: *деление с остатком всегда возможно, притом единственным образом.*

Не очень понятно, что тут доказывать (неужели это не очевидно, если подумать о связывании предметов в пачки?). Тем не менее для педантов можно привести такое доказательство.

Единственность. Если $a = bq + r = bq' + r'$, то $r - r' = b(q' - q)$ и потому $r - r'$ делится на b. Но оба числа r, r' находятся в интервале $0, 1, \ldots, b - 1$, так что их разность (если из большего вычесть меньше) не больше $b - 1$, и может делиться на b, лишь если равна нулю. Поэтому $r = r'$, откуда и $b = b'$.

Существование можно доказать индукцией по a. Для $a = 0$ частное и остаток равны нулю: $0 = 0 \cdot b + 0$. Если $a = bq + r$, то $a + 1 = bq + (r + 1)$. При этом $r + 1 \leqslant b$, так как $r < b$. Если $r + 1 < b$, то для $a + 1$ получаем частное q и остаток $r + 1$. Если же $r + 1 = b$, то $a + 1 = bq + b = b(q + 1) + 0$, получаем частное $q + 1$ и остаток 0.

Программистам будет ближе другое доказательство существования:

```
q:=0; r:=a;
{a=bq+r; r>=0}
пока (r>=b):
   q:=q+1; r:=r-b;
{a=bq+r; r>=0; r<b}
```

Здесь равенство $a = bq + r$ и неравенство $r \geqslant 0$ являются, как говорят, *инвариантом цикла*, они выполняются после любого числа итераций. В самом деле, начальные значения $q = 0$ и $r = a$ удовлетворяют условию $a = qa + r$; итерация цикла происходит при $r \geqslant b$ и потому r

после уменьшения на b остаётся неотрицательным, а условие $a = bq + r$ не нарушается, если увеличить q на единицу и одновременно уменьшить r на b. Выход из цикла происходит, когда условие нарушается, то есть $r < b$ (в дополнение к инварианту) — что и требуется определением остатка.

Внимательные читатели, наверно, уже заметили ошибку в рассуждении: оно годится лишь при $a \geqslant 0$, поскольку иначе инвариант $r \geqslant 0$ будет нарушен. Собственно, и индуктивное рассуждение годилось тоже только при $a \geqslant 0$. Так что для отрицательных a возможность деления с остатком надо доказывать отдельно. Пусть $a = -a'$, разделим a' на b с остатком: $a' = bq' + r'$. Если $r' = 0$, то $a = -a' = b(-q') + 0$, так что можно взять $q = -q'$ и $r = 0$. Если $r' > 0$, то можно записать

$$a = -a' = -bq' - r' = b(-q' - 1) + (b - r'),$$

так что можно взять $q = -q' - q$ и $r = b - r'$: поскольку мы предположили, что $0 < r' < b$, то $0 < b - r' < b$, и r попадает в нужный диапазон.

Аналогичная проблема возникает во многих языках программирования: целочисленное деление a на b даёт неполное частное в нашем смысле лишь при $a \geqslant 0$, а, скажем, -7 `div` 3 оказывается равным -2, а не -3, как требует наше определение. Тем не менее математически наше определение удобнее, так как при этом сохраняется общая формула для чисел с данным остатком: скажем, числа $3k + 1$ при всех целых k при делении на 3 дают остаток 1 (и частное k).

► Можно ли разрезать шахматную доску 8×8 на прямоугольники 3×1? ◄

► Найти число вида $100 * **$ (звёздочки обозначают некоторые цифры), которое делится на 547. ◄

► Книги на столе пытались связывать в пачки по 2, по 3, по 4 и по 5 книг, и каждый раз оставалась одна лишняя. Сколько книг было на столе? (Известно, что их было больше одной и не больше 100.) ◄

► Число a даёт остаток 6 при делении на 12. Может ли оно давать остаток 12 при делении на 20? ◄|

4. Сравнения по модулю

Если два числа a и b дают одинаковые остатки при делении на положительное число N, то говорят, что они *сравнимы* по модулю N, и пишут $a \equiv b \pmod N$.

Эквивалентное определение: a и b сравнимы по модулю N, если разность $a - b$ делится на N. (В самом деле, если они дают одинаковый остаток r, то $a = kN + r$, $b = lN + r$, и $a - b = kN - lN = (k - l)N$. Наоборот, если $a - b = mN$, и b даёт остаток r, то $b = lN + r$ и $a = (a - b) + b = mN + lN + r = (m + l)N + r$, то есть a даёт тот же остаток r.)

Можно сказать, что при данном N все целые числа разбиваются на N классов в зависимости от остатков по модулю N: два числа в одном классе сравнимы, а числа в разных классах — нет.

► Докажите, что числа a^2 и b^2 дают одинаковые остатки при делении на $a - b$, если a и b — положительные целые числа, и $a > b$. ◄

Важное свойство сравнений: чтобы узнать, в какой класс попадет сумма или произведение двух чисел, достаточно знать, в каком классе лежат слагаемые или сомножители: если одно из слагаемых (один из сомножителей) изменить на кратное N, то сумма (произведение) тоже изменится на кратное N.

В самом деле, если к одному из слагаемых прибавить kN, то к сумме тоже прибавится kN, аналогично для разности. С произведением: $(a + kN)b = ab + kbN \equiv ab \pmod N$.

Благодаря этому свойству в выражении, содержащем операции сложения и умножения (или возведение в целую степень, которое сводится к многократному умножению), можно заменять слагаемые или сомножители на сравнимые по модулю N — если результат нам важен лишь по модулю

N. Например, можно найти $2^{100} \bmod 7$ (остаток от деления 2^{100} на 7): поскольку $2^3 = 8 \equiv 1 \pmod 7$, то $2^{99} \equiv (2^3)^{33} \equiv 1^{33} = 1 \pmod 7$, так что $2^{100} \equiv 2^{99} \cdot 2 \equiv 1 \cdot 2 = 2 \pmod 7$.

► Какой остаток даёт число 100^{100} при делении на 99? ◄

► Найдите две последние цифры числа 99^{1000}. ◄

Сравнения по модулю (пусть не под таким названием) часто встречаются в быту. Скажем, циферблат показывает количество прошедших часов по модулю 12, а также количество минут по модулю 60. Измеряя углы в градусах, мы фактически измеряем число градусов по модулю 360.

► Как построить угол в 1°, если задан угол в 19°? ◄

Последняя цифра (для положительного целого числа) сравнима в этим числом по модулю 10, а две последние цифры — по модулю 100. В музыке двенадцать полутонов составляют целую октаву, и потом названия нот (до, до диез и так далее) повторяются.

Известный признак делимости на 9 (число делится на 9 тогда и только тогда, когда его сумма цифр делится на 9) можно обобщить и сказать, что число и его сумма цифр сравнимы по модулю 9. Это легко следует из наших рассуждений. Скажем, для четырёхзначного числа:

$$\overline{abcd} = 1000a + 100b + 10c + d \equiv 1a + 1b + 1c + 1d = a + b + c + d \pmod 9,$$

поскольку $10 \equiv 1 \pmod 9$ и, следовательно, 10^2, 10^3,... все сравнимы с 1 по модулю 9.

► Придумайте признак делимости на 11, использующий знакопеременную сумму цифр. ◄

Если мы хотим представить себе наглядно числа по модулю N, можно вообразить кольцевое шоссе длиной в N километров: на нём километровые столбы будут $0, 1, 2, \ldots, N-1$, далее идёт столб N, который на том же месте, где 0, затем $N+1$ на том же месте, где 1, и так далее. Можно пойти в другую сторону и поставить столб -1 на том же месте, где $N-1$: это соответствует тому, что $(N-1) \equiv (-1) \pmod N$.

Для действительных (не целых) чисел равенство дробных частей можно назвать сравнением по модулю 1. Точнее говоря, целой частью числа x называют наибольшее целое число, не превосходящее x; целую часть обозначают обычно $\lfloor x \rfloor$. Разницу $x - \lfloor x \rfloor$ называют дробной частью и иногда обозначают $\{x\}$. Числа с одинаковой дробной частью отличаются на целое число единиц; можно сказать, что они «сравнимы по модулю 1». Обратите внимание, что с отрицательными числами та же ситуация: $\lfloor -2.3 \rfloor = -3$ и $\{-2.3\} = 0.7$.

► Нарисуйте на плоскости пары (x, y), для которых $\lfloor x \rfloor = \lfloor y \rfloor$. Тот же вопрос для условия $\{x\} = \{y\}$.◄

5. Таблицы сложения и умножения по модулю N

Мы уже видели таблицы сложения и умножения по модулю 2 и 3. Теперь мы знаем, что аналогичную таблицу можно составить по любому модулю. Например, по модулю 10 получатся таблицы сложения и умножения, которые получаются из обычных, если оставить только последнюю цифру:

+	0	1	2	3	4	5	6	7	8	9
0	0	1	2	3	4	5	6	7	8	9
1	1	2	3	4	5	6	7	8	9	0
2	2	3	4	5	6	7	8	9	0	1
3	3	4	5	6	7	8	9	0	1	2
4	4	5	6	7	8	9	0	1	2	3
5	5	6	7	8	9	0	1	2	3	4
6	6	7	8	9	0	1	2	3	4	5
7	7	8	9	0	1	2	3	4	5	6
8	8	9	0	1	2	3	4	5	6	7
9	9	0	1	2	3	4	5	6	7	8

×	0	1	2	3	4	5	6	7	8	9
0	0	0	0	0	0	0	0	0	0	0
1	0	1	2	3	4	5	6	7	8	9
2	0	2	4	6	8	0	2	4	6	8
3	0	3	6	9	2	5	8	1	4	7
4	0	4	8	2	6	0	4	8	2	6
5	0	5	0	5	0	5	0	5	0	5
6	0	6	2	8	4	0	6	2	8	4
7	0	7	4	1	8	5	2	9	6	3
8	0	8	6	4	2	0	8	6	4	2
9	0	9	8	7	6	5	4	3	2	1

Скажем, $7 \times 8 = 56$, поэтому в восьмой клетке седьмого ряда мы пишем $56 \bmod 10 = 6$.

► Найдите последнюю ненулевую цифру числа $10! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot 10$? Тот же вопрос для числа $100!$ ◄

► На какие цифры могут оканчиваться точные квадраты (=квадраты целых чисел)? ◄

► В обычной алгебре уравнение $x^2 = x$ имеет только решения $x = 0$ и $x = 1$. Покажите, что по модулям 10, 100 и 1000 оно имеет ещё два решения. (Скажем, существуют два трёхзначных числа x, для которых $x^2 \equiv x \pmod{1000}$, то есть x^2 оканчивается на x.) ◄

Имея такое «сложение» и «умножение», стоит задуматься о том, верны ли для них знакомые из школьного курса алгебры правила. Скажем, будет ли сложение коммутативно ($a + b = b + a$)? Это можно проверить по таблице: она симметрична относительно главной диагонали. Но можно и так сообразить: чтобы найти, скажем, число в седьмой клетке третьего ряда, надо сложить числа с остатками 3 и 7 и взять остаток по модулю 10. А в третьей клетке седьмого ряда мы складываем числа с остатками 7 и 3. Поскольку обычное сложение коммутативно, то получится одно и то же.

Аналогичное рассуждение показывает, что и другие обычные свойства сложения и умножения сохраняются:

- $a + (b + c) = (a + b) + c$ (ассоциативность сложения);
- $ab = ba$ (коммутативность умножения)
- $a(bc) = (ab)c$ (ассоциативность умножения)
- $a(b + c) = ab + ac$ (дистрибутивность)

Числа 0 и 1 (точнее, классы чисел, дающие остаток 0 и 1 при делении на N) обладают обычными свойствами:

- $0 + a = 0$;
- $1 \cdot a = a$;
- $0 \cdot a = 0$.

Для каждого остатка a есть противоположный, который в сумме с a равен нулю (а именно, $N - a$ при $a \neq 0$, и нуль при $a = 0$). Его естественно обозначить $-a$. Как обычно, противоположные остатки можно использовать при вычитании. В самом деле, по определению вычитания, $b - a$ это такое x, что $a + x = b$. В качестве такого x годится $b + (-a)$, потому что $(b + (-a)) + a = b + ((-a) + a) = b + 0 = b$ (пользуемся ассоциативностью, свойствами противоположного, свойствами нуля). Решение уравнения $a + x = b$ не только существует, но и единственно: прибавляя $(-a)$ к обеим частям, получаем $(-a) + (a + x) = (-a) + b$, но левая часть равна $((-a) + a) + x = 0 + x = x$.

Так что с вычитанием проблем нет и всё как обычно. А вот с делением, то есть с решением уравнений $ax = b$, всё сложнее. Решая такое уравнение, мы должны в a-ой строке найти b, и посмотреть, в каком столбце это b окажется (то есть на что надо умножить a, чтобы получить b).

► Разделить 3 на 7 по модулю 10, пользуясь таблицей. ◄

При делении 3 на 7 проблем не возникло, потому что число 3 встречается в строке 7 ровно один раз. Вот если бы мы делили 3 на 6, то ничего бы не вышло: в строке 6 таблицы числа 3 нет (там стоят только чётные числа). А при делении 4 на 6 мы не знали бы, что выбрать: и $6 \cdot 4$, и $6 \cdot 9$ кончаются на 4, так что уравнение $6x = 4$ в остатках (или, как говорят ещё, «вычетах») по модулю 10 имеет два решения.

Чтобы в этом разобраться, нам понадобится понятие обратимого элемента.

6. Обратимые элементы по модулю N

Остаток (вычет) по модулю N называется *обратимым*, если в произведении с каким-то другим остатком он даёт 1. Другими словами, a обратим, если уравнение $ax = 1$ имеет решение, то есть если в строке a таблицы умножения встречается единица.

► По таблице умножения по модулю 10 найдите все обратимые элементы. ◄

На обратимые элементы можно делить: *если a обратим, то уравнение $ax = b \pmod N$ имеет единственное решение при любом b.*

В самом деле, обозначим через a^{-1} элемент, который в произведении с a даёт единицу. (Пока мы не знаем, что такой только один, так что обозначим какой-то.) Положим $x = a^{-1}b$. Тогда $ax = a(a^{-1}b) = (aa^{-1})b = 1 \cdot b = b$, так что одно решение уравнения $ax = b$ мы нашли. Оно будет единственным: если $ax = b$, то $a^{-1}(ax) = a^{-1}b$, но $a^{-1}(ax) = (a^{-1}a)x = 1 \cdot x = x$, так что для x есть только одна возможность.

Остаётся выяснить, какие элементы (остатки, вычеты) обратимы по модулю N. Решив задачу в начале этого раздела, мы знаем, что по модулю 10 это будут $1, 3, 7, 9$.

► При любом N легко указать как минимум два обратимых элемента. Что это за элементы? ◄

Ответ на этот вопрос составляет первый существенный результат этого раздела.

Теорема 1. *Обратимыми по модулю N являются те и только те остатки, которые взаимно просты с N.*

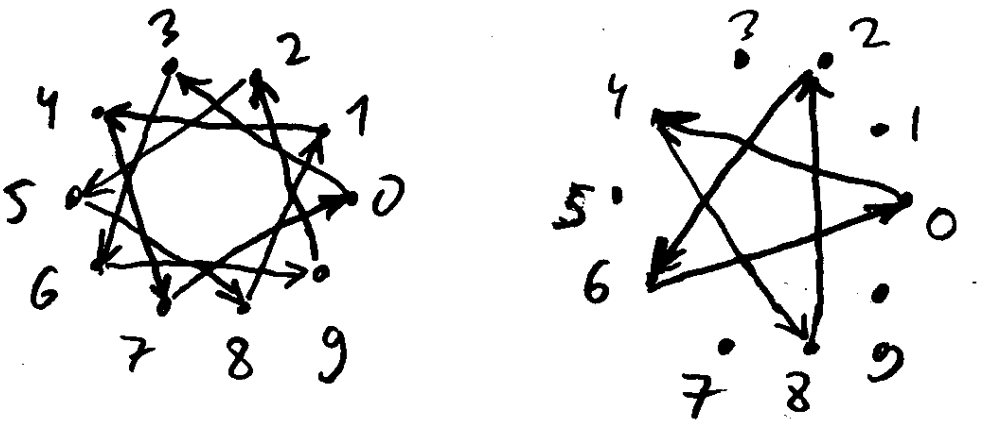
Взаимно простыми называются числа, которые не имеют общего делителя, не считая 1.

Например, по модулю 10 отпадают все чётные числа (у которых общий делитель 2), а также число 5 (общий делитель 5), остаются как раз $1, 3, 7, 9$.

В частности, если N простое (не разлагается в произведение меньших чисел), то общим делителем могут быть только 1 и N, так что все остатки, кроме нуля, взаимно просты с N и обратимы. Для простого модуля всё как в обычной алгебре: делить можно на всё кроме нуля. (Математики выражают это словами: вычеты по простому модулю p образуют *поле*.)

Прежде чем доказывать эту теорему, представим себе наглядно, что означает обратимость остатка a. Для этого вспомним, что остатки по модулю N можно расположить на круговом шоссе длины N, как автобусные остановки. Запустим маршрут автобуса, которые делает остановки каждые a километров: у него первая остановка будет в a, вторая в $a + a = 2a$, третья в $3a$ (всё по модулю N, естественно). Обратимость означает, что таким странным способом все остановки будут обслужены (в строке таблицы умножения встретятся все остатки).

Вот две картинки, показывающие, что по модулю 10 остаток 3 будет обратимым, а остаток 4 нет:



в первом случае мы проходим все остановки, а во втором не все (только чётные).

Кстати, из этой картинки хорошо видно, что если элемент обратим (мы попадаем в соседнюю остановку), то деление всегда возможно (мы попадём во все остановки). В самом деле, если через k шагов мы попали в соседнюю, то ещё через k мы попадём в следующую и так далее, пройдя через все остановки.

Теперь легко доказать простую часть утверждения: если N и шаг a имеют общий делитель d, то элемент a необратим (мы не попадём в соседнюю остановку). В самом деле, будем отмечать остановки через $d, 2d, 3d$ и так далее от начальной. Поскольку N делится на d, то мы дойдём до N-й (то есть начальной) остановки, пройдя весь круг. Если a кратно d, то a-автобус будет останавливаться только в выделенных остановках, и в соседнюю никогда не попадёт.

Осталось доказать сложную часть утверждения: если N и шаг a не имеют общих делителей, то все остановки будут обслужены. Прежде чем это сделать, мы ненадолго отвлечёмся и сделаем несколько простых замечаний об ориентированных графах.

7. Перестановки, ориентированные графы и циклы

Ориентированный граф состоит из нескольких (конечного числа) точек, называемых *вершинами* графа. Между некоторыми из них проведены стрелки, называемые *рёбрами* графа. Такая стрелка (ребро) выходит из одной вершины и ведёт в другую (или в ту же самую — разрешаются циклы). Для каждой вершины можно подсчитать, сколько стрелок из неё выходит, это число называют *исходящей степенью* вершины. Можно также подсчитать, сколько стрелок в неё входит, это число называют *входящей степенью*.

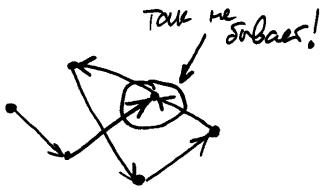
► Докажите, что сумма исходящих степеней всех вершин равна сумме входящих степеней всех вершин.◄

Нас сейчас интересует лишь частный случай, когда обе степени равны 1, то есть из каждой вершины выходит одна стрелка и в каждую вершину входит одна стрелка. Примером такого графа является *цикл* $a \to b \to c \to \ldots \to p \to q \to a$.

Теорема 2. *Ориентированный граф, в котором каждая вершина имеет входящую и исходящую степень* 1*, разбивается на непересекающиеся* (*не имеющие общих вершин и рёбер*) *циклы.*

На языке теории множеств каждый такой граф представляет собой взаимно однозначное отображение множества всех его вершин в себя, или, как говорят, *перестановку* множества вершин. Утверждение этой теоремы формулируют тогда так: *всякая перестановка разлагается в произведение непересекающихся циклов.*

Доказательство совсем простое: возьмём какую-то вершину и пойдём по стрелкам. Рано или поздно мы впервые попадём в уже посещённую вершину. Эта вершина должна быть начальной, иначе в одну вершину ведут две стрелки, что противоречит предположению.



Тем самым возникает цикл, и из его вершин больше никаких стрелок не выходит и в них не входит. Если в цикл вошли не все вершины графа, повторяем рассуждение с остатком, пока граф не разобьётся на циклы.

Вернёмся к нашей задаче об обратимости элемента a. В нашем случае вершины графа расположены по кругу, как вершины правильного N-угольника. Стрелки (рёбра графа) соответствуют движению автобуса, проезжающего a перегонов до следующей остановки. Другими словами, из

вершины x (остатка по модулю N) стрелка ведёт в вершину $x + a$. По построению из каждой вершины выходит только одна стрелка, и входит тоже только одна, из вершины $x - a$ (однозначность вычитания). Нам надо доказать, что если a взаимно просто с N, то есть только один цикл, включающий все вершины.

Пусть цикл, начатый из вершины 0, включает только часть вершин. Посмотрим, в какие вершины он попадает. Пусть ближайшая из них (по кругу) имеет номер d. Ясно, что построение цикла можно начать с любой вершины (круг везде одинаков), поэтому если мы начнём с d, то следующая обслуженная вершина будет $2d$. Значит, в цикл входят вершины 0, потом (по кругу) d, потом $2d$ и так далее, пока мы не вернёмся обратно в начальную вершину 0. В результате мы пройдём полный круг в N вершин, двигаясь шагами по d, поэтому N кратно d. С другой стороны, вершина a обслужена (на первой же остановке автобуса), поэтому и a тоже кратно d. Получается, что d — общий делитель N и a. А мы предположили, что они взаимно просты, то есть $d = 1$, что означает, что все вершины (все остатки по модулю N) попадают в один цикл.

▶ Покажите, что построенное в этом рассуждении число d делится на любой общий делитель чисел a и N и является их наибольшим общим делителем. ◀

Последняя задача показывает, что уравнение $ax = b \pmod N$ разрешимо тогда и только тогда, когда b делится на НОД(a, N) (наибольший общий делитель чисел a и N).

▶ Покажите, что в общем случае граф отображения $x \mapsto x + a$ состоит из НОД(a, N) циклов одинаковой длины $N/\,\text{НОД}(a, N)$. ◀

Если наше доказательство с циклами и многоугольниками, вписанными в окружность, кажется непонятным или недостаточно строгим, ничего страшного: другое доказательство будет приведено в следующем разделе.

8. Обратимые элементы и диофантовы уравнения

Мы только что доказали, что всякий остаток a, взаимно простой с N, имеет обратный. Но как этот обратный найти? Можно пробовать все возможности по очереди. При небольших N это реально, но что делать, если, скажем $N = 5704689200685129054721$ (хотите верьте, хотите проверьте, но это простое число, один из делителей числа $2^{128} + 1$). Проверить столько вариантов даже и с помощью какого-нибудь суперкомпьютера не удастся. Нет ли какого-то более быстрого способа? (Научно говоря, нет ли полиномиального алгоритма — время работы которого было бы ограничено многочленом от числа цифр в N?)

Оказывается, что такой алгоритм есть, и по существу он был известен ещё Евклиду в древней Греции. Для начала переформулируем нашу задачу более симметричным образом. Мы хотим решить сравнение $ax \equiv b \pmod N$. Это сравнение означает, что разность $b - ax$ должна быть целым числом, кратным N, то есть должна равняться Ny при каком-то y. Таким образом, нам надо решить в целых числах уравнение $ax + Ny = b$. Здесь a, N, y — известные целые числа, и мы ищем целые x и y, при которых левая часть равна правой. Найдя их, мы можем про y забыть, а x будет решением.

Уравнения, в которых нам нужны целые решения, называются *диофантовыми*, в честь другого древнего грека — Диофанта. (Он, правда, не такой древний, как Евклид, и жил уже после Р.Х.)

Таким образом, теорема об обратных элементах сводится к такому утверждению: *уравнение* $ax + Ny = b$ *имеет решение в целых числах тогда и только тогда, когда* b *кратно* НОД(a, N). Из него следует, что при взаимно простых a и N это уравнение имеет решение при любом b.

Чтобы сделать запись ещё более симметричной и забыть о неравноправии a и N в исходной постановке, заменим букву N на b, а b на c. Таким образом, нам надо доказать такое утверждение:

Теорема 3. *Пусть* a, b, c *— произвольные целые числа. Уравнение* $ax + by = c$ *имеет целочисленное решение тогда и только тогда, когда число* c *кратно* НОД(a, b).

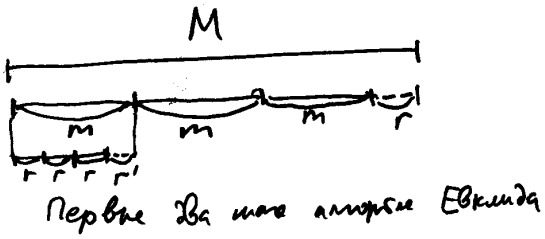
Опять в одну сторону это очевидно: если d есть (наибольший) общий делитель a и b, то $ax + by$

всегда делится на d, так что уравнение может иметь решение только при c, кратном d. Интересно обратное утверждение, и мы его докажем с помощью алгоритма Евклида, заодно показав, как можно искать это самое решение на практике.

9. Алгоритм Евклида

Давайте сначала посмотрим на этот алгоритм в том виде, как это было у Евклида в его учебнике геометрии («Начала»). Пусть нам надо найти общую меру двух отрезков, то есть третий отрезок, который укладывается целое число раз в первом и во втором. (Другими словами, мы хотим найти меру длины, при которой длины обоих данных нам отрезков будут целыми числами.)

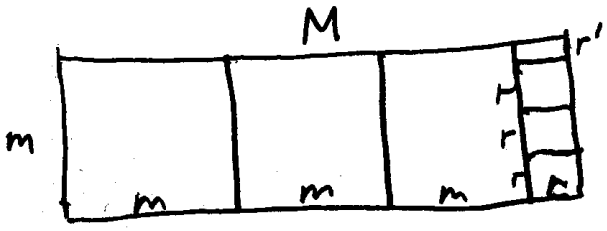
Евклид предлагает делать это так: будем откладывать меньший отрезок m внутри большего M. Если нам повезёт и он уложится целое число раз, то меньший отрезок m и будет общей мерой. Если нет, то он уложится сколько-то раз и что-то (уже меньшее m) останется. Обозначим этот остаток за r. Теперь повторим эту процедуру с отрезками m и r, укладывая меньший из них (то есть r) в большем. Снова либо он уложится без остатка, либо получится остаток r', меньший r, и мы применяем алгоритм к r и r', и так далее.



Первые два шага алгоритма Евклида

Алгоритм заканчивает свою работу, когда и если меньший отрезок укладывается в большем без остатка.

Объясняя эту процедуру школьникам, часто считают отрезки сторонами прямоугольников: сначала есть прямоугольник $M \times m$, от которого отрезают квадраты $m \times m$, пока это возможно. Когда останется прямоугольник $r \times m$, в котором $r < m$, от него отрезают квадраты $r \times r$, остаётся прямоугольник $r \times r'$ с $r' < r$. Можно считать, что у нас есть автомат, который отрезает от прямоугольника квадрат со стороной, равной меньшей стороне прямоугольника (он сам разбирается, какая сторона меньше). Мы кромсаем прямоугольник на квадратные части, засовывая остаток снова и снова в этот автомат.



По существу это, конечно, тот же самый процесс, может быть, в немного более наглядной форме.

Основное свойство алгоритма Евклида теперь можно сформулировать так:

Теорема 4. *Если исходные отрезки имеют общую меру, то алгоритм заканчивает работу и последний отрезок (тот, что уложится целое число раз) будет наибольшей их общей мерой. Если же исходные отрезки не имеют общей меры, то алгоритм никогда не остановится.*

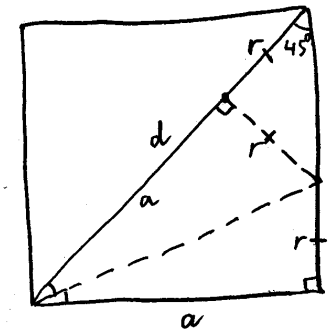
Доказательство. (1) Если исходные отрезки имеют общую меру d, примем её за единицу измерения. Тогда оба наших отрезка имеют целую длину, и мы делим одно целое число (M) на другое

(m) с остатком r, потом делим m на r с остатком r' и так дальше, при этом $m > r > r' > \ldots$ (остаток меньше делителя) и все они целые, так что бесконечно это продолжаться не может. В терминах прямоугольников: нарисуем всё по клеточкам на клетчатой бумаге с шагом d, тогда и все разрезы пройдут по клеточкам, и возможных мест разреза конечное число.

(2) Теперь покажем, что если алгоритм заканчивает работу, то последний отрезок будет общей мерой. В самом деле, если принять его за единицу измерения, то предыдущий отрезок будет целым числом (ведь последний укладывается целое число раз), перед ним тоже будет целый отрезок (равный сумме нескольких целых чисел) и пр. В терминах квадратов: если принять самый маленький квадрат при разрезании за клетку на клетчатой бумаге, то все большие квадраты и составленные их них прямоугольники пойдут по линиям сетки, и их стороны будет целыми, то есть сторона маленького квадрата будет общей мерой. (Отсюда уже следует последнее утверждение: если общей меры нет, то алгоритм не заканчивает работу.)

(3) Осталось показать, что последний отрезок будет *наибольшей* общей мерой. Более того, он даже будет кратен любой другой общей мере d. Почему? Приняв d за единицу измерения, мы видим, что оба наших отрезка имеют целую длину, и все последующие отрезки будут целыми. Значит, ответ алгоритма Евклида тоже целый (кратен d), как мы и утверждали.

Алгоритм Евклида можно использовать, чтобы показать, что сторона квадрата и его диагональ не имеют общей меры. На рисунке сторона бумажного квадрата загнута по биссектрисе так, чтобы она пошла по диагонали. Видно, что сторона a укладывается в диагонали d один раз и остаётся отрезок r. Этот же отрезок можно найти ещё в двух местах, и если его один раз (уже на следующем шаге алгоритма Евклида) отложить на стороне, то останется как раз диагональ квадрата со стороной r. Таким образом, через два шага мы приходим к тому же соотношению между диагональю квадрата и его стороной, и всё повторяется, так что алгоритм никогда не закончит работу.



► Посмотрите внимательно на картинку и восстановите пропущенные геометрические детали. ◄

► Каково должно быть отношение сторон прямоугольника, чтобы после отрезания от него одного квадрата получался прямоугольник, подобный исходному? Покажите, что стороны такого прямоугольника несоизмеримы (=не имеют общей меры). ◄

Это отношение называют *золотым сечением*; говорят, что оно встречается в красивых постройках и картинах.

► Найдите целые положительные числа x, y, z, t, при которых

$$\frac{17}{10} = x + \cfrac{1}{y + \cfrac{1}{z + \cfrac{1}{t}}}$$

При чём тут алгоритм Евклида? ◄

10. Алгоритм Евклида и диофантовы уравнения

Весь этот экскурс в историю нужен нам не сам по себе, а чтобы научиться решать уравнения $ax + by = c$ при c, кратном НОД(a, b). Ясно, что достаточно уметь это делать при $c = $ НОД(a, b), потом можно умножить на отношение $c/$ НОД(a, b), раз оно целое.

Мы можем найти $d = $ НОД(a, b), разрезая на квадраты прямоугольник $a \times b$, это будет сторона наименьшего из квадратов. (Про квадраты мы говорим только для наглядности, можно было говорить про отрезки.) Ключевое наблюдение: *все отрезки, появляющиеся в ходе алгоритма, представляются в виде $ax + by$ с некоторыми целыми a и b*. Как говорят, они являются «целочисленными линейными комбинациями» a и b. Это же относится и к последнему отрезку, то есть d, и мы получаем искомое решение.

Почему они будут целочисленными линейными комбинациями? Пусть мы сначала делим a на b с остатком r, тогда $a = bq + r$, и $r = a - bq$ представлен такой комбинацией. Теперь мы делим b на r, получаем остаток r', то есть $b = q'r + r'$, и $r' = b - q'r$ есть целочисленная комбинация b и r. Вспомним, что само r есть комбинация a и b, получится

$$r' = b - q'r = b - q'(a - bq) = b - q'a + q'bq = (qq' + 1)b - q'a,$$

то есть r' тоже есть целочисленная комбинация a и b, и так далее.

► Какая целочисленная комбинация получается для r' на приведённой выше картинке? ◄

Для программистов всё сказанное можно заменить алгоритмом:

```
{a,b - целые положительные числа}
делим a на b, получаем частное q и остаток r
M:=b; m:=r;
Ma:=0; Mb:=1; ma:=1; mb:=-q;
{M>m>=0; НОД(M,m)=НОД(a,b); M=Ma*a+Mb*b; m=ma*a+mb*b}
пока m>0:
    делим M на m, получаем частное q и остаток r
    M, Ma, Mb, m, ma, mb <- m, ma, mb, r, Ma-q*ma,Mb-q*mb
{m=0; НОД(a,b)=НОД(M,m)=НОД(M,0)=M=Ma*a+Mb*b}
ответ: M=НОД(a,b); (x,y)=(Ma,Mb) - решение ax+by=НОД(a,b)
```

Здесь стрелка `<-` означает одновременное присваивание: шесть переменных получают новые значения, указанные справа, при этом при вычислении выражений в правой части используются старые значения. Если, как это принято, выполнять присваивания последовательно, то надо написать

```
M_new:=m; Ma_new:=ma; Mb_new:=mb
m_new:=r; ma_new:=Ma-q*ma; mb_new:=Mb-q*mb
M:=M_new; Ma:=Ma_new; Mb:=Mb_new
m:=m_new; ma:=ma_new; mb:=mb_new
```

Этот алгоритм иногда называют «расширенным алгоритмом Евклида» (по-английски “extended Euclidean algorithm”). Что можно сказать о количестве операций в этом алгоритме? Можно заметить, что за два шага алгоритма Евклида больший отрезок (большее число для алгоритма с целыми числами) уменьшается по крайней мере вдвое. В самом деле, если меньший отрезок не превосходит половины большего, то это случится уже за один шаг; если же нет, то на первом шаге частное равно 1, а остаток не больше половины большего отрезка, так что за два шага это всё равно случится. Поэтому число шагов для n-битовых чисел равно $O(n)$. Каждый шаг требует деления с остатком — если это делать «уголком», как в школе, то число действий будет $O(n^2)$, так что всего получается $O(n^3)$. Конечно, если следовать определению буквально и делить с помощью последовательного вычитания (как мы делали, доказывая существование частного и остатка), то будет плохо (число действий будет экспоненциальным).

► Напишите алгоритм решения уравнения $ax + by = \text{НОД}(a, b)$, использующий три соотношения:

- $\text{НОД}(2a, 2b) = 2\,\text{НОД}(a, b)$;
- $\text{НОД}(2a, b) = \text{НОД}(a, b)$ при нечётном b;
- $\text{НОД}(a, b) = \text{НОД}(a - b, b)$, которое верно всегда, но нужно при нечётных a и b и $a \geqslant b$.

В каждом случае задачу решения уравнения можно рекурсивно свести к той же задаче для новой пары. Сколько действий требует этот вариант алгоритма Евклида? ◄

► Известно, что алгоритм Евклида для пары целых положительных чисел (a, b) требует n шагов. Каково наименьшее возможное значение a? (Считаем, что $a \geqslant b$.) ◄

Если не интересоваться алгоритмической стороной дела, то можно доказать теорему о разрешимости уравнений $ax + by = c$ следующим образом. Напомним, что мы называли *идеалом* множество целых чисел, которое вместе с любыми двумя элементами содержит их сумму и разность, и вместе с любым элементом содержит все его кратные. В качестве примера идеала мы приводили множество I_c всех кратных некоторого числа c. Такие идеалы алгебраисты называют *главными*. Оказывается, что (для целых чисел) других и нет. В самом деле, пусть I — некоторый идеал. Если он состоит только из нуля, то $c = 0$. Если в нём есть ненулевые числа, возьмём минимальное по модулю число c. Будем считать, что оно положительно (перейдя к $-c$, если надо). Теперь ясно, что все числа в I кратны c: если какое-то $x \in I$ даёт ненулевой остаток r при делении на c, то $x = qc + r$ и $r = x - qc$ есть разность двух чисел из I и потому тоже принадлежит I, что противоречит минимальности c. Такое число c, при котором $I = I_c$, называется *образующей* идеала I.

Пусть теперь a, b — произвольные целые числа, а I состоит из всех целых чисел, которые можно представить в виде $ax + by$. Это будет идеал, и по доказанному он совпадает с некоторым I_c. Поскольку $a, b \in I$, то они кратны c, так что c будет общим делителем. С другой стороны, любой общий делитель d чисел a, b делит все элементы I, в том числе и c, так что c будет наибольшим общим делителем (и кратен любому делителю — это мы тоже доказали заодно). Наконец, c представим в виде $ax + by$ по построению I.

► Покажите, что общие кратные двух чисел a и b образуют идеал, и выведите отсюда, что наименьшее общее кратное $\text{НОК}(a, b)$ является делителем любого общего кратного. ◄

► Покажите, что $\text{НОД}(a, b) \cdot \text{НОК}(a, b) = ab$. ◄

► Покажите, что для данных целых a, b числа x, при которых ax кратно b, образуют идеал. Как найти его образующую, зная a, b и $\text{НОД}(a, b)$? ◄

► При каких целых a, b, c, d уравнение $ax + by + cz = d$ разрешимо в целых числах? ◄

11. Однозначность разложения на множители

Теперь мы можем доказать теорему об однозначности разложения целых чисел на простые множители, которую иногда торжественно называют «основной теоремой арифметики». Напомним, что целое число $p > 1$ называется *простым*, если оно не разлагается в произведение меньших чисел (то есть не имеет делителей, кроме 1 и p).

Теорема 5. *Всякое целое положительное число, большее* 1*, разлагается на простые множители, причём единственным образом: любые два разложения отличаются только перестановкой сомножителей.*

(Можно сказать, что единица тоже разлагается в произведение нуля простых множителей — если принять, что произведение нуля сомножителей равно 1. Как вы думаете, кстати, что следует считать суммой нуля слагаемых?)

Трудность с этой теоремой в том, что она (видимо, со школы) кажется само собой разумеющейся, и непонятно, что тут доказывать. Тем не менее доказательство требует некоторых усилий (которые мы по большей части уже проделали).

Существование разложения совсем просто. Если данное число N простое, то получилось разложение из одного сомножителя. Если нет, то $N = ab$ для каких-то меньших a, b. Если a и b простые, то хорошо, если нет, то разложим их в произведение меньших и так далее до тех пор, пока дальше уже ничего не раскладывается, поскольку числа простые. (Формально говоря, мы рассуждаем по индукции и считаем, что для меньших чисел a и b существование разложения уже известно.)

Единственность разложения. Пусть некоторое число N имеет два разложения

$$N = p_1 \cdot p_2 \cdot \ldots \cdot p_n = q_1 \cdot q_2 \cdot \ldots \cdot q_m$$

(они могут отличаться и числом сомножителей, m не обязано равняться n). Мы хотим получить противоречие. Сократим на общие сомножители (если они есть). Если сократится не всё, то получим два разложения одного числа, не имеющих общих сомножителей

$$p_1 \cdot p_2 \cdot \ldots \cdot p_n = q_1 \cdot q_2 \cdot \ldots \cdot q_m.$$

Как говорят, «без ограничения общности» можно предположить, что общих сомножителей нет (сократив на них, если есть).

В чём тут противоречие? С одной стороны, левая часть делится на p_1 (можно было бы взять любой другой p_i, если там несколько сомножителей). А правая часть равна произведению чисел, ни одно из которых не делится на p_1: они ведь простые и p_1 среди них по предположению нет. Осталось доказать, что такого не бывает, то есть доказать следующую лемму.

Лемма 1. *Если p — простое число, то произведение чисел, не делящихся на p, не может делиться на p.*

По существу мы уже это доказали: в терминах вычетов по модулю p нужно доказать, что произведение ненулевых вычетов не равно нулю. А мы знаем, что если $a \neq 0 \pmod p$, то a взаимно просто с p, поэтому a обратим и уравнение $ax = 0$ имеет единственное решение (нулевое).

Можно повторить эти рассуждения более подробно. Во-первых, достаточно доказать лемму для двух сомножителей. Если есть, скажем, три числа a, b, c, не делящиеся на p, то мы применяем лемму для двух сомножителей a и b и заключаем, что ab не делится на p. После этого уже можно применить лемму к двум сомножителями ab и c и заключить, что $(ab)c$ не делится на p. (Аналогично для любого числа сомножителей — формально говоря, мы используем индукцию по числу сомножителей.)

Для двух сомножителей мы можем доказать более общий факт:

Лемма 2. *Если ab делится на n, и при этом a взаимно просто с n, то b делится на n.*

(Более общий он потому, что если a не делится на простое p, то a взаимно просто с p — других общих делителей быть не может.)

Доказательство. Если $\text{НОД}(a, n) = 1$, можно найти x, y, для которых $ax + ny = 1$. Умножим это равенство на b, получим, что

$$b = abx + ny.$$

Осталось заметить, что оба слагаемых в правой части делятся на n: по предположению ab делится на n, и очевидным образом ny делится на n. Значит, и сумма (то есть b) делится на n, что и требовалось доказать.

► Как вычислить $\text{НОД}(a, b)$ и $\text{НОК}(a, b)$, зная разложения a и b на простые множители? Докажите, что $\text{НОД}(a, b) \cdot \text{НОК}(a, b) = ab$, используя (очевидное) равенство $\min(x, y) + \max(x, y) = x + y$. Можно ли использовать эту задачу, чтобы найти более эффективный алгоритм поиска наибольшего общего делителя? ◄

Однозначность разложения на множители делает очевидными многие утверждения, например, такое: *если число a делится на b и на c, и при этом b и c взаимно просты, то a делится на bc.* В самом деле, если a делится на b, то разложение a на множители содержит все множители из разложения b (и ещё что-то), поскольку можно разложить по отдельности b и a/b. Аналогичным образом разложение a содержит все множители из разложения c. Но общих множителей у b и c нет, так что это разные множители, и в разложении a можно выделить разложение для bc.

Можно, конечно, обойтись и без теоремы об единственности разложения: поскольку b и c взаимно просты, то $bx + cy = 1$ при некоторых x и y, так что $a = a(bx + cy) = abx + acy$, и оба слагаемых abx и acy делятся на bc (почему?) — но это всё же выглядит более искусственно...

► Докажите, что если a взаимно просто с b и с c по отдельности, то оно взаимно просто с bc. (Можно воспользоваться разложением на множители, а можно и перемножить равенства $ax+by = 1$ и $au + cv = 1$.) ◄

12. Китайская теорема об остатках

Как найти число, которое даёт остаток 1 при делении на 23 и остаток 1 при делении на 37? Понятно как: надо взять число, делящееся и на 23, и на 37, например, $23 \cdot 37$, и прибавить единицу.

Немного сложнее найти число, которое даёт остаток 22 при делении на 23 и 36 при делении на 37. Надо заметить, что если прибавить 1, то получится число, делящееся и на 23, и на 37, так что можно попробовать $23 \cdot 37 - 1$, и всё получится.

А что делать, если надо найти число, дающее (скажем) остаток 17 при делении на 23, и одновременно остаток 24 при делении на 37? И вообще, возможно ли это? Оказывается, что возможно, причём для любой комбинации остатков, поскольку 23 и 37 взаимно просты. Это гарантирует следующая теорема, традиционно называемая «китайской теоремой об остатках» (Chinese remainder theorem).

Теорема 6. *Пусть числа m и n взаимно просты, и пусть u и v — любые целые числа. Тогда можно найти число x, для которого $x \equiv u \pmod m$ и одновременно $x \equiv v \pmod n$.*

Прежде чем доказывать это, посмотрим на какой-нибудь пример. Числа 23 и 37 слишком большие, возьмём, скажем, 3 и 4 и составим таблицу:

x	0	1	2	3	4	5	6	7	8	9	10	11	
$x \bmod 3$	0	1	2	0	1	2	0	1	2	0	1	2	
$x \bmod 4$	0	1	2	3	0	1	2	3	0	1	2	3	

(Дальше можно уже не продолжать, поскольку 12 делится и на 3, и на 4, и всё повторится с начала.)

Так вот, китайская теорема учит, что в двух нижних строках встретятся все возможные комбинации остатков: любой из остатков $0, 1, 2$ комбинируется с любым из остатков $0, 1, 2, 3$ (всего как раз 12 вариантов: три варианта mod 3 комбинируются с четырьмя вариантами mod 4, так что каждая комбинация встречается ровно по одному разу.)

Эта таблица подсказывает доказательство китайской теоремы об остатках. Построим аналогичную таблицу, записав в первой строке числа $0, 1, 2, \ldots, mn - 1$, во второй строке их остатки при делении на m (получится $0, 1, 2, \ldots, m - 1$, повторённое n раз), а в третьей строке их остатки при делении на n (получится $0, 1, 2, \ldots, n-1$, повторённое m раз). Покажем, что все комбинации остатков (их будет mn) встретятся ровно по одному разу. Для этого заметим, что никакая комбинация не может повториться дважды: если числа u и v дают одинаковые остатки и при делении на m, и при делении на n, то их разность $u - v$ делится и на m, и на n. Как мы уже видели, это значит, что $u - v$ кратно mn (поскольку m и n взаимно просты), а у нас все остатки при делении на mn представлены по одному разу.

Осталось заметить, что если ни одна из mn комбинаций не повторяется в имеющихся mn столбцах, то придётся использовать все mn комбинаций. (Если N голубей разместить в N норах, причём в каждой норе не больше одного голубя, то все норы будут заполнены. Это называется по-английски pigeon-hole principle, а по-русски принципом Дирихле, поскольку Дирихле использовал это соображение, изучая приближения действительных чисел рациональными.)

► Докажите, что если числа m, n, k попарно взаимно просты, то для любых u, v, w найдётся такое x, что

$$x \equiv u \pmod{m}; \quad x \equiv v \pmod{n}; \quad x \equiv w \pmod{k}.$$

◄

Приведённое нами доказательство китайской теоремы об остатках не даёт алгоритма нахождения искомого x. Но легко понять, что эта задача сводится к уже известным. В самом деле, нам нужно найти x, для которого $x \equiv u \pmod{m}$ и $x \equiv v \pmod{n}$. Первое условие означает, что $x = u + tm$, и надо искать t. Второе условие, переписанное в терминах t, означает, что $u + tm \equiv v \pmod{n}$, то есть надо решить сравнение $tm \equiv -u \pmod{n}$, что мы уже умеем делать с помощью алгоритма Евклида (надо найти обратный для m в вычетах по модулю m, используя взаимную простоту m и n).

Заодно мы получили другое доказательство китайской теоремы об остатках.

13. Малая теорема Ферма

Этот раздел называется «малая теорема Ферма», хотя исторически это не вполне справедливо: как и в случае «большой», или «последней» теоремы Ферма (о том, что уравнение $x^n + y^n = z^n$ при $n > 2$ не имеет решений в целых положительных числах), сам Ферма не предложил доказательства — то ему полей не хватало (с большой теоремой Ферма), то он боялся надоесть адресату (je vous envoierois la démonstration, si je n'appréhendois d'être trop long). Но в этом случае не понадобилось ждать до конца XX века, доказательство вроде бы знал Лейбниц в том же XVII веке, а опубликовано оно было Эйлером в 1736 году, и вполне можно допустить, что это доказательство было известно Ферма.

Вот о каком утверждении идёт речь:

Теорема 7. *Если p — простое число, то*

$$a^{p-1} \equiv 1 \pmod{p}$$

при любом a, не делящемся на p.

Вот одно из доказательств (пожалуй, наиболее естественное). Рассмотрим все ненулевые остатки по модулю p. Их всего $p-1$: от единицы до $p-1$. Сделаем их вершинами графа, проведя стрелки из x в ax (умножение по модулю p). Разница с тем графом, что у нас уже был, двоякая: во-первых, мы рассматриваем не все p остатков, а только $p-1$ ненулевых остатков. Но главное, мы не прибавляем a, а умножаем на a: стрелка из x ведёт не в $x + a$, а в ax. Так что красивых картинок с правильными звёздчатыми многоугольниками, как раньше, не получится.

Но тем не менее из каждой вершины выходит одна стрелка (по построению) и в каждую вершину входит одна стрелка (поскольку a не делится на p, оно обратимо и уравнение $ax = b$ имеет единственное решение при любом b, то есть в b входит ровно одна стрелка). Значит, граф, как и раньше, разбивается, на циклы.

Что это за циклы? Начнём с какой-то вершины, например, с остатка 1, и будем идти по стрелкам, пока цикл не замкнётся:

$$1 \mapsto a \mapsto a^2 \mapsto a^{k-1} \mapsto a^k = 1$$

Здесь k — минимальная степень a, равная 1 (как мы знаем, цикл может замкнуться, только придя в начальную вершину). Если начать построение с какой-то вершины b вместо 1, то получится цикл

того же размера:

$$b \mapsto ab \mapsto a^2b \mapsto a^{k-1}b \mapsto a^kb = 1b = b.$$

Почему? Ясно, что если $a^k = 1$, то и $a^kb = b$, так что после k шагов цикл замкнётся. Но надо понять, почему этот цикл не замкнётся раньше. В самом деле, если $a^lb = b$, то можно домножить справа на b^{-1} (ведь остаток b тоже обратим), и получится $a^l = 1$, а мы предположили, что k минимальное.

Таким образом, $p-1$ ненулевых остатков разбиваются на циклы одинакового размера k, так что $p-1$ делится на k, то есть $p-1 = km$ при каком-то m. Тогда

$$a^{p-1} = a^{km} = (a^k)^m = 1^m = 1 \pmod p,$$

что и требовалось доказать.

Есть и другие доказательства, приведём два любопытных.

Достаточно доказать, что $a^p \equiv a \pmod p$, потом можно сократить на обратимый элемент a. Представим себе, что имеется a цветов, и ими нужно раскрасить круг из p равных секторов (каждый сектор в какой-то цвет), причём раскраски, отличающиеся лишь поворотом круга, считаются за одну. Сколькими способами это можно сделать? Можно все сектора красить в один цвет, это можно сделать a способами (по числу цветов). А можно использовать более одного цвета, на это остаётся $a^p - a$ способов, но они делятся на группы по p, отличающихся поворотами (каждую неоднородную раскраску можно повернуть p различными способами). Получается ответ $a + \frac{a^p-a}{p}$. Но число способов должно быть целым, так что $a^p - a$ делится на p.

Второе доказательство использует отображение остатков по модулю p в себя, заданное формулой $f(x) = x^p$. Нам надо доказать, что оно тождественное, то есть что $f(x) = x$ при всех x. Это вытекает из трёх его свойств:

- $f(0) = 0$;
- $f(1) = 1$;
- $f(x+y) = f(x) + f(y)$.

В самом деле, третье свойство можно по индукции распространить на любое число слагаемых, скажем,

$$f(x+y+z) = f((x+y)+z) = f(x+y) + f(z) = f(x) + f(y) + f(z)$$

и так далее, поэтому

$$f(x) = f(1+1+\ldots+1) = f(1) + f(1) + \ldots + f(1) = 1 + 1 + \ldots + 1 = x$$

(многоточия обозначают суммы из x слагаемых). Так что достаточно доказать эти три свойства. Первые два очевидны, а третье следует из бинома Ньютона: в разложении

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

все биномиальные коэффициенты, кроме двух крайних, делятся на p (у них есть p в числителе, и это простое число, которое не может сократиться с меньшими числами в знаменателе), так что по модулю p остаются только два крайних члена, что и требовалось.

Алгебраисты сказали бы, что эти три свойства верны в любом поле «характеристики p», где сумма p единиц равна нулю — их можно ещё дополнить четвёртым свойством $f(xy) = f(x)f(y)$. Но отображение f уже не будет (вообще говоря) тождественным; его называют «автоморфизмом Фробениуса» (слово «автоморфизм» означает, что выполнены эти четыре свойства, а Фробениус — немецкий математик, в честь которого он назван).

С точки зрения практиков, теорема Ферма может использоваться как способ убедиться, что число не простое. Скажем, можно вычислить $2^8 = 256 \equiv 4 \pmod 9$ и заключить отсюда, что число

9 не простое (если бы мы этого и так не знали). Это выглядит глупо, но на самом деле имеет вычислительный смысл: возвести какое-то a в степень $p-1$ по модулю p не так сложно (надо вычислять $a, a^2, a^4, a^8, \ldots \bmod p$ последовательными возведениями в квадрат, а потом перемножить нужные из них, следуя двоичному разложению для $p-1$), это требует полиномиального числа действий (от n для n-битового числа p), и если нам повезёт и получится не 1 по модулю p, то мы сможем убедиться, что число p составное. Можно назвать число a, для которого $a^{p-1} \not\equiv 1 \pmod p$, «свидетелем» того, что p составное; как мы обсудили, если такой свидетель известен, «проверить его показания» можно быстро.

Собственно говоря, проверка разложения на множители (перемножение сомножителей) тоже полиномиальна, но преимущество нового способа в том, что свидетелей обычно бывает много, и есть шанс наткнуться на них, взяв наугад число от 1 до $p-1$. (А на разложение на множители так просто не наткнуться; полиномиального алгоритма разложения на множители не известно, и для чисел из нескольких тысяч цифр искать разложение на множители никто не умеет.) Так что если вам нужно быстро найти какое-нибудь большое простое число, можно выбрать случайную последовательность цифр и проверить теорему Ферма для (скажем) $a = 2, 3, 5$; если она не выполнена, число не простое и надо взять другую случайную последовательность цифр, пока этот тест не будет пройден. С большой вероятностью этот способ даёт простые числа.

► Мы предлагали проверять теорему Ферма для $2, 3, 5$, но почему-то пропустили 4. Как вы думаете, почему? ◄

14. Функция Эйлера и теорема Эйлера

На самом деле приведённое доказательство малой теоремы Ферма имеет смысл не только для простых модулей. Пусть N — какое-то число, не обязательно простое. Мы можем по-прежнему рассмотреть остатки по модулю N, но взять только взаимно простые с N. (В алгебре это называют «мультипликативной группой кольца вычетов по модулю N».) Произведение двух таких остатков тоже взаимно просто с N, и все они обратимы, так что взяв какой-то остаток a среди них, мы можем построить ориентированный граф $x \mapsto xa$, и в нём из каждой вершины будет исходить одна стрелка и в каждую вершину будет входить одна стрелка. Этот граф разбивается на циклы, и по тем же причинам все циклы будут одинаковой длины. Эта длина видна на примере цикла, начинающегося с 1:

$$1 \mapsto a \mapsto a^2 \mapsto \ldots \mapsto a^k = 1,$$

где k — наименьшая степень a, равная 1 по модулю N. Значит, k является делителем числа остатков, взаимно простых с N, и мы получаем такой результат, называемый «теоремой Эйлера».

Теорема 8. *Пусть $N > 1$ — произвольное целое число, а $\varphi(N)$ — количество остатков среди $0, 1, \ldots, N-1$, взаимно простых с N. Пусть a — один из этих остатков. Тогда*

$$a^{\varphi(N)} \equiv 1 \pmod N.$$

Определённую таким образом функцию φ называют *функцией Эйлера* и традиционно обозначают буквой φ. Если число N простое, то $\varphi(N) = N-1$, и теорема Эйлера превращается в малую теорему Ферма.

Как вычислить функцию Эйлера $\varphi(N)$, зная N? Это несложно, если мы дополнительно знаем разложение числа $N-1$ на простые множители (как мы уже говорили, разложение на множители вычислительно сложно), и делается с помощью таких свойств:

- $\varphi(p^n) = p^n(1 - 1/p)$ для простого p;
- $\varphi(uv) = \varphi(u)\varphi(v)$, если u и v взаимно просты.

Первое свойство позволяет найти φ для степеней простых чисел, а второе свойство позволяет собрать из них любое число. Свойства эти нам по существу уже известны. Первое говорит, что

доля чисел, не взаимно простых с p^n (то есть делящихся на p) составляет $1/p$ среди всех чисел, и их надо вычеркнуть. Второе легко доказать, сославшись на китайскую теорему об остатках и на то, что взаимная простота с uv означает одновременную взаимную простоту с u и v. В терминах теории вероятностей можно сказать, что определённые на вычетах по модулю uv функции «остаток по модулю u» и «остаток по модулю v» независимы и имеют равномерное распределение, и вероятность события «быть взаимно простым с uv» можно вычислять как произведение вероятностей событий «быть взаимно простым с u» и «быть взаимно простым с v».

Второе свойство называют *мультипликативностью* функции Эйлера. В теории чисел встречаются разные мультипликативные функции.

▶ Покажите, что функция $d(n)$, равная числу положительных делителей числа n, мультипликативна. ◀

▶ Покажите, что функция $\sigma(n)$, равная сумме всех положительных делителей числа n, мультипликативна. ◀

15. Что дальше?

Мы познакомились с базовыми свойствами сложения и умножения по модулю N, или, как говорят, *кольца вычетов по модулю* N, стараясь не использовать алгебраического языка. Есть и много других интересных (и не таких сложных) результатов. Скажем, есть такая теорема: при простом N найдётся остаток a по модулю N, при котором все элементы $1, a, a^2, \ldots, a^{N-2}$ различны (следующий элемент по теореме Ферма равен 1). Или другой результат: ровно половина всех ненулевых вычетов по простому модулю являются точными квадратами (квадратами других вычетов). В принципе и их можно было бы изложить, не используя языка алгебры (группы, подгруппы, кольца, гомоморфизмы, многочлены и пр.), но в этом нет смысла: пора с ним познакомиться. Это можно сделать по любому учебнику алгебры — есть старинный «Курс высшей алгебры» Куроша (когда-то бывший стандартным мехматским учебником), или несколько более новый учебник Кострикина (тоже по материалам мехматского курса). Стандартные учебники, покрывающие существенно больший материал — классические книги ван дер Вардена и Ленга. В программе ВШЭ курс алгебры предусмотрен в 4 модуле первого курса.

Где все эти алгебраические премудрости используются «на практике»? Вот игрушечный пример: пусть вы хотите послать кому-то номер своей кредитной карточки, но боитесь, что ваше сообщение будет перехвачено. Тогда есть такой способ: каждую цифру a_i этого номера можно представить в виде суммы по модулю 10 случайно выбранной цифры r_i и недостающей цифры $s_i = a_i - r_i$, и послать последовательности цифр r_i и s_i (той же длины, что и номер карточки) в двух разных электронных письмах. Получив оба этих письма, адресат легко восстановил $a_i - r_i + s_i$. Но если враг перехватит только одно из них, а второе нет — то у него не будет никакой информации о вашей карточке. В самом деле, сообщение r не содержит никакой информации (это просто случайная последовательность цифр), и сообщение s в отдельности тоже (потому что при данном a это сообщение может быть любым, и все они равновероятны).

Конечно, защита тут слабая — лишь на случай перехвата одного из писем. Но, скажем, если одно посылается по электронной почте, а другое — как сообщение в Skype, то враг должен получить доступ к обоим сервисам (они разные, и каналы передачи сообщений разные). Существуют и более надёжные методы передачи сообщений по открытой сети, прежде всего так называемая система RSA (она используется в банках, в протоколах https, ssh, цифровой подписи PGP и вообще на каждом шагу). И она тоже основана на арифметике остатков. Наших знаний достаточно, чтобы понять, почему она не искажает сообщения — но вот доказать, что её нельзя взломать, мы не можем, и никто пока не может — если кто-то вдруг научится быстро разлагать большие числа на множители, эта система, а вместе с ней и все банки, рухнет. Но для этого (быстрого разложения) нужны либо новые алгоритмы, либо квантовые компьютеры (которые, если вдруг их удастся построить, могут разлагать числа на множители — это обнаружил P. Shore). Другие системы шифрования основаны на более сложных алгебраических методах (эллиптические кривые) —

так что знания по алгебре лишними не будут.

Вообще алгебраические методы незаменимы там, где нужно построить «регулярно устроенное» семейство объектов с хорошими свойствами. Попробуйте, например, найти в 7-элементном множестве как можно больше 3-элементных подмножеств, любые два из которых имеют общий элемент. (Сколько дней подряд можно назначать по три дежурных в походе из 7 человек, чтобы наряды на любые два дня имели ровно одного общего?) Не так просто — если не знать, что двумерные подпространства в трёхмерном пространстве над полем вычетов по модулю 2 содержат (не считая нуля) по три элемента и любые два их них пересекаются по одномерной прямой, содержащей ровно один ненулевой элемент.