Основы высшей алгебры и теории кодирования

Ю. И. Журавлёв, Ю. А. Флёров, М. Н. Вялый

Учебные материалы к курсу
«Основы высшей алгебры и теории кодирования»
Вариант от 28 апреля 2019 г.

Основным учебником для курса «Основы высшей алгебры и теории кодирования» ФУПМ МФТИ является учебник Ю. И. Журавлёва, Ю. А. Флёрова, М. Н. Вялого «Дискретный анализ. Основы высшей алгебры». Материалы из этого учебника излагаются в данном тексте с существенными дополнениями. Разбираются решения многих типовых задач курса.

Для облегчения работы с учебником, в текст вставлены так называемые «контрольные вопросы». Это простые упражнения, которые должны легко решаться. Если решение контрольного вопроса вызывает затруднения, это верный признак того, что нужно перечитать предшествующее изложение.

Текущая версия текста черновая, в ней возможны опечатки, ошибки и пропуски. По мере исправления текущий текст будет заменяться улучшенными версиями. Чтобы было легче различать версии, внизу страниц текста указана дата компиляции. Просьба сообщать о замеченных неточностях М. Н. Вялому по адресу `vyalyi@gmail.com`

Содержание

1 Примеры

Из элементарной математики известны свойства основных числовых систем: целых чисел $\mathbb{Z}$, рациональных чисел $\mathbb{Q}$, действительных чисел $\mathbb{R}$ и комплексных чисел $\mathbb{C}$.

Нас будут интересовать обобщения этих числовых систем, сохраняющие те или иные их свойства.

В этом разделе мы напоминаем кратко свойства основных математических понятий, знание которых предполагается при изучении курса.

1.1 Множества

Значительная часть математики основана на понятии множества. Хотя мы предполагаем знакомство читателя с этим понятием, в данном разделе напомним те сведения о множествах, которые будут использоваться в дальнейшем.

Неформально, множество — это совокупность элементов, не имеющая никакой структуры. Элементы могут быть произвольной природы, в том числе элементы множества сами по себе могут быть множествами. Отношение «элемент x принадлежит множеству M» обозначается $x \in M$. Множество однозначно задаётся своими элементами. Это означает, что множества A и B равны тогда и только тогда, когда любой элемент, принадлежащий множеству A, принадлежит множеству B, и наоборот: любой элемент принадлежащий множеству B принадлежит A.

Пустое множество $\varnothing$ не содержит ни одного элемента.

Контрольный вопрос 1.1. Проверьте, пользуясь определением равенства множеств, что пустое множество определено однозначно.

Конечные множества часто задают списком элементов, взятым в фигурные скобки. Например, $\{1, 2, 3, 4, 5\}$ — это множество из 5 элементов 1, 2, 3, 4 и 5.

Контрольный вопрос 1.2. Сколько элементов в множестве $\{\varnothing, \{1, 2\}, \{2, 3, 4, 5\}\}$?

Более общая форма этой записи: фигурные скобки с условием. Мы не перечисляем все элементы (для бесконечных множеств это попросту невозможно), а указываем свойство, которому удовлетворяют элементы множества и только они. Например, множество чётных чисел можно задать так:

$$2\mathbb{Z} = \{x : x = 2k,\ k \in \mathbb{Z}\}.$$

Пример 1.3. Какое множество задаётся записью

$$\{x : x^2 + 1 = 0,\ x \in \mathbb{R}\}\,? \tag{1.1}$$

Это множество состоит из тех действительных чисел, квадрат которых равен -1. Таких чисел нет, поэтому запись (1.1) — это ещё одно имя для пустого множества. □

В определении равенства множеств указаны два свойства. Если оставить лишь одно из них, получаем *отношение включения* $A \subseteq B$: по определению множество A включено в B (другое название — множество A является *подмножеством* множества B), если любой элемент множества A принадлежит множеству B.

Упражнение 1.4. Проверьте, что отношение включения удовлетворяет свойствам рефлексивности ($A \subseteq A$), транзитивности (из $A \subseteq B$, $B \subseteq C$ следует $A \subseteq C$) и антисимметричности (если $A \subseteq B$ и $B \subseteq A$, то $A = B$).

Как и в случае сравнения чисел, используется также отношение строгого включения: $A \subset B$ означает, что A является подмножеством B, но не равно ему (то есть, в множестве B есть такие элементы, которые не принадлежат A).

На множествах определены операции объединения, пересечения и разности. Напомним их определения и стандартные обозначения для них:

Объединение множеств A и B (обозначается $A \cup B$) состоит из тех и только тех элементов, которые входят[1)] хотя бы в одно из множеств A, B.

Аналогично определяется объединение произвольного семейства множеств.

Пересечение множеств A и B (обозначается $A \cap B$) состоит из тех и только тех элементов, которые входят в оба множества A, B.

Аналогично определяется пересечение произвольного семейства множеств.

Разность множеств A и B (обозначается $A \setminus B$) состоит из тех и только тех элементов, которые входят в множество множества A, но не входят в множество B.

Ещё одна операция с множествами — *декартово произведение*. Элементами декартова произведения $A \times B$ двух множеств A и B являются все упорядоченные пары (a, b), где $a \in A$ и $b \in B$, и только они.

Аналогично определяется декартово произведение нескольких множеств. В частности, декартова степень A^n — это множество, состоящее из последовательностей элементов множества, длина которых равна n.

1.2 Числовые системы

Начнём с повторения основных свойств числовых систем, упомянутых выше. Изложение в этом разделе неполно, цель состоит лишь в том, чтобы напомнить определения числовых систем и их свойства. Упражнения в этом разделе носят технический характер и не добавят ничего к пониманию свойств числовых систем. Однако выполнение этих упражнений — очень полезная тренировка в умении рассуждать логически, опираясь только на небольшое количество явно сформулированных свойств (в общем случае такие рассуждения называются аксиоматическим методом). Для алгебры такой тип рассуждений характерен и будет постоянно встречаться в дальнейшем.

1.2.1 Натуральные числа

Все числовые системы строятся из множества натуральных чисел $\mathbb{N} = \{0, 1, \dots\}$. Натуральные числа называются так потому, что они являются результатом подсчёта

[1)] «Элемент входит в множество» — это то же самое, что «элемент принадлежит множеству».

количества предметов в некоторой совокупности (то есть в некотором множестве). Поэтому основная операция с натуральными числами — прибавление единицы (добавили ещё один элемент). На основе этой операции определяются также сложение и умножение натуральных чисел. Они обладают следующими свойствами: для любых натуральных чисел x, y, z выполняются равенства

$$a + (b + c) = (a + b) + c \quad \text{(ассоциативность сложения)}, \tag{1.2}$$

$$a(bc) = (ab)c \quad \text{(ассоциативность умножения)}, \tag{1.3}$$

$$a + b = b + a \quad \text{(коммутативность сложения)}, \tag{1.4}$$

$$ab = ba \quad \text{(коммутативность умножения)}, \tag{1.5}$$

$$a(b + c) = ab + ac \quad \text{(дистрибутивность)}, \tag{1.6}$$

$$0 \cdot a = 0 \quad \text{(нуль и умножение)}, \tag{1.7}$$

$$0 + a = a \quad \text{(нейтральность нуля)}, \tag{1.8}$$

$$1 \cdot a = a \quad \text{(нейтральность единицы)}. \tag{1.9}$$

Последние два свойства означают, что существуют такие натуральные числа (они обозначены 0 и 1 соответственно), что прибавление нуля или умножение на единицу не меняет числа.

Помимо арифметических операций на множестве натуральных чисел определено отношение сравнения «число a меньше числа b». Оно обозначается $a < b$ и обладает следующими свойствами:

$$\text{если } a < b, \text{ то неверно, что } b < a \quad \text{(антисимметричность)}, \tag{1.10}$$

$$\text{если } a < b \text{ и } b < c, \text{ то } a < c \quad \text{(транзитивность)}, \tag{1.11}$$

$$\text{если } a \neq b, \text{ то } a < b \text{ или } b < a \quad \text{(линейность порядка)}, \tag{1.12}$$

$$\text{если } a < b, \text{ то } a + c < b + c \text{ для любого } c \quad \text{(монотонность сложения)}, \tag{1.13}$$

$$\text{если } a < b, \text{ то } ac < bc \text{ для любого } c \neq 0 \quad \text{(монотонность умножения)}, \tag{1.14}$$

$$0 < 1 \quad \text{(положительность единицы)} \tag{1.15}$$

Конечно, свойств арифметических операций и сравнения чисел гораздо больше. Но из сформулированных выше простых свойств можно выводить более сложные.

Для этого часто используется принцип математической индукции. Пусть есть некоторое утверждение о натуральных числах, обозначим его $A(n)$. **Принцип математической индукции**: из справедливости двух утверждений — «$A(0)$» (база индукции) и «для любого n из $A(n)$ следует $A(n+1)$» (шаг индукции) — следует, что для любого n верно $A(n)$.

Мы приведём примеры доказательств по индукции. Утверждения, которые доказываются ниже, очевидны сами по себе. Смысл этих формальных доказательств в том, что они показывают, какие утверждения логически следуют из сформулированных свойств арифметических операций и сравнения натуральных чисел, а также

принципа математической индукции.

Утверждение 1.5. *Для любого натурального x выполняется неравенство $x < x + 1$.*

Доказательство. Используя коммутативность сложения, нейтральность нуля относительно сложения, положительность единицы и монотонность сложения, получаем искомое неравенство $x = 0 + x < 1 + x$. □

Утверждение 1.6. 0 — *наименьшее натуральное число.*

Доказательство. Докажем по индукции утверждение $0 \leqslant x$ для любого натурального x. (Знак $\leqslant$ как обычно обозначает комбинированное утверждение «меньше или равно».)

База индукции $0 \leqslant 0$ очевидна. Если $0 \leqslant x$, то $0 \leqslant x < x + 1$ в силу утверждения 1.5 и транзитивность сравнения. Это шаг индукции.

Осталось применить принцип математической индукции. □

Утверждение 1.7. *Если натуральное число x отлично от 0, то существует такое натуральное число y, что $x = y + 1$.*

Доказательство. База индукции: число 0. Для него утверждение выполняется, так как посылка ложна (по правилам логики в этом случае истинно составное высказывание «если ..., то ...», которое называется *импликацией* или логическим следованием).

Шаг индукции. Пусть $x = y + 1$. Тогда $x + 1 = (y + 1) + 1$. □

Утверждение 1.8. *Для каждого натурального числа x число $x + 1$ — наименьшее из натуральных чисел, бóльших x.*

Доказательство. База индукции: 1 — наименьшее из положительных натуральных чисел (то есть не равных 0). Доказательство базы само происходит по индукции, доказываем утверждение «$x \geqslant 1$ или $x = 0$» для всех натуральных чисел. База индукции для этого утверждения очевидна. Для шага индукции нужно неравенство $x+1 \geqslant 1$, которое следует из минимальности нуля (утверждение 1.6) и монотонности сложения.

Шаг индукции. Пусть для натурального числа x число $x + 1$ — наименьшее из натуральных чисел, бóльших x. Из монотонности сложения следует, что $x+2 > x+1$. Докажем, что $x + 2$ — наименьшее из натуральных чисел, бóльших $x + 1$.

Пусть $x + 2 > y > x + 1 > 0$. Утверждение 1.7 говорит, что $y = k + 1$ для некоторого натурального числа k.

Сравним числа k и x. Из неравенства $k + 1 > x + 1$ следует, что $k \neq x$. Если $k < x$, то также $k + 1 < x + 1$ по монотонности сложения. Значит, $k > x$ и по предположению индукции $k \geqslant x + 1$. Но тогда $y = k + 1 \geqslant (x + 1) + 1 = x + 2$ (здесь 2 — это обозначение суммы двух единиц).

Получили противоречие с неравенством $x+2 > y$. Это противоречие показывает, что такого y не существует. Шаг индукции доказан. Утверждение справедливо по принципу математической индукции. □

Часто оказывается удобной другая формулировка принципа математической индукции.

Лемма 1.9. *Всякое непустое множество натуральных чисел имеет наименьший элемент.*

Доказательство. Рассмотрим некоторое множество натуральных чисел S. Пусть в нём нет наименьшего числа. Докажем, что тогда S пусто.

В качестве утверждения $A(n)$ возьмём такое: «если $0 \leqslant x \leqslant n$, то x не принадлежит S».

База индукции $A(0)$ выполняется, так как если бы 0 принадлежал S, то это было бы наименьшее в S натуральное число (выше доказано, что 0 — наименьшее среди всех натуральных чисел, значит, и среди чисел из множества S).

Шаг индукции: если $A(n)$ истинно, то $A(n+1)$ тоже истинно. Докажем от противного. Предположим, что $A(n+1)$ ложно. Это означает, что для какого-то $x \in S$ выполняются неравенства $0 \leqslant x \leqslant n+1$. Так как в S нет наименьшего числа, то какое-то $y < x$ также должно входить в множество S. При этом утверждение $A(n)$ истинно, то есть $n < y < x \leqslant n+1$. Но такие неравенства несовместны в силу утверждения 1.8.

По принципу математической индукции для любых натуральных чисел $x \leqslant n$ справедливо, что x не принадлежит S. Для произвольного x выберем $n = x+1$. По утверждению 1.5 выполняется неравенство $x < x+1 = n$. Таким образом, ни одно натуральное число не принадлежит S, то есть это множество пусто, что и требовалось доказать. □

На самом деле лемма 1.9 равносильна принципу математической индукции.

Упражнение 1.10. Выведите принцип математической индукции из леммы 1.9. (Это проще доказательства леммы 1.9.)

Очень часто оказывается удобной более общая форма принципа математической индукции.

Теорема 1.11 (индукция по всем меньшим числам)**.** *Пусть для утверждения о натуральных числах $B(n)$ выполняются два свойства: (база индукции). «$B(0)$» истинно и (индуктивный переход) «если $B(k)$ истинно для любого $k < n$, то $B(n)$ истинно».*

Тогда $B(n)$ истинно для любого n.

Доказательство. Применим стандартный принцип математической индукции к такому утверждению: $A(n)$= «$B(k)$ истинно для любого $k < n$».

Утверждение $A(0)$ истинно, поскольку нет ни одного натурального числа, меньшего нуля.[2)]

[2)] Это следует из законов формальной логики, мы здесь опускаем подробности. Сомневающийся читатель может усложнить доказательство так, чтобы избегнуть суждений о пустых множествах натуральных чисел.

Пусть выполнено $A(n)$. Тогда по условию теоремы (индуктивный переход) выполняется $B(n)$. То есть $B(k)$ истинно для любого $k < n$ и для $k = n$. Значит, $B(k)$ истинно для для любого $k < n + 1$. То есть $A(n + 1)$ истинно.

Применяя стандартный принцип математической индукции, приходим к выводу, что $A(n)$ истинно для всех n. Но из истинности $A(n + 1)$ следует истинность $B(n)$. Поэтому $B(n)$ истинно для любого n. □

1.2.2 Целые числа

Обратные операции: вычитание и деление не всегда определены для натуральных чисел. Напомним, что обратные операции задаются как решения уравнений

$$a + x = b,$$
$$ax = b.$$

Если существует решение первого уравнения, оно называется разностью чисел b и a и обозначается $b - a$. Если существует решение второго уравнения, оно называется частным чисел b и a и обозначается b/a (или $\frac{b}{a}$, это одно и то же обозначение, которое записывается по-разному для удобства чтения формул).

Чтобы определить обратные операции для произвольных пар чисел, натуральные числа нужно расширить. В целых числах $\mathbb{Z} = \{\dots, -1, 0, 1, \dots\}$ для любой пары чисел определена операция вычитания. Целые числа — это натуральные числа и отрицательные целые числа, которые получаются добавлением знака «$-$» к ненулевому натуральному числу (добавление знака к нулю ничего не меняет: $-0 = 0$).

Несложная, хотя и утомительная, проверка показывает, что свойства (1.2–1.9) выполняются и для целых чисел.

Упражнение 1.12. Вспомните определения операций сложения и умножения для целых чисел. Проверьте выполнение свойств (1.2–1.9) для целых чисел.

Аналогично (за одним уточнением) и для сравнения целых чисел.

Упражнение 1.13. Вспомните определение сравнения целых чисел. Проверьте для целых чисел выполнение свойств (1.10–1.13) и уточнённого свойства (1.14):

$$\text{если } a < b, \quad \text{то } ac < bc \text{ для любого } c > 0,$$
$$\text{если } a < b, \quad \text{то } ac > bc \text{ для любого } c < 0.$$

Другими словами, умножение на положительное число не меняет знак неравенства, а умножение на отрицательное число меняет знак неравенства на противоположный.

Для целых чисел лемма 1.9 неверна. В частности, нет наименьшего целого числа: для любого натурального числа x выполняется неравенство $x < x + 1$, поэтому из монотонности умножения на -1 следует, что $-x - 1 < -x$. Однако аналог леммы 1.9 выполняется для ограниченных подмножеств целых чисел.

Упражнение 1.14. Докажите, что если все числа из непустого множества $S \subset \mathbb{Z}$ больше числа c (множество S ограничено снизу), то в S есть наименьший элемент.

Аналогично, если все числа из непустого множества $S \subset \mathbb{Z}$ меньше числа c (множество S ограничено сверху), то в S есть наибольший элемент.

Подсказка. В первом случае рассмотрите множество $S-c = \{x : x = y-c,\ y \in S\}$. Во втором — множество $c-S = \{x : x = c-y,\ y \in S\}$. Примените к этим множествам лемму 1.9.

Деление в целых числах не всегда возможно.

Пример 1.15. Проверим, что уравнение $2x = 1$ не имеет решений в целых числах. Поскольку $2 \cdot 0 = 0 < 2 \cdot 1 = 2$ (монотонность умножения на положительное число 2), то для корня уравнения $2x = 1$ получаем неравенства $0 < x < 1$. Такого целого числа не существует (утверждение 1.8). □

Однако для целых чисел выполняется более слабое, но очень полезное свойство: *закон сокращения*

$$\text{из } ax = ay \text{ и } a \neq 0 \text{ следует } x = y. \tag{1.16}$$

Упражнение 1.16. Докажите закон сокращения, используя только указанные выше свойства целых чисел.

1.2.3 Рациональные числа

На множестве рациональных чисел $\mathbb{Q}$ уже определена операция деления, хотя и только для ненулевых делителей. При этом свойства арифметических операций (1.2–1.9) для рациональных чисел по-прежнему выполняются.

Напомним, как строятся рациональные числа. Для этого используется множество *обыкновенных дробей* (обозначение такое же, как для операции деления: b/a, $a \neq 0$, a, b — целые). Первый элемент в паре называется *числителем дроби*, второй — знаменателем.

Важное отличие от предыдущего расширения (от $\mathbb{N}$ к $\mathbb{Z}$) состоит в том, что дроби задают рациональные числа неоднозначно (вспомните, что $1/2$ и $2/4$ — это одно и то же число).

Будем считать дроби a/b и c/d равными, если выполняется равенство в целых числах $ad = bc$.

Лемма 1.17. *Для равенства дробей выполняются следующие свойства:*

$$\text{если } \frac{a}{b} = \frac{c}{d}, \qquad \text{то } \frac{c}{d} = \frac{a}{b} \qquad \textit{(симметричность)}, \tag{1.17}$$

$$\text{если } \frac{a}{b} = \frac{c}{d} \text{ и } \frac{c}{d} = \frac{e}{f}, \qquad \text{то } \frac{a}{b} = \frac{e}{f} \qquad \textit{(транзитивность)}, \tag{1.18}$$

$$\frac{a}{b} = \frac{a}{b}, \qquad \textit{(рефлексивность)}. \tag{1.19}$$

Доказательство. Рефлексивность очевидна из определения.

Симметричность следует из симметричности равенства для целых чисел: если $ad = bc$, то $bc = ad$.

Для доказательства транзитивности нужно повозиться. Основная идея — применить закон сокращения для целых чисел. Действительно, запишем посылку свойства транзитивности, используя определение равенства дробей:

$$\text{если } ad = bc \text{ и } cf = ed, \text{ то } af = eb.$$

Перемножим два равенства в посылке, получим $adcf = bced$ или $(cd)af = (cd)eb$ (тут мы используем коммутативность и ассоциативность умножения целых чисел). Остаётся применить закон сокращения.

Однако приведённое выше рассуждение ошибочно, точнее говоря, неполно. Ведь закон сокращения выполняется только для ненулевых множителей. Случай $cd = 0$ требует отдельного анализа. В этом случае $c = 0$, так как $d \neq 0$ по определению обыкновенной дроби.

Контрольный вопрос 1.18. Как вывести из $cd = 0$, $d \neq 0$ равенство $c = 0$ с помощью закона сокращения?

Раз $c = 0$, то и $a = 0$, и $e = 0$. Значит $af = 0 \cdot f = 0 = 0 \cdot b = eb$. На этом доказательство транзитивности заканчивается. □

Отношение, которое обладает свойствами (1.17–1.19), называется *отношением эквивалентности*. Отношение эквивалентности похоже на равенство. Разница в том, что может быть много разных, но попарно эквивалентных, элементов.

Пример 1.19. Разобьём множество десятичных цифр на три группы:

$$\{0, 3, 6, 9\}; \quad \{1, 4, 7\}; \quad \{2, 5, 8\}.$$

Будем считать эквивалентными те цифры, которые попадают в одну группу. Это отношение эквивалентности, как легко проверить: рефлексивность и симметричность очевидны из определения; транзитивность также понятна: если a и b принадлежат множеству S, b и c принадлежат множеству S, то тогда все три элемента принадлежат множеству S, то есть a и c также принадлежат множеству S. □

Пример 1.19 легко обобщить. Пусть множество A разбито на множества A_i, то есть в объединении A_i дают всё множество A и при этом попарно не пересекаются: $A_i \cap A_j = \varnothing$ при $i \neq j$. Тогда принадлежность пары элементов одному множеству разбиения задаёт отношение эквивалентности, рассуждение точно такое же как в примере 1.19.

Такое обобщение по существу исчерпывает возможные отношения эквивалентности.

Лемма 1.20. *Пусть $a \sim b$ — отношение эквивалентности на множестве A. Тогда существует такое разбиение A на подмножества A_i, что $a \sim b$ равносильно тому, что $a, b \in A_i$ для какого-то i.*

Множества A_i называются *классами эквивалентности*.

Доказательство. Для каждого a определим множество $[a] = \{x : x \sim a\}$ тех элементов, которые эквивалентны a. Из рефлексивности отношения эквивалентности следует, что $a \in [a]$, а также из $a \in [b]$ следует $b \in [a]$.

Тогда $[a] = [b]$ равносильно тому, что $a \sim b$. Действительно, применив определение множества $[a]$, получаем, что из $a \in [b]$ следует $a \sim b$. В обратную сторону используем транзитивность: пусть $a \sim b$ и $x \in [a]$, то есть $x \sim a$; тогда из транзитивности $x \sim b$, то есть $x \in [b]$. Это означает, что $[a] \subseteq [b]$. Включение в обратную сторону доказывается аналогично.

С другой стороны, если $[a] \neq [b]$, то $[a] \cap [b] = \varnothing$. Действительно, если $x \in [a] \cap [b]$, то $x \sim a$, $x \sim b$. Из транзитивности и симметричности получаем, что $a \sim b$, по доказанному выше отсюда получаем, что $[a] = [b]$, и приходим к противоречию.

Итак, построенные множества $[a]$ либо совпадают, либо не пересекаются, причём эквивалентность $a \sim b$ равносильна тому, что эти элементы лежат в одном из построенных классов. Таким образом, мы получили разбиение на классы эквивалентности. □

Множества $[a]$, возникающие в этом доказательстве, называются *классами эквивалентности с представителем* a.

Вернёмся к рациональным числам. По определению, рациональное число — это класс эквивалентности введённого выше отношения равенства дробей. Для простоты обозначений классы эквивалентности не отличаются на письме от дробей. Мы не пишем $[1/2]$, когда говорим о рациональном числе, которое является классом эквивалентности, содержащим дроби $a/2a$, $a \in \mathbb{Z} \setminus \{0\}$.

Точно так же поступают во многих книгах и с другими, более сложными, отношениями эквивалентности. Ниже мы в основном используем более развёрнутые обозначения, чтобы облегчить понимание текста. Но, конечно, в обозначениях рациональных чисел следуем традиции, опять-таки для облегчения понимания.

Теперь определим арифметические операции для рациональных чисел и отношение сравнения.

Арифметические операции определяются для обыкновенных дробей следующим образом:

$$\frac{a}{b} + \frac{c}{d} = \frac{ad + bc}{bd}\,,$$
$$\frac{a}{b} \cdot \frac{c}{d} = \frac{ac}{bd}\,.$$

Отношение сравнения определяется так: $a/b < c/d$ равносильно тому, что $ad < bc$, если $bd > 0$, и $ad > bc$, если $bd < 0$.

Лемма 1.21. *Эти операции согласованы с отношением равенства дробей: если дроби α и β равны и дроби γ и δ равны, то*

$$\alpha + \gamma = \beta + \delta; \quad \alpha \cdot \gamma = \beta \cdot \delta$$

и $\alpha < \gamma$ равносильно $\beta < \delta$.

Эта лемма должна доказываться в школьном курсе математики, так как без неё невозможны корректные рассуждения с рациональными числами. Её доказательство несложно и использует приведённые выше свойства операций с целыми числами. Для полноты мы приводим краткое изложение доказательства, но настоятельно советуем читателю провести его самостоятельно во всех деталях.

Доказательство. Возьмём две пары равных дробей

$$\frac{a}{b} = \frac{c}{d}; \quad \frac{e}{f} = \frac{g}{h}.$$

По определению равенства дробей это означает, что выполняются равенства

$$ad = bc, \quad eh = fg. \tag{1.20}$$

Перемножим эти равенства и перегруппируем множители. Получим равенство

$$(ae) \cdot (dh) = (bf) \cdot (cg),$$

которое по определению равенства дробей и операции умножения дробей равносильно искомому равенству произведений

$$\frac{a}{b} \cdot \frac{e}{f} = \frac{c}{d} \cdot \frac{g}{h}\ .$$

Равенство сумм

$$\frac{a}{b} + \frac{e}{f} = \frac{c}{d} + \frac{g}{h}$$

равносильно равенству

$$dh(af + be) = bf(ch + dg)$$

в целых числах. Это равенство легко следует из равенств (1.20):

$$dh(af + be) = ad \cdot fh + eh \cdot bd = bcfh + fgbd = bf(ch + dg).$$

Осталось разобраться со сравнениями дробей. Тут удобно вначале рассмотреть частный случай и доказать, что

$$\frac{a}{b} < \frac{e}{f} \quad \text{равносильно} \quad \frac{-a}{-b} < \frac{e}{f},$$

использовав монотонность умножения на -1 целых чисел: $af < eb$ равносильно $-af > -eb$. После этого можно свести общий случай к случаю положительных знаменателей, для которого проверка неравенства прямолинейна: для положительных b, d, f, h из $ad = bc$ и $eh = fg$ следует цепочка равносильных равенств с целыми числами, каждый переход в которой состоит в умножении на положительное число:

$$af < be \ \Leftrightarrow\ bcf = adf < ebd \ \Leftrightarrow\ bcfh < bedh = bdfg \ \Leftrightarrow\ ch < dg.$$

Крайние неравенства в этой цепочке равносильны $a/b < e/f$ (первое) и $c/d < g/h$ (второе). Знаменатели, напомним, положительные. □

Упражнение 1.22. Проверьте выполнение свойств (1.2–1.9) для операций с рациональными числами.

Упражнение 1.23. Проверьте для рациональных чисел выполнение свойств (1.10–1.13) и уточнённого свойства (1.14), указанного в упражнении 1.13.

Отметим, что в отличие от целых чисел, рациональные числа относительно операции сравнения обладают свойством плотности: между любыми двумя различными рациональными числами есть рациональное число (на самом деле, бесконечно много чисел, как легко проверить).

Упражнение 1.24. Докажите, что в рациональных числах уравнение $a+x=b$ имеет единственное решение при любых a, b, а уравнение $ax=b$ имеет единственное решение при любом b и любом $a\neq 0$.

1.2.4 Действительные числа и числовая прямая

Теперь напомним как возникают действительные числа. Если предыдущие переходы возникали из алгебраических соображений (расширить числовую систему так, чтобы некоторые уравнения стали разрешимыми), то переход от рациональных чисел к действительным преследует другую цель: гарантировать существование пределов для монотонных ограниченных последовательностей.

Это предмет математического анализа. Мы пропускаем аккуратные определения действительных чисел. Заметим лишь, что для них также выполняются отмеченные выше свойства арифметических операций и сравнения чисел.

Приведём также геометрическую интерпретацию действительных чисел. Если выбрать на прямой точки P_0 и P_1, то точки такой *числовой прямой* взаимно однозначно соответствуют действительным числам: числу x соответствует такая точка P_x, что (1) отношение расстояний P_0P_x и P_0P_1 равно x; (2) если число x положительное, то P_x лежит на том же луче с концом в P_0, что и P_1, а если отрицательное, то P_x и P_1 лежат на разных лучах с концом в P_0. Это соответствие показано на рисунке 1.

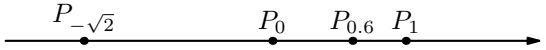


Рис. 1: Числовая прямая

1.2.5 Комплексные числа

Последняя числовая система в этом кратком обзоре: комплексные числа. Её можно строить двумя способами и удивительным образом получается одно и то же.

Первый способ состоит в том, что к действительным числам добавляется корень уравнения $x^2+1=0$ (обозначается i, называется мнимой единицей) и все линейные комбинации $a+bi$, $a,b\in\mathbb{R}$. Операции определяются как

$$(a+bi)+(c+di)=(a+c)+(b+d)i;$$
$$(a+bi)\cdot(c+di)=(ac-bd)+(ad+bc)i.$$

Сравнение комплексных чисел не определено.

Другой способ определить комплексные числа — геометрический. Рассмотрим координатную плоскость $\mathbb{R}^2$. Векторы на этой плоскости и будут комплексными числами.

Операции определяются так. Сложение чисел — это обычное сложение векторов. А умножение определяется более сложным образом. Назовём аргументом $\arg z$ комплексного числа z (сейчас это вектор на плоскости) угол, который он образует с положительным лучом оси абсцисс. Модулем $|z|$ вектора z назовём длину вектора.

Умножение комплексных чисел определяется так: модули перемножаются, а аргументы складываются. См. рис. 2.

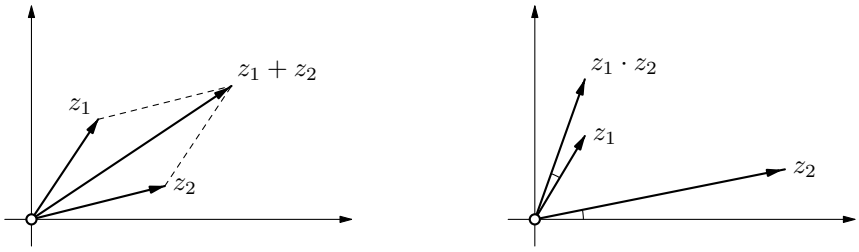


Рис. 2: Операции с комплексными числами

Соответствие между этими двумя разными определениями устанавливается следующим образом: вектору с координатами (a, b) сопоставляется число $a + bi$.

Таким образом, ось абсцисс отождествляется с действительными числами, а ось ординат — с чисто мнимыми, то есть числами вида ai, $a \in \mathbb{R}$.

Операция сложения в этих двух определениях задаётся согласовано, как нетрудно видеть. Про операцию умножения это не так понятно, но она тоже согласована. Если аргумент числа z равен φ, а модуль равен r, то из определений тригонометрических функций получаем

$$z = r(\cos\varphi + i\sin\varphi).$$

Теперь согласованность умножения в двух разных определениях получается с помощью тригонометрических формул для суммы углов:

$$\begin{aligned}(r_1\cos\varphi_1 + r_1 i\sin\varphi_1)(r_2\cos\varphi_2 + r_2 i\sin\varphi_2) = \\ = r_1r_2(\cos\varphi_1\cos\varphi_2 - \sin\varphi_1\sin\varphi_2) + r_1r_2(\cos\varphi_1\sin\varphi_2 + \sin\varphi_1\cos\varphi_2)i = \\ = r_1r_2\big(\cos(\varphi_1+\varphi_2) + i\sin(\varphi_1+\varphi_2)\big).\end{aligned}$$

Упражнение 1.25. Проверьте, что для операций с комплексными числами выполняются свойства (1.2–1.9).

Существование противоположного числа к комплексному числу достаточно очевидно. Геометрически это противоположный вектор. Существование обратного числа к ненулевому числу z также нетрудно доказать. У числа z^{-1} аргумент противоположен аргументу z, а модуль обратен модулю z. В результате умножения $z \cdot z^{-1}$ получается число с аргументом 0 и модулем 1, т.е. число 1.

Пример 1.26. Найдём обратное к комплексному числу $1 + i\sqrt{3}$.

Его модуль равен $\sqrt{1+3} = 2$, а аргумент равен $\pi/3$. Поэтому обратное число имеет модуль $1/2$ и аргумент $-\pi/3$, то есть оно равно

$$\frac{1}{2}\big(\cos(\pi/3) - i\sin(\pi/3)\big) = \frac{1 - i\sqrt{3}}{4}\,. \qquad \square$$

Есть другой способ находить обратное к комплексному числу. Для этого способа нужно ввести ещё одно понятие — комплексного сопряжения. Геометрически комплексное сопряжение — это отражение относительно оси абсцисс. То есть, сопряжённое к $z = a + bi$ число равно $\bar{z} = a - bi$.

Упражнение 1.27. Проверьте свойства комплексного сопряжения: (1) $z\bar{z} = |z|^2$, (2) $\overline{z_1 + z_2} = \bar{z}_1 + \bar{z}_2$, (3) $\overline{z_1 z_2} = \bar{z}_1 \bar{z}_2$.

Из свойства (1) легко следует формула

$$z^{-1} = \frac{\bar{z}}{|z|^2}\,.$$

Для уравнений в комплексных чисел выполняется очень сильное свойство.

Основная теорема алгебры. У всякого алгебраического уравнения

$$a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0 = 0, \quad a_n \neq 0,\ n > 0,$$

с комплексными коэффициентами есть по крайней мере один корень.

Доказательство этой теоремы здесь не приводится. Заметим лишь, что оно так или иначе использует анализ.

Мы рассмотрим один важный случай уравнений, а именно уравнения вида

$$z^n = 1.$$

Корни этого уравнения называются *корнями из единицы степени* n, множество этих корней обозначим U_n.

Лемма 1.28. *Для любого натурального* $n > 0$ *существует ровно* n *различных корней из единицы степени* n.

Доказательство. Из геометрического определения умножения комплексных чисел следует, что модуль корня из единицы должен равняться 1, так как уравнение $r^n = 1$ имеет единственное решение в неотрицательных действительных числах.

При возведении в степень n аргумент φ корня из единицы умножается на n. Чтобы получить в результате единицу, $n\varphi$ должно быть кратно 2π. Получаем соотношение

$$n\varphi = 2\pi k, \quad k \in \mathbb{Z},$$

то есть $\varphi = 2\pi k/n$, $k \in \mathbb{Z}$. Все числа с такими аргументами и модулем 1 являются корнями из единицы степени n. Среди них ровно n различных. Действительно, $2\pi k/n$ отличается от $2\pi(k + qn)/n$ на целое кратное 2π, то есть комплексные числа с такими аргументами и модулем 1 совпадают.

Осталось заметить, что любое k отличается от одного из чисел 0, 1, ..., $n-1$ на целое кратное n. Это вполне очевидно, но ниже будет разобрано подробнее (см. раздел 1.3). □

Геометрически корни из единицы образуют вершины правильного n-угольника на комплексной плоскости, см. рис. 3.

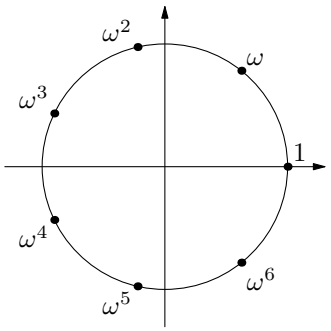


Рис. 3: Корни 7й степени из 1

Контрольный вопрос 1.29. Укажите на рис. 3 число ω^{11}.

Множества корней из единицы разных степеней могут пересекаться и даже содержаться одно в другом. Рассмотрим несколько примеров. В этих примерах мы по сути забегаем вперёд и используем свойства делимости целых чисел и свойства деления с остатком, о которых идёт речь в следующем разделе.

Пример 1.30. Докажем, что $U_2 \subseteq U_{2n}$. Это очень просто: если $z \in U_2$, то по определению $z^2 = 1$. Поэтому $z^{2n} = (z^2)^n = 1^n = 1$, то есть $z \in U_{2n}$. □

Пример 1.31. Известно, что $z \in U_{48}$. Верно ли, что $z^{18} \in U_{24}$?

Ответ положительный, в чём легко убедиться. Если $z \in U_{48}$, то есть $z^{48} = 1$, то $(z^{18})^{24} = z^{18\cdot 24} = z^{48\cdot 9} = (z^{48})^9 = 1^9 = 1$. Значит, $z^{18} \in U_{24}$. □

Пример 1.32. Найдём количество элементов в $U_{360} \cap U_{48}$.

Если $z^{360} = 1$ и $z^{48} = 1$, то в силу равенства $360 - 7 \cdot 48 = 24$ аналогично рассуждениям в предыдущих примерах получаем $z^{24} = 1$. Значит, $U_{360} \cap U_{48} \subseteq U_{24}$.

Пусть $z^{24} = 1$. Тогда $z^{360} = (z^{24})^{15} = 1$ и $z^{48} = (z^{24})^2 = 1$. Отсюда получаем включение в другую сторону $U_{24} \subseteq U_{360} \cap U_{48}$. Таким образом, выполняется равенство $U_{360} \cap U_{48} = U_{24}$.

Ответ: в $U_{360} \cap U_{48}$ ровно 24 элемента (столько же, сколько в U_{24}). □

Пример 1.33. Найдём, сколько чисел из U_{360} являются 48-ми степенями чисел из U_{360}.

Пусть $z = v^{48}$, $v \in U_{360}$. Тогда $z^{15} = (v^{360})^2 = 1$. То есть все такие z являются корнями 15-й степени из единицы.

Проверим обратное: всякий корень 15-й степени из единицы, во-первых, является корнем 360-й степени из единицы; во-вторых, является 48-й степенью некоторого корня 360-й степени из единицы.

Первое утверждение следует из равенства $360 = 15 \cdot 24$ (аналогично предыдущим вычислениям с последовательными возведениями в степень).

Докажем второе утверждение. Пусть ω_{15} имеет модуль 1 и аргумент $2\pi/15$ (это корень 15-й степени из единицы), а ω_{360} имеет модуль 1 и аргумент $2\pi/360$ (это корень 360-й степени из единицы). Из геометрического правила умножения комплексных чисел следует, что $\omega_{15} = \omega_{360}^{24}$ и $\omega_{15}^2 = \omega_{360}^{48}$.

Теперь докажем, что каждый корень 15-й степени из единицы является положительной целой степенью числа ω_{15}^2. Отсюда будет следовать наше второе утверждение, так как целые положительные степени корня 360-й степени из единицы являются корнями 360-й степени из единицы.

Фактически нужно доказать, что для любого аргумента $2\pi k/15$, k целое, найдётся такое целое положительное ℓ, что $\ell \cdot 2\pi \cdot 2/15$ отличается от $2\pi k/15$ на целое кратное 2π. Другими словами, нужно подобрать ℓ по заданному k так, чтобы разность $2k - 4\ell$ делилась бы на 30. Это равносильно тому, что $2\ell - k$ делится на 15. Годится, например, $\ell = 8k$:

$$2 \cdot 8k - k = 15k. \qquad \square$$

1.3 Элементарная теория чисел

В этом разделе мы разберём очень важный для дальнейшего пример.

Деление на множестве целых не всегда определено. Если оно все же возможно, то есть $a = qb$ для целых чисел a, b, q, то говорят, что a является *кратным* b, а b является *делителем* a. Отношение делимости принято обозначать вертикальной чертой: $b \mid a$ (b делитель a)[3)].

Однако всегда определена более сложная операция *деления с остатком*.

Определение 1.34. Если для целых чисел a, $b \neq 0$, q, r выполняется равенство

$$a = qb + r,$$

причём $0 \leqslant r < b$, то говорят, что r — *остаток* при делении числа a на число b, а q — *неполное частное*.

Пример 1.35. Равенство $8 = 2 \cdot 3 + 2$ показывает, что 2 является остатком при делении 8 на 3. Однако неправильно считать, исходя из этого равенства, что 2 является остатком при делении 8 на 2 (не выполняется неравенство $2 < 2$). $\square$

Лемма 1.36. *Деление с остатком однозначно определено для любых целых a, $b \neq 0$.*

[3)] Обратите внимание, что в книгах по элементарной математике часто используется другое обозначение делимости $a \vdots b$. В этом обозначении не только вертикальная черта заменяется на вертикальное троеточие, но и порядок другой — кратное предшествует делителю.

Доказательство. Пусть $b > 0$. Рассмотрим те кратные числа b, которые не превосходят a. Это ограниченное сверху множество и потому в нём есть наибольший элемент qb. Для него выполняются неравенства $qb \leqslant a < (q+1)b$, так как это максимальное кратное b, не превосходящее a. Поэтому для $r = a - qb$ выполняются неравенства $0 \leqslant r < b$.

Аналогично для $b < 0$. Рассмотрим те кратные числа b, которые больше a. Это ограниченное снизу множество и потому в нём есть наименьший элемент $(q+1)b$. Для него выполняются неравенства $(q+1)b > a \geqslant qb$, поэтому $0 \leqslant r = a - qb < b$. □

Легко видеть, что отношение делимости выражается через деление с остатком: $b \mid a$ равносильно $a = qb = qb + 0$ для некоторого целого q, что равносильно, в свою очередь, тому, что a даёт остаток 0 при делении на b.

Введём отношение сравнимости по модулю n и классы вычетов по модулю n.

Определение 1.37. $a \equiv b \pmod{n}$ означает, что $n \mid (a - b)$. Другими словами, a и b дают одинаковый остаток при делении на n.

Лемма 1.38. *Отношение сравнимости по модулю n — это отношение эквивалентности.*

Доказательство. Проверка всех трёх свойств отношения эквивалентности выполняется прямолинейно.

Рефлексивность: $a \equiv a \pmod{n}$, так как $(a - a) = 0 = 0 \cdot n$, то есть $n \mid (a - a)$.

Симметричность: если $a - b = qn$, то $b - a = -qn$.

Транзитивность: пусть $a - b = qn$, $b - c = tn$. Тогда $a - c = (a - b) + (b - c) = qn + tn = (q + t)n$. □

Определение 1.39. *Класс вычетов по модулю n* — это класс эквивалентности отношения сравнимости. Для краткости будем также говорить сокращённо «вычет», имея в виду класс вычетов.

Другими словами, класс вычетов образуют те целые числа, которые дают один и тот же остаток при делении на n.

Пример 1.40. Числа $\ldots, -13, -6, 1, 8, \ldots$ дают остаток 1 по модулю 7. Они образуют класс вычетов. □

Как и раньше, будем использовать обозначение $[a]$ для класса вычетов, содержащего число a. При необходимости будем писать $[a]_n$, если нужно указать модуль, по которому выполняются сравнения.

На классах вычетов можно определить арифметические операции так, чтобы для них выполнялись те же свойства, что и для операций с целыми числами. Это похоже на определение операций с дробями, но по сути даже проще.

Определение 1.41. $[a]_n + [b]_n = [a + b]_n$; $[a]_n \cdot [b]_n = [ab]_n$.

Это определение можно понимать так, что мы задаём операции на множестве остатков (ведь каждому вычету однозначно соответствует какой-то остаток). Правило естественное: выполняем сложение остатков и берём остаток при делении результата на n, аналогично с умножением. Однако определение 1.41 гораздо удобнее для вычислений и рассуждений.

Например, из этого определения сразу следует, что для сложения и умножения по модулю n выполняются те же свойства, что и для сложения и умножения целых чисел: ассоциативность сложения и умножения, коммутативность, дистрибутивность, нейтральность нуля и единицы.

Контрольный вопрос 1.42. Докажите, что для любого вычета $[x]_n$ по модулю n есть противоположный, то есть такой вычет $[y]_n$, что $[x]_n + [y]_n = [0]_n$.

Однако с определением 1.41 есть проблема: нужно доказать корректность определения, то есть доказать, что результат операции с классами вычетов не зависит от выбора представителей в классе вычетов, к которым применяется определение 1.41. Выполним эту проверку (в дальнейшем нам придётся несколько раз выполнять такую проверку для операций на классах эквивалентности).

Утверждение 1.43. *Если* $a_1 \equiv b_1 \pmod n$ *и* $a_2 \equiv b_2 \pmod n$, *то* $a_1 + a_2 \equiv b_1 + b_2 \pmod n$ *и* $a_1 a_2 \equiv b_1 b_2 \pmod n$.

Доказательство. Раскрыв определение сравнения, перепишем условия утверждения как

$$b_1 = a_1 + q_1 n, \quad b_2 = a_2 + q_2 n, \qquad q_1, q_2 \in \mathbb{Z}.$$

Поэтому

$$b_1 + b_2 = (a_1 + a_2) + (q_1 + q_2)n,$$
$$b_1 b_2 = (a_1 + q_1 n)(a_2 + q_2 n) = a_1 a_2 + n(q_1 a_2 + q_2 a_1 + q_1 q_2 n),$$

то есть выполняются сравнения $a_1 + a_2 \equiv b_1 + b_2 \pmod n$ и $a_1 a_2 \equiv b_1 b_2 \pmod n$. □

Утверждение 1.43 во многих случаях упрощает вычисления. Вместо операций с остатками можно выполнять те же операции с числами, сравнимыми с этими остатками по модулю n и это зачастую оказывается проще.

Пример 1.44. Какой остаток даёт 101^{101} при делении на 51? Прежде всего заметим, что $101 \equiv -1 \pmod{51}$. А возводить в степень -1 легко:

$$101^{101} \equiv (-1)^{101} = -1 \equiv 50 \pmod{51}.$$

Какой остаток даёт $20^2 + 21^2 + 22^2$ при делении на 23? Заменяя числа в формуле на более удобные с теми же остатками по модулю 23, получаем

$$20^2 + 21^2 + 22^2 = (-3)^2 + (-2)^2 + (-1)^2 = 9 + 4 + 1 = 14. \qquad \square$$

Однако не всегда такие замены приводят к правильному ответу. Попробуем найти остаток $\binom{32}{3}$ по модулю 33, действуя аналогично предыдущему. Раскрываем

биномиальный коэффициент и получаем:

$$\binom{32}{3} = \frac{32 \cdot 31 \cdot 30}{1 \cdot 2 \cdot 3} \equiv \frac{(-1) \cdot (-2) \cdot (-3)}{1 \cdot 2 \cdot 3} = -1 \equiv 32 \pmod{33}. \tag{1.21}$$

Проверим эту выкладку непосредственным вычислением и делением с остатком:

$$\binom{32}{3} = \frac{32 \cdot 31 \cdot 30}{1 \cdot 2 \cdot 3} = 16 \cdot 31 \cdot 10 = 4960 = 150 \cdot 33 + 10.$$

Получили остаток 10, а не 32, как в предыдущей выкладке? Одно из двух вычислений неверное, нужно разбираться, какое именно.

Утверждение 1.43 гарантирует, что $32 \cdot 31 \cdot 30 \equiv (-1) \cdot (-2) \cdot (-3) = -6 \pmod{33}$. Проблема возникает на следующем шаге, когда мы пытаемся выполнить операцию деления. Давайте запишем более подробно те рассуждения, которые скрыты за выкладкой в (1.21):

$$\frac{-6}{6} = (-1) \cdot 6 \cdot 6^{-1} = (-1) \cdot 1.$$

Второе равенство справедливо и для вычислений по модулю. А вот для придания смысла первому равенству нужно иметь вычет, который играет роль обратного к 6, то есть такой вычет $[x]$, что

$$6[x] \equiv 1 \pmod{33}$$

Нетрудно убедиться, что такого вычета нет. Написанное сравнение равносильно равенству $6x = 1 + 33y$ для некоторого целого y, которое, в свою очередь, равносильно равенству $3 \cdot (2x - 11y) = 1$. Но, как нетрудно видеть, такое равенство не может выполняться при целом значении $2x - 11$.

Чтобы существовал обратный вычет, должны выполняться некоторые условия. Сейчас мы их сформулируем и докажем. В этих рассуждениях удобно использовать порядок на вычетах. Порядок на остатках наследуется из порядка на целых числах. Но теперь $[n-1]_n + [1]_n = [0]_n$, так что правильный порядок на вычетах — циклический. Поэтому вычеты удобно изображать на окружности, как показано на рисунке 4.

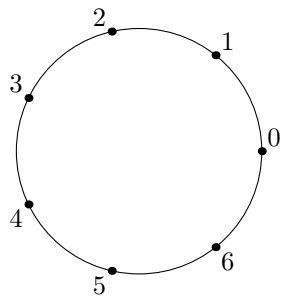


Рис. 4: Вычеты по модулю 7

В общем случае вычетам по модулю n будем сопоставлять вершины правильного n-угольника. (Аналогия с корнями n-й степени из единицы, которую мог заметить

кто-нибудь из читателей, не случайна: позже мы обсудим смысл этой аналогии и придадим ей точную формулировку.)

Будем анализировать обратимость вычета $[a]$, опираясь на эту графическую иллюстрацию.

Сложение вычета $[x]$ с вычетом $[a]$ на рисунке будет означать сдвиг на a против часовой стрелки. Начиная с $[x] = [0]$ и последовательно прибавляя $[a]$, будем получать различные вычеты, кратные вычету $[a]$, см. рис. 5 слева, на котором показано несколько первых шагов этого процесса.

Всего вычетов конечное число. Поэтому рано или поздно они станут повторяться. Первое повторение обязательно случится с нулевым вычетом (с которого мы начали строить множество кратных): $[ka] = [0]$. Это следует из обратимости сложения (то есть из того, что определены противоположные вычеты и операция вычитания). Более формально, пусть k — номер первого повторения, то есть $[ka] = [sa]$, $s < k$. Но тогда $[(k - s)a] = [0]$, то есть повторение уже произошло на шаге $k - s < k$.

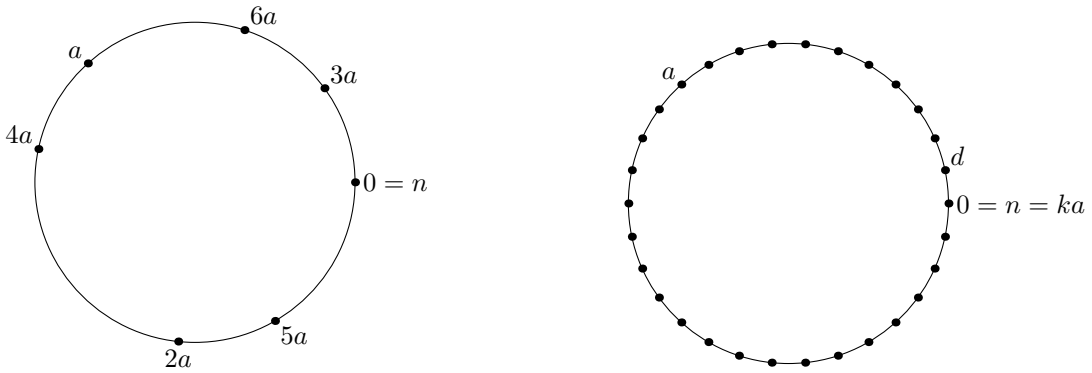


Рис. 5: Вычеты, кратные вычету a

На рисунке 5 справа изображено окончательное множество S_a всех вычетов, кратных вычету $[a]$. Выделим среди этих вычетов тот, который имеет наименьший ненулевой остаток d.

Утверждение 1.45. $d \mid a$.

Доказательство. Проверим, что множество S_a вычетов, кратных вычету $[a]$, замкнуто относительно сложения и вычитания: если $[x] \in S_a$ и $[y] \in S_a$, то $[x \pm y] \in S_a$. Вычеты, кратные $[a]$, — это в точности те вычеты, которые представляются в виде $[qa]$. Но из $[x] = [q_1 a]$, $[y] = [q_2 a]$ следует, что $[x \pm y] = [(q_1 \pm q_2)a] \in S_a$.

Аналогично проверяется, что вычеты, кратные $[a]$, замкнуты относительно умножения: если $[x] \in S_a$, то $[tx] \in S_a$ для любого целого числа t.

Разделим a на d с остатком: $a = sd + r$. Так как $[a] \in S_a$ и $d \in S_a$, то $[r] = [a - sd] \in S_a$ в силу замкнутости относительно умножения и вычитания.

По определению деления с остатком, для остатка r выполняются неравенства $0 \leqslant r < d$. Но мы выбрали d наименьшим ненулевым остатком в множестве S_a.

Поэтому остаётся единственная возможность $r = 0$, что и означает делимость a на d. □

Утверждение 1.46. $d \mid n$.

Доказательство. Аналогично предыдущему. Разделим n на d с остатком: $n = qd + r$. Так как $[n] = [ka] \in S_a$ и $d \in S_a$, то $[r] = [a - sd] \in S_a$ в силу замкнутости относительно умножения и вычитания.

Точно так же, как и в предыдущем рассуждении, убеждаемся, что $r = 0$ в силу выбора d как наименьшего остатка в множестве S_a. □

Утверждение 1.47. *Пусть* $d' \mid a$ *и* $d' \mid n$. *Тогда* $d' \mid d$.

Доказательство. Здесь удобно смотреть на S_a как на множество остатков (целых чисел). Рассмотрим какой-нибудь остаток x из этого множества. Поскольку $[x] = [ta]$ для какого-то кратного a, получаем равенство $ta = qn + x$ в целых числах, которое преобразуется к виду

$$x = ta'd' - qn'd' = (ta' - qn')d',$$

так как d' — общий делитель a и n.

Таким образом, все остатки из множества S_a делятся на d', в том числе и наименьший в этом множестве остаток d. □

Мы доказали в этой серии утверждений, что $d = \text{НОД}(a, n)$ — наибольший общий делитель a и n, а также, что все остатки, кратные $[a]$, делятся на d. После этой подготовки уже несложно доказать критерий обратимости вычета по модулю n.

Теорема 1.48. *Сравнение* $ax \equiv 1 \pmod{n}$ *имеет решение тогда и только тогда, когда* $\text{НОД}(a, n) = 1$.

Доказательство. Пусть $ax \equiv 1 \pmod{n}$ для некоторого x. Тогда $1 \in S_a$ и потому $d = 1$ (нет остатков между 0 и 1).

В обратную сторону. Если $d = \text{НОД}(a, n) = 1$, то $1 \in S_a$. То есть $[1] = [xa]$, что и означает существование решения у сравнения. □

Замечание 1.49 (терминологическое)**.** Мы говорим о равенстве чисел как об утверждении (которое может быть истинно или ложно) и об уравнении как о задаче поиска всех значений переменной, для которых выполняется некоторое равенство.

Для сравнений такого удобного разделения терминов нет: слово «сравнение» означает как утверждение о сравнимости двух вычетов, так и задачу поиска всех вычетов, для которых сравнение выполнено.

Обратите также внимание, что сравнению всегда удовлетворяет конечное множество вычетов (их вообще лишь конечное число). А целых чисел, отвечающих этим вычетам, обычно бесконечно много, так как любой класс вычетов содержит бесконечно много чисел.

Теорема 1.48 говорит о существовании решения сравнения $ax \equiv 1 \pmod{n}$. Нетрудно также убедиться, что если такое сравнение имеет решение, то оно единственное (напомним ещё раз, что решение сравнения — это вычет, то есть бесконечное множество целых чисел).

Следствие 1.50. *Если сравнение $ax \equiv 1 \pmod{n}$ имеет решение, то это решение единственное.*

Доказательство. Разрешимость сравнения означает, что $\text{НОД}(a, n) = 1$ и потому множество кратных вычета $[a]$ состоит из всех вычетов по модулю n. С другой стороны, $[qa] = [(n+q)a]$, то есть существует не более n различных кратных $[a]$. Поскольку каждое кратное представляется в виде $[qa]$, $0 \leqslant q < n$, оно представляется единственным образом. (Если это место непонятно, вернитесь к нему после чтения следующего раздела про отображения множеств.) □

Для полноты рассмотрим более общий случай сравнений вида $ax \equiv b \pmod{n}$. Из предыдущего ясно, что такое сравнение имеет решения тогда и только тогда, когда b кратно $\text{НОД}(a, n)$. Пусть это условие выполняется. Сколько есть решений у данного сравнения? Оказывается, решений может быть много.

Пример 1.51. Сколько решений имеет сравнение $64x \equiv 24 \pmod{360}$?

Так как $\text{НОД}(64, 360) = 8$ и $8 \mid 24$, решения у этого сравнения есть.

Если x_1, x_2 — два разных решения, то $x_1 - x_2$ будет решением сравнения $64x \equiv 0 \pmod{360}$, так как почленное вычитание сравнений $64x_1 \equiv 24 \pmod{360}$ и $64x_2 \equiv 24 \pmod{360}$ даёт сравнение $64(x_1 - x_2) \equiv 0 \pmod{360}$.

Теперь определим количество решений сравнения $64x \equiv 0 \pmod{360}$. В целых числах это сравнение равносильно уравнению $64x = 360y$, которое после сокращения на наибольший общий делитель приобретает вид $8x = 45y$. Так как $\text{НОД}(8, 45) = 1$, то сравнение $8x \equiv 0 \pmod{45}$ имеет единственное решение в вычетах по модулю 45: это доказывается аналогично следствию 1.50.

Итак, из равенства в целых числа $8x = 45y$ следует, что x кратно 45. Всего разных вычетов, кратных 45 по модулю 360 есть $360/45 = 8$ штук. Это и есть количество решений сравнения.

Эти выкладки полезно представить геометрически на картинке с окружностью, по которой расставлены вычеты, см. рис. 6. Давайте идти по этой окружности, сдвигаясь каждый раз на 64. Мы вернёмся в точку 0 после 45 шагов и будем возвращаться в неё каждые 45 шагов. То же самое верно и для остальных точек, которые мы посетим: в каждую мы будем возвращаться через 45 шагов. Поскольку всего шагов (возможных значений вычета, которые могут удовлетворять сравнению) в данном случае 45, делением получаем количество раз, которые мы побываем в каждой точке: $360/45 = 8$, что и даёт количество решений сравнения $64x \equiv b \pmod{360}$ для тех b, для которых решение вообще есть. □

Упражнение 1.52. Проведите те же рассуждения в общем случае и покажите, что если сравнение $ax \equiv b \pmod{n}$ имеет решения, то количество решений равно $\text{НОД}(a, n)$.

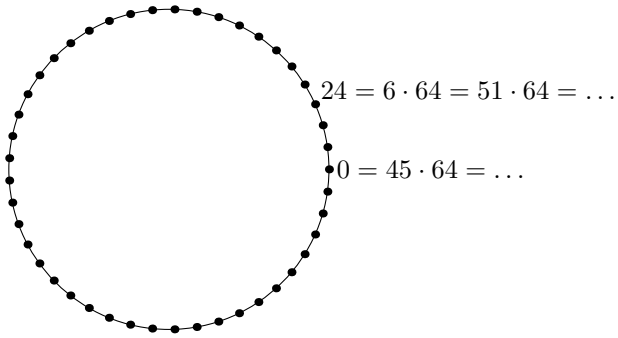


Рис. 6: О подсчёте числа решений сравнения

Позже мы вернёмся к этой задаче, сформулировав её в более общих терминах.

В завершение этого раздела рассмотрим простые числа, которые играют в теории числе и алгебре особую роль.

Определение 1.53. Целое положительное число называется *простым*, если оно больше 1 и делится только на 1 и на само себя.

Числа, которые не являются простыми, называются *составными*.

Замечание 1.54. Это определение выглядит странным: в обычном житейском смысле единица, несомненно, самое простое из положительных целых чисел. Однако дальше станет ясно, почему 1 исключается из простых чисел. Так оказывается гораздо удобнее, если рассматривать обобщения арифметики.

Среди маленьких чисел простых довольно много:

$$2,\ 3,\ 5,\ 7,\ 9,\ 11,\ 13,\ 15,\ 17,\ 19,\ \ldots$$

Нетрудно понять, что далее лакуны между простыми числами могут становиться сколь угодно большими.

Утверждение 1.55. *Для любого L найдётся такое n, что все числа $n+1$, $n+2$, …, $n+L$ составные.*

Доказательство. Возьмём $n = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot L = L!$. Тогда $n + k$ делится на k для любого k от 1 до L. □

Впрочем, простых чисел достаточно много. Очень просто доказывается, что их бесконечно много. (Это одна из самых старых теорем в математике.)

Теорема 1.56. *Простых чисел бесконечно много.*

Доказательство. Нам нужен такой факт: любое целое число > 1 делится на простое. Доказательство индукцией (по всем меньшим числам как в теореме 1.11) по величине числа. База очевидна, а шаг индукции состоит в том, что либо число n простое, либо делится на какое-то меньшее число k. Применяя индуктивное предположение к числу k, получаем простой делитель для n.

Теперь рассмотрим любое конечное множество простых чисел p_1, p_2, …, p_s. Число $p_1 \cdot p_2 \cdot \ldots \cdot p_s + 1$ даёт остаток 1 при делении на p_1, p_2, …, p_s. Значит, его простые делители (а они существуют, как мы показали выше) не принадлежат этому множеству. □

Замечание 1.57. На самом деле простых чисел довольно много. Обозначим через $\pi(x)$ количество простых чисел, не превосходящих x. Тогда *асимптотический закон распределения простых чисел* утверждает, что

$$\pi(x) \sim \frac{x}{\ln x}\,.$$

Асимптотический закон — одна из жемчужин математики. Однако его доказательство выходит за рамки этого курса.

Для полноты изложения ниже мы доказываем более слабые оценки (см. теорему 2.73 на с. 67), которых достаточно для многих приложений, например, для подавляющего большинства приложений в теоретической информатике.

1.4 Отображения, композиции

Отображением множества A в множество B (другое название: *функция* из множества A в множество B) называется произвольное соответствие, которое каждому элементу множества A сопоставляет ровно один элемент множества B.

Замечание 1.58. Если ослабить условие и потребовать лишь, чтобы соответствие сопоставляло не более одного элемента B, приходим к более общему понятию частично определённой функции. Здесь это понятие не используется: мы рассматриваем только всюду определённые функции.

Чаще всего мы будем обозначать отображение как $f\colon A \to B$. Другое часто используемое обозначение $f(x)$ для отображений пришло из анализа. Оно удобно, если известно, на каком множестве определена функция (ниже мы увидим и другие достоинства этого обозначения). Аналогичное обозначение «со стрелочками» выглядит как $x \overset{f}{\mapsto} y$ или $f\colon x \mapsto y$. (Обратите внимание на вертикальную чёрточку на левом конце стрелки — она указывает, что x элемент множества, а не само множество, как при обозначении с обычной стрелкой.)

Отображения $f\colon A \to B$ и $g\colon A \to B$ равны, если для любого $x \in A$ выполняется равенство $f(x) = g(x)$. Это определение становится частным случаем определения равенства множеств, если задать отображения как множества. Напомним, как это делается.

Отображению $f\colon A \to B$ сопоставим *график отображения*

$$\Gamma_f = \{(a, b) : b = f(a)\}.$$

Излагая эту запись словами, получаем, что график — это подмножество декартова произведения $A \times B$, состоящее в точности из пар вида $(a, f(a))$.

Контрольный вопрос 1.59. Проверьте, что отображения равны тогда и только тогда, когда их графики равны как множества.

Для двух отображений $f\colon A \to B$ и $g\colon B \to C$ определена операция *композиции* (обозначение $g \circ f$, порядок существенный). Это такое отображение из множества A в множество C, которое элементу $x \in A$ сопоставляет тот элемент $z \in C$, для которого выполняются равенства $z = g(y)$, $y = f(x)$.

В функциональных обозначениях это записывается даже проще: $(g \circ f)(x) = g\big(f(x)\big)$.

Контрольный вопрос 1.60. Пусть $f\colon \{1,2,3\} \to \{4,5,6\}$ и $g\colon \{1,2,3\} \to \{4,5,6\}$. Определена ли композиция $g \circ f$?

Пример 1.61. Пусть $f\colon \mathbb{N} \to \mathbb{N}$ — отображение, которое сопоставляет натуральному числу n его остаток по модулю 10 (последнюю цифру десятичной записи), а $g\colon \mathbb{R} \to \mathbb{R}$ — функция синуса, $g\colon x \mapsto \sin x$. Формально применить определение композиции к этим отображениям невозможно. Однако ясно, что это техническая проблема — ведь для любого натурального числа определено значение функции синуса, так как натуральные числа являются подмножеством действительных чисел.

Чтобы использовать данное выше определение композиции, нам нужно «промежуточное» отображение вложения подмножества в множество. Если $A \subseteq B$, то определим вложение $\mathrm{id}_{A,B}$ как

$$\mathrm{id}_{A,B}\colon a \mapsto a.$$

После этого применение синуса к последней цифре натурального числа выражается как композиция трёх отображений: f (из натуральных чисел в натуральные), $\mathrm{id}_{\mathbb{N},\mathbb{R}}$ (вложение натуральных в действительные) и g (из действительных в действительные).

Определение вложения выглядит как пустая формальность: элемент переходит сам в себя, но только мы его теперь считаем элементом другого множества. Однако это позволяет не вводить лишних определений и придавать точный смысл композициям таких функций, как в этом примере. □

Важным частным случаем вложения является *тождественное отображение* $\mathrm{id}_A = \mathrm{id}_{A,A}$. Это отображение определено для любого множества и переводит каждый элемент этого множества в себя.

Контрольный вопрос 1.62. Проверьте, пользуясь определениями композиции и тождественного отображения, что для любого отображения $f\colon A \to B$ выполняются равенства $f = f \circ \mathrm{id}_A = \mathrm{id}_B \circ f$.

Утверждение 1.63. *Операция композиции обладает свойством ассоциативности: для любых отображений $f\colon A \to B$, $g\colon B \to C$, $h\colon C \to D$ выполняется равенство*

$$\big((h \circ (g \circ f)\big) = \big((h \circ g) \circ f\big), \quad \textit{то есть } \big((h \circ (g \circ f)\big)(x) = \big((h \circ g) \circ f\big)(x)$$

для любого $x \in A$.

Доказательство. Прямолинейное применение определения композиции. Обозначим $y = f(x)$, $z = g(y)$, $w = h(z)$. Тогда из определения композиции получаем для любого $x \in A$ равенства

$$(g \circ f)(x) = z; \quad \big((h \circ (g \circ f)\big)(x) = w; \quad (h \circ g)(y) = w; \quad \big((h \circ g) \circ f\big)(x) = w,$$

то есть $\big((h \circ (g \circ f)\big)(x) = \big((h \circ g) \circ f\big)(x)$, что и требовалось доказать. □

Определение 1.64. Пусть $f\colon A \to B$ — некоторое отображение.

Отображение $g\colon B \to A$ называется *левым обратным* к f, если $g \circ f = \mathrm{id}_A$.

Отображение $g\colon B \to A$ называется *правым обратным* к f, если $f \circ g = \mathrm{id}_B$.

Отображение $g\colon B \to A$ называется *обратным* к f, если оно одновременно левое и правое обратное.

Определение обратного отображения аналогично определениям противоположного числа (относительно операции сложения) и обратного числа (относительно операции умножения). Несколько вариантов определения возникают из-за того, что композиция не обязательно коммутативна, даже если определены оба отображения $g \circ f$ и $f \circ g$.

Пример 1.65. Пусть $f\colon x \mapsto x + 1$, а $g\colon x \mapsto x^2$ — два отображения на множестве действительных чисел. Тогда

$$(f \circ g)\colon x \mapsto x^2 + 1,$$
$$(g \circ f)\colon x \mapsto (x + 1)^2,$$

то есть $f \circ g \neq g \circ f$. □

Определения обратных отображений неявные и используют операцию композиции. На первый взгляд неясно, для каких отображений существует обратное (левое обратное, правое обратное). Впрочем, ответ на этот вопрос нетрудно получить, внимательно изучив определение композиции. Для формулировки нам потребуются определения, которые будут дальше постоянно использоваться.

Определение 1.66. Отображение $f\colon X \to Y$ называется *инъективным* (инъекция), если из равенства $f(x_1) = f(x_2)$ следует равенство $x_1 = x_2$.

Отображение $f\colon X \to Y$ называется *сюръективным* (сюръекция), если для любого y существует такое x, что $f(x) = y$.

Отображение $f\colon X \to Y$ называется *биективным* (биекция или взаимно однозначное отображение), если для любого y существует ровно одно такое x, что $f(x) = y$ (то есть отображение является одновременно сюръективным и инъективным).

Другими словами, при инъективном отображении образы разных элементов различны; при сюръективном отображении $f\colon X \to Y$ каждый элемент множества Y является образом некоторого элемента множества X.

Наглядно эти определения представлены на схематических рисунках 7–11, на которых множества обозначаются овалами, элементы множеств — точками, а отображения — стрелками, начинающимися в точке и заканчивающимися в её образе.

Заметим, что свойства инъективности и сюръективности (а, значит, и биективности) сохраняются при композициях отображений.

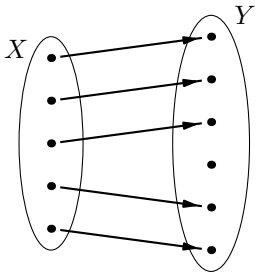


Рис. 7: инъекция

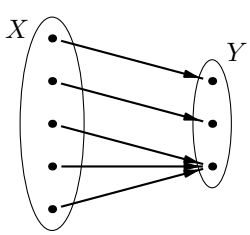


Рис. 8: сюръекция

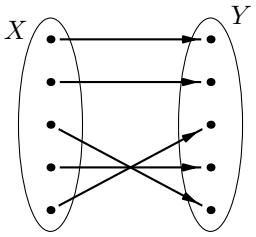


Рис. 9: биекция

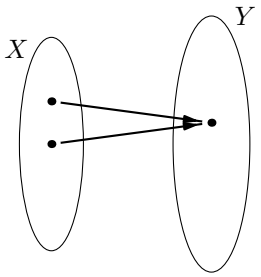


Рис. 10: не инъекция

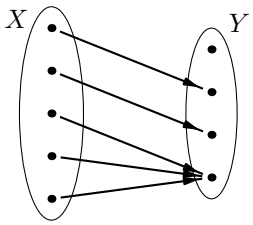


Рис. 11: не сюръекция

Утверждение 1.67. *Если отображения $f\colon Y \to Z$ и $g\colon X \to Y$ инъективны, то их композиция $f \circ g$ также инъективна. Если отображения f, g сюръективны, то их композиция $f \circ g$ также сюръективна.*

Доказательство. Пусть f, g инъективны. В силу инъективности f из равенства $(f \circ g)(x_1) = (f \circ g)(x_2)$ следует равенство $g(x_1) = g(x_2)$, из которого в силу инъективности g следует равенство $x_1 = x_2$. Значит, $f \circ g$ инъективно.

Пусть f, g сюръективны. Тогда для любого $z \in Z$ существует такое $y \in Y$, что $z = f(y)$. Для этого y существует такое $x \in X$, что $y = g(x)$. Но тогда $(f \circ g)(x) = = f(g(x)) = f(y) = z$. Таким образом, $f \circ g$ сюръективно. □

Через введённые свойства инъективности, сюръективности и биективности выражается существование обратных отображений.

Теорема 1.68 (теорема об обратной функции)**.** *У отображения есть левое обратное тогда и только тогда, когда оно инъективное.*

У отображения есть правое обратное тогда и только тогда, когда оно сюръективное.

У отображения есть обратное тогда и только тогда, когда оно биективное.

Доказательство. Третье утверждение теоремы следует из первых двух.

Докажем первое утверждение. Пусть $g\colon Y \to X$ — левое обратное к отображению $f\colon X \to Y$, то есть $g \circ f = \mathrm{id}_X$. Тогда из равенства $f(x_1) = f(x_2)$ следует равенство

$$x_1 = g(f(x_1)) = g(f(x_2)) = x_2,$$

то есть отображение f инъективное.

В обратную сторону: пусть $f\colon X \to Y$ инъективное. Рассмотрим такое отображение $g\colon Y \to X$, что если $y = f(x)$, то $g(y) = x$ (для остальных y значение g произвольное). В силу инъективности такое отображение существует, так как у каждого элемента $y \in Y$ не более одного прообраза. При этом $(g \circ f)(x) = g(f(x)) = x$, то есть такое отображение g является левым обратным.

Докажем второе утверждение. Пусть $g\colon Y \to X$ — правое обратное к отображению $f\colon X \to Y$, то есть $f \circ g = \mathrm{id}_Y$. Другими словами, $y = f(g(y))$ для любого $y \in Y$. Поэтому f сюръективное.

В обратную сторону: пусть $f\colon X \to Y$ сюръективное. Любой $y \in Y$ является образом некоторого элемента x, быть может, не одного. Выберем для каждого y какой-нибудь такой $x = g(y)$, что $y = f(x)$. Мы задали некоторое отображение $g\colon Y \to X$ и оно прямо по построению является правым обратным: $(f \circ g)(y) = f(g(y)) = y$. □

С помощью этой теоремы нетрудно выполнить следующее упражнение.

Упражнение 1.69. Приведите примеры отображений, для которых (а) есть правое обратное, но нет левого обратного; (б) есть левое обратное, но нет правого обратного.

Важный частный случай — отображения конечных множеств в себя. Для таких отображений выполняется более сильное свойство.

Лемма 1.70. *Инъективное отображение конечного множества в себя является сюръективным и потому биективным.*

Аналогично, из сюръективности отображение конечного множества в себя следует инъективность и биективность.

Доказательство. Подсчёт количества элементов в множестве двумя способами.

Пусть $f\colon X \to X$ — инъективное отображение n-элементного множества в себя. Образы всех элементов различны, значит, всего образов ровно n. То есть каждый элемент обязан быть образом какого-то элемента. Поэтому f сюръективно.

Аналогично для сюръективного отображения $f\colon X \to X$ n-элементного множества в себя. У каждого из n элементов есть прообраз. Но всего прообразов не больше n (общего количества элементов в множестве). Поэтому у каждого элемента не более одного прообраза. □

1.5 Перестановки

Перестановкой называется взаимно однозначное отображение конечного множества на себя. Поскольку элементы конечного множества из n элементов можно перенумеровать числами от 1 до n, то часто перестановкой называется взаимно однозначное

отображение множества $\{1, 2, \ldots, n\}$ на себя. В примерах будут возникать и перестановки на других множествах.

Перестановки можно задавать многими различными способами. Для комбинаторики самый естественный и основной способ состоит в том, что числа от 1 до n записываются в некотором порядке. Другими словами, перестановка задаётся последовательностью $a_1, a_2, \ldots, a_n$, в которой каждое число от 1 до n встречается ровно один раз.

Более соответствует нашему определению другой способ. Любую функцию на конечном множестве X можно задать таблицей значений: в первой строке перечислены (в произвольном порядке) элементы множества X, а во второй строке записаны значения функции от того элемента, который стоит в данном столбце.

Пример 1.71. Перестановка, которая задаётся последовательностью $2, 1, 4, 5, 3$ записывается в виде таблицы как

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 2 & 1 & 4 & 5 & 3 \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} 5 & 2 & 4 & 1 & 3 \\ 3 & 1 & 5 & 2 & 4 \end{pmatrix}.$$

(Есть ещё много таблиц, задающих ту же перестановку. Найдите их количество.) □

Табличная запись длиннее, но зато позволяет легко вычислять композицию перестановок. Чтобы вычислить композицию перестановок σ и π, нужно написать таблицу для π, под ней написать такую таблицу для σ, в которой первая строка совпадает со второй строкой таблицы для π (в каждой строке таблицы записаны ровно по одному разу числа от до n, так что такая таблица существует). После этого нужно взять первую строку первой таблицы и вторую строку второй таблицы. Это и будет таблица для композиции перестановок.

Правило звучит громоздко, его легко записать в виде формулы

$$\begin{pmatrix} t_1 & t_2 & t_3 & \ldots & t_i & \ldots & t_n \\ v_1 & v_2 & v_3 & \ldots & v_i & \ldots & v_n \end{pmatrix} \circ \begin{pmatrix} 1 & 2 & 3 & \ldots & i & \ldots & n \\ t_1 & t_2 & t_3 & \ldots & t_i & \ldots & t_n \end{pmatrix} = \\ = \begin{pmatrix} 1 & 2 & 3 & \ldots & i & \ldots & n \\ v_1 & v_2 & v_3 & \ldots & v_i & \ldots & v_n \end{pmatrix}.$$

Справедливость этого правила сразу же получается из определений: если $\pi(i) = t_i$, а $\sigma(t_i) = v_i$, то $(\sigma \circ \pi)(i) = v_i$.

По общей теореме 1.68 для для любой перестановки π есть обратная π^{-1}: такая перестановка, что $\pi^{-1} \circ \pi = \pi \circ \pi^{-1} = \mathrm{id}$. Таблица для обратной перестановки легко получается из таблицы перестановки.

Контрольный вопрос 1.72. Проверьте, что таблица для обратной перестановки получается перестановкой строк таблицы для исходной перестановки.

Для алгебры важен ещё один способ записи перестановок: цикловое разложение. В качестве полезного промежуточного шага построим граф перестановки. Это ориентированный граф на множестве вершин $\{1, 2, \ldots, n\}$, в котором из вершины

i исходит ровно одно ребро в вершину $\pi(i)$. В силу биективности в каждую вершину этого графа также входит ровно одно ребро. Пример графа перестановки из примера 1.71 приведён на рис. 12.

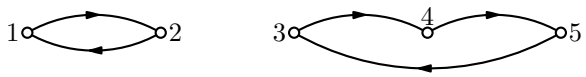


Рис. 12: Граф перестановки

В графе перестановки возможны петли, они возникают в тех вершинах, для которых $\pi(i) = i$.

В любом случае ориентированный граф, входящие и исходящие степени вершин которого равны 1, разбивается на непересекающиеся циклы (петли считаем циклами длины 1). Записывая вершины в порядке обхода этих циклов и разделяя циклы скобками, получаем *цикловое разложение перестановки*. Для перестановки из примера 1.71 цикловое разложение выглядит как

$$(1\ 2)(3\ 4\ 5), \quad \text{или} \quad (2\ 1)(4\ 5\ 3), \quad \text{или} \quad (3\ 4\ 5)(2\ 1).$$

Порядок циклов в записи циклового разложения несущественен. Внутри каждого цикла важен лишь циклический порядок: неважно, какой именно элемент цикла стоит на первом месте.

Контрольный вопрос 1.73. Сколько есть способов записать цикловое разложение перестановки из примера 1.71?

Для краткости в цикловом разложении обычно не указываются циклы длины 1. То есть, вместо записи $(1)(2\ 3)$ используется запись $(2\ 3)$, при этом подразумевается, что не указанные в записи элементы остаются на месте при перестановке. Возникает проблема с тождественно перестановкой: ведь в ней все циклы имеют длину 1. Будем записывать тождественную перестановку как пару скобок $()$.

Упражнение 1.74. Для того, чтобы освоиться с цикловым разложением, вычислите композиции перестановок, заданных цикловыми разложениями, не выписывая таблиц перестановок:

$$(1\ 4\ 5\ 6)(2\ 7) \circ (1\ 2\ 3\ 4\ 5\ 6\ 7\ 8),$$
$$(1\ 3\ 5)(2\ 4\ 6)(7\ 8\ 9) \circ (1\ 2\ 3)(4\ 5\ 6)(7\ 8\ 9),$$
$$(1\ 2) \circ (2\ 3) \circ (3\ 4) \circ (4\ 5) \circ (5\ 1).$$

Для этого удобно строить цикловое разложение композиции, вычисляя образ очередного элемента, пока не произойдёт зацикливание.

В качестве примера покажем, как вычисляется композиция двух циклов

$$\pi = (1\ 4\ 5\ 3\ 2) \circ (1\ 2\ 3\ 4\ 5)$$

Найдём $\pi(1)$. Первая перестановка переводит 1 в 2, вторая переводит 2 в 1. Поэтому $\pi(1) = 1$. Получили цикл длины 1, который не указывается в записи циклового разложения. Далее найдём $\pi(2)$. Аналогично получаем $\pi(2) = 2$.

Поскольку первая перестановка переводит 3 в 4, а вторая — 4 в 5, то $\pi(3) = 5$. Далее вычисляем $\pi(5)$, это 4. Аналогично проверяем, что $\pi(4) = 3$. Получили цикл (3 5 4) длины 3, так что $\pi = (3\ 5\ 4)$.

Пример 1.75. На множестве остатков остатков по модулю 13 рассмотрим отображение $\pi\colon x \mapsto 8x \bmod 13$. Это перестановка, так как для каждого ненулевого остатка однозначно определён обратный по модулю 13 (так как 13 простое число, все меньшие положительные целые числа с ним взаимно просты, поэтому применима теорема 1.48). Ясно также, что $\pi(0) = 0$.

Найдём $\pi^4 = \pi \circ \pi \circ \pi \circ \pi$. Подставляя в определение π, получаем

$$\pi^4(x) \equiv 8\pi^3(x) \equiv \cdots \equiv 8^4 x \bmod 13.$$

Но $8^2 = 64 \equiv -1 \bmod 13$, так что $8^4 \equiv 1 \bmod 13$. Поэтому $\pi^4 = \mathrm{id}$.

Найдём цикловое разложение π. Нулевой остаток образует цикл длины 1. Найдём цикл, в который входит 1. Вычисления по модулю 13 показывают, что

$$1 \overset{\pi}{\mapsto} 8 \overset{\pi}{\mapsto} 64 \equiv -1 \equiv 12 \overset{\pi}{\mapsto} -8 \equiv 5 \overset{\pi}{\mapsto} 1,$$

то есть в π есть цикл (1 8 12 5). Аналогично находим цикл, содержащий 2:

$$2 \overset{\pi}{\mapsto} 16 \equiv 3 \overset{\pi}{\mapsto} 24 \equiv -2 \equiv 11 \overset{\pi}{\mapsto} -16 \equiv -3 \equiv 10 \overset{\pi}{\mapsto} 80 \equiv 2,$$

то есть (2 3 11 10).

Число 4 пока не вошло ни в один из циклов, построим для него цикл точно такими же вычислениями:

$$4 \overset{\pi}{\mapsto} 32 \equiv 6 \overset{\pi}{\mapsto} 48 \equiv -4 \equiv 9 \overset{\pi}{\mapsto} -32 \equiv -6 \equiv 7 \overset{\pi}{\mapsto} 56 \equiv 4,$$

то есть (4 6 9 7).

Собирая циклы вместе, получаем ответ: $\pi = (1\ 8\ 12\ 5)(2\ 3\ 11\ 10)(4\ 6\ 9\ 7)$. Циклы получились одинаковой длины (не считая цикла длины 1, состоящего из нулевого остатка). Тому есть внятное объяснение на языке алгебры, аналогичное объяснению для примера 1.51. □

Перестановка, которая меняет местами два элемента, а остальные оставляет на месте, называется *транспозицией*. Цикловое разложение транспозиции имеет вид $(i\ j)$. Нетрудно понять, что любая перестановка является композицией транспозиций: переставляя числа попарно, можно расположить их в любом заданном порядке. Приведём, однако, доказательство этого факта, поскольку он будет нужен в дальнейшем.

Теорема 1.76. *Любая перестановка представляется как композиция транспозиций.*

Доказательство. Цикловое разложение показывает, что любая перестановка является композицией циклов (здесь под циклом мы понимаем перестановку, которая какое-то подмножество элементов переставляет циклически, а остальные элементы оставляет на месте).

Поэтому достаточно представить в виде композиции транспозиций один цикл. Сделаем это для цикла $(1\ 2\ \ldots\ k)$. Докажем, что

$$(1\ 2\ \ldots\ k) = (k\ 1) \circ ((k-1)\ 1) \circ \cdots \circ (3\ 1) \circ (2\ 1). \tag{1.22}$$

Композиции перестановок вычисляются справа налево, так же, как вычисляются функции в подстановке $f(g(h(x)))$. Самая правая транспозиция переводит 1 в 2, а все остальные оставляют 2 на месте. Эта же транспозиция переводит 2 в 1, следующая — 1 в 3, остальные оставляют 3 на месте. Далее аналогично: первые (справа) $i-2$ транспозиций оставляют i на месте; $(i-1)$-я справа переводит i в 1, а i-я справа — 1 в $i+1$. Это рассуждение справедливо для $i<k$. Для k нет k-й справа транспозиции (их всего $k-1$), поэтому k переходит в 1.

Для произвольного цикла $(i_1\ i_2\ \dots\ i_k)$ точно такое же рассуждение доказывает, что

$$(i_1\ i_2\ \dots\ i_k) = (i_k\ i_1)\circ(i_{k-1}\ i_1)\circ\dots\circ(i_3\ i_1)\circ(i_2\ i_1),$$

нужно только говорить об индексах, а не о самих числах. (Эта симметрия перестановок будет дальше изучена более подробно.) □

Конструкция из доказательства теоремы 1.76 использует все транспозиции. Это необязательно, можно обойтись меньшим количеством. Чтобы анализировать возможность представления перестановок композициями, удобно изменить формулировку. Заметим, что транспозиция обратна самой себе:

$$(i\ j)\circ(i\ j) = (). \tag{1.23}$$

Поэтому для представления перестановки π как композиции (каких-то) транспозиций можно искать представление тождественной перестановки в виде композиции π и какого-то количества транспозиций. Если

$$\mathrm{id} = \tau_1\circ\tau_2\circ\dots\circ\tau_N\circ\pi,$$

где τ_i — транспозиции, то

$$\begin{aligned}
\tau_1 &= \tau_1\circ(\tau_1\circ\tau_2\circ\dots\circ\tau_N\circ\pi) = \tau_2\circ\dots\circ\tau_N\circ\pi,\\
\tau_2\circ\tau_1 &= \tau_2\circ(\tau_2\circ\dots\circ\tau_N\circ\pi) = \tau_3\circ\dots\circ\tau_N\circ\pi,\\
&\dots\\
\tau_N\circ\tau_{N-1}\circ\dots\circ\tau_2\circ\tau_1 &= \pi
\end{aligned}$$

(здесь мы использовали ассоциативность композиции отображений и (1.23)).

Такое изменение формулировки позволяет строить доказательства по индукции: домножим на транспозиции так, чтобы один из элементов оставался на месте, и применим индуктивное предположение.

Лемма 1.77. *Любая перестановка чисел $\{1,2,\dots,n\}$ представляется как композиция транспозиций соседних чисел, то есть транспозиций вида* (1 2), (2 3), …, $((n-1)\ n)$.

Доказательство. Индукция по n. База индукции $n=2$ очевидна: есть только две перестановки на двух элементах, одна из которых — транспозиция (1 2), а вторая — тождественная перестановка, которая равна (1 2) ∘ (1 2).

Шаг индукции: предполагаем, что лемма верна для перестановок $n-1$ чисел. Докажем, что тогда она верна и для перестановок n чисел.

Рассмотрим какую-нибудь перестановку π. Обозначим $i = \pi(n)$.

Композиция

$$\tau_i = (n\ (n-1)) \circ ((n-1)\ (n-2)) \circ \cdots \circ ((i+1)\ i)!$$

переводит i в n, как легко видеть. Поэтому $(\tau_i \circ \pi)(n) = \tau_i(i) = n$. Отбросим n и применим индуктивное предположение к полученной перестановке $n-1$ чисел. $\square$

Рассуждая аналогично, нетрудно решить вопрос о представлении всех перестановок композициями транспозиций из заданного семейства. Приведём формулировку, оставляя доказательство в качестве хорошего упражнения для читателя.

Задача 1.78. По множеству транспозиций T построим неориентированный граф $G(T)$, вершины которого — это числа от 1 до n, а рёбра задаются транспозициями из T: вершины i, j соединены ребром, если $(i\ j) \in T$.

Любая перестановка представляется композициями транспозиций из множества T в том и только том случае, когда граф $G(T)$ связный.

1.6 Движения пространства и симметрии многогранников

1.6.1 Движения

Под пространством в этом разделе мы понимаем d-мерное евклидово пространство. Случай $d = 2$ отвечает обычной евклидовой плоскости, а $d = 3$ — обычному евклидову трёхмерному пространству, в котором мы живём.

Мы напомним основные определения и свойства движений. Более подробное изложение можно найти в любом учебнике геометрии и линейной алгебры. (См. также ниже раздел 8.)

Движения пространства — это преобразования, сохраняющие расстояние между точками: если T — движение, то $d(Tx, Ty) = d(x, y)$, где $d(\cdot, \cdot)$ обозначает расстояние между точками пространства.

Легко понять, что движение пространства инъективно, так как расстояние между различными точками положительное.

Сложнее доказать, что движение сюръективно. Например, можно ввести в пространстве декартову систему координат и доказать следующую лемму.

Лемма 1.79. *Движение пространства в любой декартовой системе координат задаётся линейной неоднородной формулой $T\colon x \mapsto Ax + b$, где A — $d \times d$ матрица, а b — вектор размерности d.*

Из условия сохранения расстояний получается условие на матрицу A:

$$A^T A = I, \tag{1.24}$$

здесь A^T — транспонированная матрица. Удовлетворяющие условию (1.24) матрицы называются *ортогональными*. Ортогональные матрицы невырождены: обратная к ортогональной, как видно из условия, совпадает с транспонированной. Обратное отображение задаётся формулой $T^{-1}\colon x \mapsto A^T x - A^T b$.

Упражнение 1.80. Докажите, что движение d-мерного пространства задаётся образами любого множества из $d+1$ точки, которое не лежит в пространстве меньшей размерности.

Итак, движения биективны. Поэтому композиция движений — также движение.

Из линейной алгебры известно, что определитель невырожденной матрицы не равен нулю. Более того, из условия ортогональности (1.24) легко вывести, что определитель матрицы A, задающей движение, по модулю равен 1 (его квадрат равен определителю единичной матрицы, то есть 1). Если $\det A = +1$, то движение называют *собственным*, в противном случае его называют *несобственным*.

Важный пример несобственного движения: отражение относительно гиперплоскости. При отражении относительно гиперплоскости H точка x переходит в точку x', лежащую на прямой, перпендикулярной H, на том же расстоянии от H, что и x, но на другом луче этой прямой с вершиной в точке пересечения перпендикуляра и гиперплоскости.

В подходящей системе координат отражение относительно гиперплоскости задаётся очень простыми формулами: $x_1 \mapsto -x_1$, $x_i \mapsto x_i$, $i > 1$.

Контрольный вопрос 1.81. Отражение относительно какой гиперплоскости задают эти формулы?

Далее мы будем в основном рассматривать только 2-мерный и 3-мерный случаи. Более того, нас будут интересовать только движения с неподвижной точкой (в подходящей системе координат они задаются однородной линейной формулой $x \mapsto Ax$).

Более точно, нас будут интересовать *симметрии* фигур и тел. По определению, движение T является симметрией множества X, если $T(X) = X$. Нам потребуются следующие факты о симметриях фигур и тел.

Теорема 1.82. *Если X — ограниченное множество, то любая его симметрия имеет неподвижную точку.*

Теорема 1.83. *Все собственные движения плоскости, имеющие неподвижную точку, — это повороты относительно точки.*

Все несобственные движения плоскости, имеющие неподвижную точку, — это отражения относительно прямой.

Теорема 1.84. *Все собственные движения 3-мерного пространства, имеющие неподвижную точку, — это повороты относительно прямой.*

1.6.2 Правильные многоугольники и их симметрии

Правильный многоугольник — это такой выпуклый многоугольник, у которого длины всех сторон равны и все углы равны. На рис. 13 изображены правильные 3-угольник (треугольник), 4-угольник (квадрат), 12-угольник.

Каковы симметрии правильных многоугольников? У правильного многоугольника есть *центр*. Это такая точка, которая находится на равных расстояниях от всех его вершин.

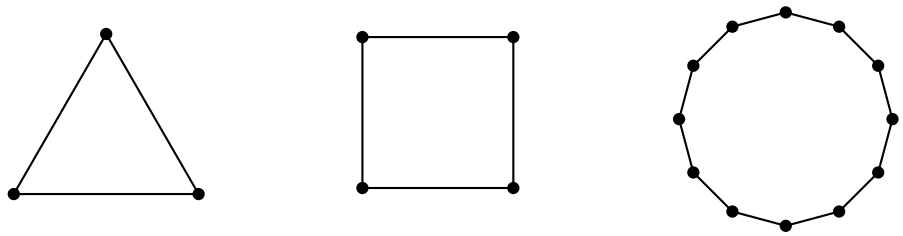


Рис. 13: Правильные многоугольники

Повороты на углы $2\pi/n$ относительно центра являются собственными симметриями правильного n-угольника. Таких симметрий n штук.

Отражения относительно прямых, проходящих через центр и одну из вершин правильного n-угольника являются его несобственными симметриями при нечётном n. В этом случае вторая точка пересечения оси симметрии с границей n-угольника — это середина стороны, противоположной вершине, через которую проходит ось симметрии. Таких симметрий также n штук.

При чётном n несобственные симметрии бывают двух типов: ось симметрии проходит либо через пару противоположных вершин, либо через середины противоположных сторон. Симметрий каждого типа $n/2$, поэтому общее количество несобственных симметрий по-прежнему равно n.

Других симметрий у правильных многоугольников нет. Чтобы в этом убедиться, заметим, что симметрия правильного n-угольника однозначно задаётся тем, куда переходят вершины одной из сторон.

Здесь мы опираемся на очень важный факт: симметрия многоугольника переводит вершины в вершины, а стороны в стороны.

Упражнение 1.85. Докажите это утверждение.

Отсюда получаем следствие: центр правильного n-угольника остаётся на месте при любой симметрии (это единственная точка плоскости, которая находится на одинаковых расстояниях от всех его вершин).

Второе важное наблюдение: композиция симметрий является симметрией. Более того, композиция поворотов является поворотом, композиция отражений является поворотом на удвоенный угол между осями симметрии, а композиция поворота и отражения является отражением.

Упражнение 1.86. Докажите эти утверждения. Опишите ось симметрии в случае композиции поворота и отражения.

Далее, найдём симметрии, которые оставляют вершину a правильного n-угольника на месте. Это нетрудно сделать. Рассуждение иллюстрируется рис. 14.

Так как вершина a и центр O остаются на месте, то и все точки прямой ℓ, проходящей через a и O, остаются на месте. Но тогда и любая прямая, перпендикулярная ℓ,

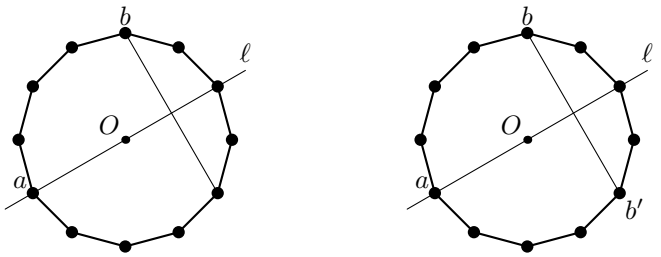


Рис. 14: Симметрии, сохраняющие вершину a

переходит в себя, причём точка пересечения перпендикуляра с ℓ остаётся на месте (все точки ℓ остаются на месте).

Легко проверить, что движение прямой, имеющее неподвижную точку, либо тождественное, либо является отражением относительно этой неподвижной точки (проверьте!). Поэтому получаем два случая.

I. Если все точки какого-нибудь перпендикуляра к ℓ остаются на месте (рис. 14 слева), то такая симметрия тождественная (три точки a, O, b, не лежащие на одной прямой, остаются на месте).

II. Если точки перпендикуляра отражаются относительно точки пересечения с прямой ℓ (рис. 14 слева), то такая симметрия является отражением относительно прямой ℓ (образы точек a, O, b такие же, как при отражении относительно ℓ).

Рассмотрим теперь какую-нибудь симметрию T, которая переводит вершину a в вершину a' (рис. 15).

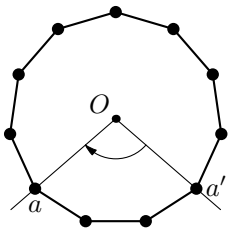


Рис. 15: Симметрия 11-угольника

Существует поворот R, который переводит вершину a' в a. Тогда композиция $S = R \circ T$ оставляет вершину a на месте. Применяя к этому равенству обратный поворот R^{-1}, получаем, что $R^{-1} \circ S = R^{-1} \circ R \circ T = T$. Таким образом, T является либо поворотом, либо композицией поворота и отражения, а такая композиция является отражением.

Итак, мы убедились, что нет симметрий правильных многоугольников, отличающихся от перечисленных выше.

Замечание 1.87. Если поместить правильный многоугольник в 3-мерное простран-

ство, то все найденные симметрии реализуются поворотами относительно прямых, проходящих через его центр. Поворотные симметрии на плоскости отвечают поворотам вокруг прямой, перпендикулярной плоскости многоугольника. Отражения в плоскости в пространстве реализуются как повороты вокруг прямой (оси отражения) на 180°. На рис. 16 показаны соответствующие прямые для правильных 6- и 5-угольника.

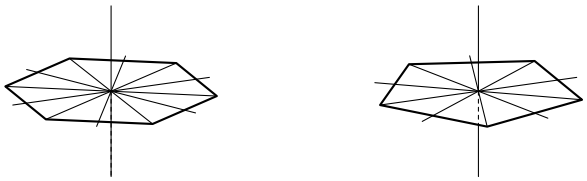


Рис. 16: Симметрии многоугольников как повороты в пространстве

1.6.3 Правильные многогранники и их симметрии

Как хорошо известно, в 3-мерном пространстве есть 5 разных (с точностью до подобия) видов правильных многогранников (они ещё называются платоновыми телами, в честь философа Платона, который их обожал). Они приведены на рисунке 17.

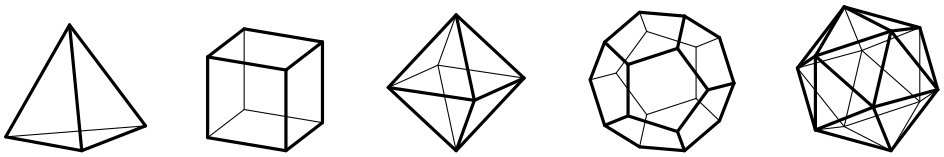


Рис. 17: Правильные многогранники. Слева направо: тетраэдр, куб, октаэдр, додекаэдр, икосаэдр

Для правильных многогранников есть много определений. Например, правильный многогранник — это такой выпуклый многогранник, грани которого — равные правильные многоугольники, и в каждой вершине сходится одинаковое количество рёбер.

Нас интересуют симметрии многогранников, поэтому более подходящим является определение, которое сразу указывает на свойства симметрии. Правильный многогранник — это такой выпуклый многогранник, каждый из любого его флага можно получить любой другой симметрией этого многогранника. Флагом мы называем вершину, исходящее из неё ребро, и грань, которая содержит эти вершину и ребро.

На рис. 18 изображены два флага на додекаэдре. По второму определению должна найтись симметрия додекаэдра, которая переводит один флаг в другой. Это

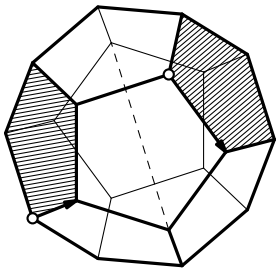


Рис. 18: Два флага на додекаэдре

поворот на 120° относительно прямой, отмеченной на рисунке штриховой линией. Эта прямая проходит через две противоположные вершины додекаэдра.

Мы не приводим доказательство равносильности этих определений и не обсуждаем подробности классификации правильных многогранников.

Сделаем лишь несколько замечаний о построении правильных многогранников. Построение куба не вызывает обычно трудностей. Октаэдр получается из куба так: его вершины — это центры граней куба. И наоборот, центрами граней октаэдра являются вершины куба. Такое соответствие называется двойственностью. Додекаэдр и икосаэдр также двойственны.

Правильный тетраэдр также легко построить или увидеть его на кубе, см. рис. 19 слева. Додекаэдр можно получить надстройкой над кубом, как показано на рис. 19 в центре. Икосаэдр можно получить, выбрав центры граней додекаэдра. Альтернативный способ показан на рис. 19: икосаэдр можно вписать в октаэдр.

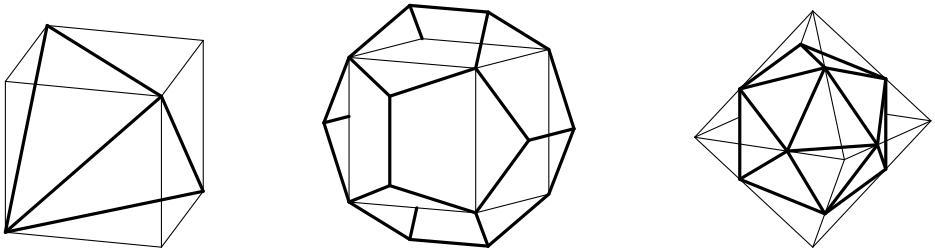


Рис. 19: Подсказки к построению правильных многогранников

Замечание 1.88. Существует ровно пять кубов, вписанных в додекаэдр как на рис. 19 в центре. Выбирая одну из пяти диагоналей в какой-нибудь грани, можно построить куб аналогично изображённому на рисунке. Других кубов с вершинами в вершинах додекаэдра нет. Мы не доказываем это утверждение, но используем его ниже при изучении алгебраической структуры собственных симметрий додекаэдра.

2 Группы и подгруппы

В этой главе вводится одна из самых важных алгебраических структур — группы. Фактически, в предыдущей главе мы уже рассматривали различные примеры групп, не вводя самого понятия.

2.1 Определение группы, примеры групп

Определение 2.1. *Бинарная алгебраическая операция* на множестве M — это отображение $*\colon M \times M \to M$.

Другими словами, бинарная операция однозначно сопоставляет упорядоченной паре элементов M (операндам) результат операции. Результат операции по традиции записывается в инфиксной записи: запись $c = a * b$ означает, что паре элементов a и b сопоставлен результат операции c. Зачастую результат операции называется *произведением* элементов a и b и даже обозначается как $a \cdot b$, хотя сами элементы могут быть не числовой природы или операция на числах может отличатся от обычного умножения. Скажем, сложение чисел — это тоже бинарная операция на соответствующем числовом множестве. Обычно сложение обозначается знаком $+$. Поэтому иногда применяется альтернативный способ записи операции с помощью знака $+$. В этом случае результат операции называется по понятным причинам суммой.

По традиции при изучении одной бинарной операции на множестве инфиксная запись сокращается. Если ясно из контекста, о какой операции идёт речь, то произведение записывается как ab, аналогично обычному произведению чисел.

Определение 2.2. Группа $G = \langle M, * \rangle$ — это такая пара из множества M и бинарной операции $*$ на этом множестве, что выполняются следующие свойства (аксиомы группы):

G1: $(x * y) * z = x * (y * z)$ (ассоциативность);

G2: (аксиома единицы) существует единственный *нейтральный* (или единичный) элемент e такой, что для любого x выполняется $e * x = x * e = x$;

G3: для любого элемента x существует ровно один *обратный элемент*, то есть такой элемент y, для которого $y * x = x * y = e$ (обратный элемент обозначается x^{-1}).

Таким образом, чтобы задать группу, нужно указать множество и такую операцию на этом множестве, для которой выполняются указанные выше аксиомы группы.

Пример 2.3 (Числовые группы)**.** Числовые системы, рассмотренные в разделе 1.2: целые числа $\mathbb{Z}$, рациональные числа $\mathbb{Q}$, действительные числа $\mathbb{R}$, комплексные числа $\mathbb{C}$, образуют группы относительно операции сложения.

Множества отличных от нуля чисел (рациональных $\mathbb{Q}^*$, действительных $\mathbb{R}^*$, комплексных $\mathbb{C}^*$) образуют группу относительно операции умножения.

Групповые аксиомы во всех этих случаях означают хорошо известные свойства числовых систем, которые также указаны в разделе 1.2. □

Все числовые группы удовлетворяют дополнительному свойству коммутативности: $a * b = b * a$ для любых a, b из группы. Группы со свойством коммутативности называются *абелевыми* или коммутативными.

При использовании мультипликативной записи $x \cdot y$ или xy нейтральный элемент группы традиционно называется единицей и обозначается e или 1. При аддитивной записи единичный элемент называется нулём и обозначается 0. Вместо термина «обратный» при аддитивной записи используется термин «противоположный» как и для числовых систем. Противоположный к элементу y обозначается $-y$. Аддитивная запись обычно (но далеко не всегда) используется для обозначения коммутативных операций.

Группа называется *конечной*, если в ней (а точнее — в множестве M) конечное число элементов. Это число называется *порядком* группы.

Пример 2.4. Перестановки множества $\{1, 2, \ldots, n\}$ с операцией композиции перестановок образуют группу. Мы проверили все групповые аксиомы в разделах 1.4 и 1.5. Обозначается группа перестановок n элементов как S_n. Другое её название — *симметрическая группа*.

Перестановки дают пример некоммутативных групп. Действительно, легко проверить, что

$$(1\ 2) \circ (2 \circ 3) = (1\ 2\ 3) \neq (1\ 3\ 2) = (2 \circ 3) \circ (1\ 2)$$

(здесь, как и в большинстве дальнейших вычислений, мы используем цикловую запись перестановок).

Перестановки образуют конечную группу. Её порядок, то есть количество перестановок, нетрудно найти обычными методами элементарной перечислительной комбинаторики. Он равен $n! = 1 \cdot 2 \cdot \ldots \cdot n$. □

Для элементарной теории чисел интересны группы вычетов по модулю n.

Пример 2.5 (аддитивная группа вычетов)**.** Вычеты по модулю n с операцией сложения образуют группу, мы это проверили в разделе 1.3. Обозначается эта группа Z_n. Это конечная абелева группа. Её порядок, то есть количество вычетов, равен n. □

Пример 2.6 (мультипликативная группа вычетов)**.** Взаимно простые с n вычеты по модулю n (то есть отвечающие взаимно простым с n остаткам), с операцией умножения образуют группу. Ассоциативность и коммутативность умножения вычетов доказана в разделе 1.3.

Нейтральным элементом является вычет [1]: $[1]_n \cdot [x]_n = [1 \cdot x]_n = [x]_n$ для любого вычета x, даже необязательно взаимно простого с n.

Существование обратного вычета гарантирует теорема 1.48, доказанная в разделе 1.3.

Мультипликативная группа вычетов по модулю n обозначается Z_n^*. Её порядок, то есть количество остатков по модулю n, которые взаимно просты с n, обозначается $\varphi(n)$ и называется *функцией Эйлера.* □

Контрольный вопрос 2.7. Проверьте, что $\varphi(2) = 1$, $\varphi(3) = 2$, $\varphi(4) = 2$.

Все вычеты по модулю n с операцией умножения не образуют группу при $n > 1$. Действительно, $[0] \cdot [x] = [0]$, поэтому у вычета $[0]$ нет обратного, если есть хотя бы два разных вычета.

Пример 2.8 (матричные группы)**.** Матрицы с ненулевым определителем образуют группу относительно операции матричного умножения. Эта группа обозначается $GL(n)$ (n — размер матриц). Она имеет прямое отношение к группам преобразований, поскольку матрицами размера $n \times n$ записываются линейные преобразования n-мерного пространства, а матричное умножение соответствует композиции линейных преобразований.

Группа обратимых матриц неабелева, как видно из примера

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \neq \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

Выделением матриц специального вида получаются другие примеры матричных групп. В частности, если ограничиться ортогональными матрицами, удовлетворяющими условию $X^T = X^{-1}$ (транспонированная матрица совпадает с обратной), то получим группу $O(n)$, которая соответствует движениям n-мерного пространства, оставляющим на месте начало координат (*ортогональная группа*).

Мы ничего не сказали об элементах матриц. Это ещё одна степень свободы при выборе матричных групп. От элементов матриц требуется немногое — их нужно складывать и умножать по обычным законам арифметики. Например, можно рассматривать группы матриц с рациональными, действительными или комплексными коэффициентами (и это разные группы!).

Позже мы рассмотрим другие примеры алгебраических систем, допускающих такие операции. Пока будем считать (если не оговорено противное), что элементы матриц — действительные числа. □

Пример 2.9 (группы движений)**.** Движения пространства также образуют группу, как и симметрии многоугольников и многогранников.

Обычно группа симметрий правильного n-угольника (как собственных, так и несобственных) обозначается D_n и называется *группой диэдра.* Как мы проверили выше в разделе 1.6.2, в этой группе $2n$ элементов.

Эта группа также неабелева. Если обозначить через r поворот вокруг центра многоугольника против часовой стрелки на угол $\varphi = 2\pi/n$, а через p — отражение относительно прямой ℓ (любое из n возможных), то

$$p' = p \circ r \neq r \circ p = p'',$$

где p' — отражение относительно прямой ℓ', полученной из ℓ поворотом на угол π/n по часовой стрелке, а p'' — отражение относительно прямой ℓ'', полученной из ℓ

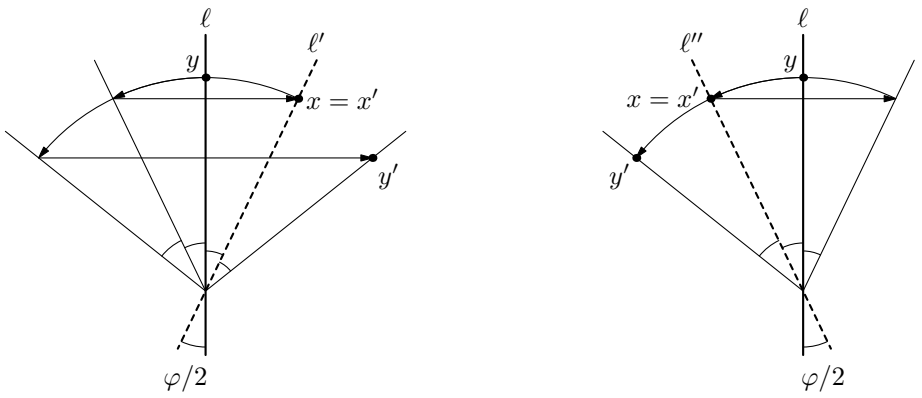


Рис. 20: Неабелевость группы диэдра

поворотом на угол π/n против часовой стрелки. Это вычисление иллюстрируется рис. 20 (слева вычисление $p \circ r$, справа — $r \circ p$).

Контрольный вопрос 2.10. Поворачивая на угол π/n, нужно сделать $2n$ шагов, чтобы вернуться в исходную точку. Почему осей симметрии у правильного n-угольника в два раза меньше?

Группы собственных симметрий (поворотов вокруг прямой) правильных многогранников так и называются: группа тетраэдра, группа куба и т.д. Как мы увидим далее, по сути есть три различных группы симметрий правильных многогранников: группа тетраэдра, группа куба и группа додекаэдра.

Заметим, что D_n можно рассматривать как группу собственных симметрий правильного многоугольника в пространстве. На правильный многоугольник в пространстве можно смотреть как на вырожденный многогранник с двумя гранями. Отсюда и название — группа диэдра (двугранника по-русски). □

Позже у нас появится много других примеров групп.

Про любое множество с заданной на нём бинарной операцией можно поставить вопрос, является ли это множество с данной операцией группой. Ответ получается проверкой аксиом группы. Приведём примеры, когда нарушается ассоциативность.

Пример 2.11. Множество целых чисел с операцией вычитания. Ассоциативность нарушается, так как

$$1 - (2 - 3) = 2 \neq -4 = (1 - 2) - 3.$$

Что касается аксиомы единицы, то ситуация более сложная. Так как $a - 0 = a$ для любого целого числа, 0 является *правой единицей* для операции вычитания. Но, конечно же, 0 не является *левой единицей*: в общем случае $0 - a = -a \neq a$. □

Пример 2.12. Приведём интересный пример множества с операцией, в котором выполняются аксиомы единицы и существования обратного, но не выполняется свойство ассоциативности.

Определим на множестве целых чисел операцию $*$:

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

Легко видеть, что $x * y = y * x$, то есть эта операция коммутативная. Нейтральным элементом относительно этой операции будет 0:

$$0 * x = (1 + 0 \cdot x)(0 + x) = x.$$

Противоположным элементом будет $-x$:

$$(-x) * x = (1 + (-x) \cdot x)(-x + x) = 0.$$

Свойство ассоциативности не выполняется, что легко следует из прямого вычисления

$$1 * 2 = (1 + 2)(1 + 2) = 9, \qquad 2 * 3 = (1 + 6)(2 + 3) = 35,$$
$$9 * 3 = (1 + 27)(9 + 3) = 7 \cdot 48, \quad 1 * 35 = (1 + 35)(1 + 35) = 36^2,$$

показывающего, что $(1 * 2) * 3 = 7 \cdot 48 \neq 36 \cdot 36 = 1 * (2 * 3)$. □

2.2 Общие свойства групп и бинарных операций

Ассоциативность операции означает независимость результата применения двух операций от порядка, в котором они выполняются: $x(yz) = (xy)z$. (Используем сокращённую мультипликативную запись операции, то есть опускаем знак операции между сомножителями).

Из ассоциативности следует аналогичное свойство для сколь угодно длинных выражений. Сформулируем это свойство более точно. Рассмотрим множество A с бинарной операцией. Если дана последовательность $a_1, a_2, \ldots, a_n, a_{n+1}$ элементов этого множества, то n-кратным применением операции из неё можно получить один элемент. Результат в общем случае зависит от порядка применения операций к элементам последовательности. (См. пример 2.12.) Чтобы выделить одну из возможных последовательностей применения операции, нужно указать порядок применения операций. Это принято делать расстановкой скобок, как в левой и правой частях равенства ассоциативности. В общем случае расстановок скобок очень много. Количество способов расставить скобки при выполнении n операций равно *числу Каталана* C_n. Это известная последовательность чисел, дающая ответ к очень многим комбинаторным задачам (известны сотни эквивалентных определений чисел Каталана). Числа Каталана выражаются через биномиальные коэффициенты

$$C_n = \frac{1}{n+1}\binom{2n}{n},$$

асимптотически они растут экспоненциально быстро.

Мы сейчас покажем, что всё это разнообразие возможных ответов вычислений вырождается в случае ассоциативной операции до одного. Приведём для удобства формулировки стандартное определение.

Определение 2.13. Множество M с ассоциативной операцией называется *полугруппой*.

Лемма 2.14. *Результат применения n операций к последовательности элементов полугруппы не зависит от расстановки скобок.*

Доказательство. Будем доказывать индукцией по числу сомножителей n, что при любой расстановке скобок произведение элементов $a_1, \dots, a_n$ равно $a_1(a_2(\dots a_n)\dots)$. База индукции $n = 3$ — это в точности свойство ассоциативности $(a_1a_2)a_3 = a_1(a_2a_3)$.

Пусть утверждение доказано для произведений n элементов. Рассмотрим произведение $(n+1)$-го элемента. Оно имеет вид $L \cdot R$, где в L и R элементов не больше n. Поэтому по предположению индукции $L = a_1 \cdot L'$, а $L \cdot R = (a_1 \cdot L') \cdot R = a_1 \cdot (L' \cdot R)$. В $L' \cdot R$ всего n элементов, так что можно ещё раз применить предположение индукции и вывести справедливость доказываемого утверждения для произведений $(n+1)$-го элемента. □

Следующее свойство также удобно применять в вычислениях.

Лемма 2.15. $(xy)^{-1} = y^{-1}x^{-1}$.

Доказательство. Прямое вычисление, в котором множители группируются по-разному и применяются групповые аксиомы (указаны над знаками равенства):

$$(y^{-1}x^{-1})(xy) \overset{\text{G1}}{=} y^{-1}\big(x^{-1}(xy)\big) \overset{\text{G1}}{=} y^{-1}\big((x^{-1}x)y\big) \overset{\text{G3}}{=} y^{-1}(ey) \overset{\text{G2}}{=} y^{-1}y \overset{\text{G2}}{=} e. \qquad \square$$

Приведём пример использования этого свойства.

Пример 2.16. Докажем, что в любой группе $(xy)^{-1}x^2(yx)^{-1}y^2 = e$ для любых элементов группы x, y. Используя лемму 2.15 и ассоциативность умножения, получаем

$$(xy)^{-1}x^2(yx)^{-1}x^2 = (y^{-1}x^{-1})x^2(x^{-1}y^{-}1)y^2 = y^{-1}(x^{-1})x^2(x^{-1})y^{-1}y^2 = y^{-2}y^2 = e. \square$$

Ещё один пример применения этого свойства возникает в доказательстве следующей леммы.

Лемма 2.17. *Пусть для любого элемента x группы G выполняется равенство $x^2 = e$. Тогда группа G абелева.*

Доказательство. Из равенства $x^2 = e$ следует (после умножения обеих частей на x^{-1}), что $x^{-1} = x$ для любого элемента группы. Применим это равенства к равенству из леммы 2.15: $xy = (xy)^{-1} = y^{-1}x^{-1} = yx$. Получили свойство коммутативности. □

Аксиомы единицы и обратного элемента гарантируют, что в любой группе разрешимы линейные уравнения $ax = b$ и $xa = b$ (относительно неизвестной x):

$$\begin{aligned} ax = b \quad &\Leftrightarrow \quad a^{-1}ax = a^{-1}b \quad \Leftrightarrow \quad x = a^{-1}b, \\ xa = b \quad &\Leftrightarrow \quad xaa^{-1} = ba^{-1} \quad \Leftrightarrow \quad x = ba^{-1} \end{aligned}$$

(знаком $\Leftrightarrow$ обозначается равносильность утверждений).

Пример 2.18. Решим уравнение $(1456)(27) \circ x = (12345678)$ в группе перестановок S_8. Умножая равенство слева на $\big((1456)(27)\big)^{-1} = (6541)(72)$, получаем

$$x = (6541)(72) \circ (12345678) = (178623),$$

это и есть единственный корень уравнения. □

Ещё одно полезное в вычислениях с группами свойство — *закон сокращения*: из равенства $ac = bc$ следует равенство $a = b$. Это совершенно очевидно: если $ac = bc$, то и $acc^{-1} = bcc^{-1}$, поэтому $a = b$. Аналогичный закон сокращения выполняется и для умножения слева.

Замечание 2.19. В школьной математике сокращение множителей выполняется более сложным образом: нужно указывать, что $c \neq 0$. Это происходит из-за того, что все (действительные) числа не образуют группу относительно умножения, а ненулевые числа образуют.

Пусть x — элемент группы G. Определим неотрицательные степени этого элемента соотношениями

$$x^0 = e;\ x^1 = x;\ x^k = x \cdot x^{k-1} = \underbrace{x \cdot x \cdot \ldots \cdot x}_{k \text{ раз}}. \tag{2.1}$$

Аналогично определим отрицательные степени x:

$$x^{-k-1} = x^{-1} \cdot x^{-k}.$$

Если групповая операция записывается аддитивно, с помощью знака сложения $+$, то степени элемента обозначаются как целые кратные: nx, $n \in \mathbb{Z}$, $x \in G$. Это соответствует данным выше определениям, если применить их к операции сложения.

Для степеней элемента в любой группе справедливы те же соотношения, которые выполняются для степеней чисел.

Лемма 2.20. *Для любой группы G, любого элемента $x \in G$ и любых целых показателей n, m выполняются равенства*

$$x^n \cdot x^m = x^{n+m},$$
$$\big(x^n\big)^m = x^{nm}.$$

Доказательство. Если одно из чисел равно 0, то оба равенства леммы очевидны. Они также очевидны для положительных n, m (см. формулу (2.1)).

Теперь проверим частный случай второго равенства

$$(x^n)^{-1} = (x^{-1})^n = x^{-n}, \quad n > 0. \tag{2.2}$$

Проверка прямым вычислением, в котором мы пользуемся тем, что x и x^{-1} коммутируют (в любом порядке их произведение равно нейтральному элементу).

$$\big(\underbrace{x \cdot x \cdot \ldots \cdot x}_{n \text{ раз}}\big) \cdot \big(\underbrace{x^{-1} \cdot x^{-1} \cdot \ldots \cdot x^{-1}}_{n \text{ раз}}\big) = \big(\underbrace{(x \cdot x^{-1}) \cdot \ldots \cdot (x \cdot x^{-1})}_{n \text{ раз}}\big) = e.$$

Проверка остальных случаев производится рутинными рассуждениями в духе школьных доказательств тех же равенств.

Для отрицательных n и m сделанного вычисления сводит первое равенство к случаю положительных n и m.

Не ограничивая общности считаем, что $n \geqslant m$. Поэтому для доказательства первого равенства остаётся рассмотреть случай $n > 0 > m$. В этом случае

$$x^n \cdot x^m = \big(\underbrace{x \cdot x \cdot \ldots \cdot x}_{n \text{ раз}}\big) \cdot \big(\underbrace{x^{-1} \cdot x^{-1} \cdot \ldots \cdot x^{-1}}_{-m \text{ раз}}\big) =$$
$$= x^{n-(-m)} \cdot \big(\underbrace{(x \cdot x^{-1}) \cdot \ldots \cdot (x \cdot x^{-1})}_{-m \text{ раз}}\big) = x^{n+m}.$$

Проверим второе равенство для отрицательных m и положительных n, используя (2.2):

$$\big(x^n\big)^m = \big((x^n)^{-1}\big)^{-m} = \big((x^{-1})^n\big)^{-m} = (x^{-1})^{-nm} = x^{-(-nm)} = x^{nm}.$$

Аналогично рассуждаем и в случае отрицательного n: в силу (2.2) имеем $x^n = (x^{-1})^{-n}$, поэтому

$$\big(x^n\big)^m = \big((x^{-1})^{-n}\big)^m = (x^{-1})^{-nm} = x^{-(-nm)} = x^{nm}.$$

На этом разбор случаев закончен. □

Закончим этот раздел необязательным замечанием об избыточности выбранного нами определения группы. В определении группы мы указывали по два равенства в аксиомах единицы и обратного элемента. Это сделано для простоты но, вообще говоря, не обязательно.

Для некоммутативных операций иногда бывают нужны *левые* и *правые* единицы. Элемент e называется левой единицей, если для любого x выполнено $ex = x$. Аналогично, для правой единицы выполняется тождество $xe = x$.

Аналогичные определения можно дать и для обратных элементов. Мы их уже фактически использовали в разделе 1.4 для левых и правых обратных отображений.

Утверждение 2.21. *Пусть на множестве G задана ассоциативная операция $*$, для которой есть правая единица e (то есть, для любого $x \in G$ выполнено $x * e = x$) и для любого $x \in G$ есть правый обратный x^{-1} ($x * x^{-1} = e$). Тогда G с операцией $*$ является группой. (в частности, единица в ней единственна и для каждого элемента обратный тоже единственен).*

Доказательство. Докажем, что всякий правый обратный является левым обратным: умножим равенство $x^{-1} * x * x^{-1} = x^{-1}$ справа на $(x^{-1})^{-1}$, получаем $x^{-1} * x = e$.

Докажем, что всякая правая единица является левой единицей: $e * x = x * x^{-1} * x = x * e = x$.

Теперь легко проверяется единственность единицы и обратного.

Пусть e, e' — две единицы. Тогда $e = e * e' = e'$.

Пусть $x * y = x * z = e$. Поскольку y, z являются также левыми обратными, получаем цепочку равенств: $y = y * e = y * x * z = e * z = z$. □

2.3 Подгруппы и смежные классы

Группа действительных чисел по сложению содержит в себе группу целых чисел по сложению. Существенно, что не только одно множество включено в другое, но и что ограничение операции на меньшее множество даёт в точности операцию в меньшей группе. Зафиксируем это важное отношение между группами в формальном определении.

Определение 2.22. Подмножество H элементов группы G называется *подгруппой*, если для него выполняются следующие свойства

1. если $a, b \in H$, то $a \cdot b \in H$;
2. $e \in H$;
3. если $a \in H$, то $a^{-1} \in H$.

Обозначение $H < G$ указывает, что H — подгруппа G.

Эти свойства в точности и означают, что на множестве H корректно определена операция, задаваемая ограничением групповой операции G (первое свойство в определении). Эта операция автоматически ассоциативна, так как ассоциативна групповая операция в самой группе G. Второе и третье свойства в определении подгруппы утверждают выполнение для неё аксиомы единицы и аксиомы обратного элемента.

Пример 2.23. В разделе 1.3 при построении множества вычетов по модулю n использовалось множество

$$n\mathbb{Z} = \{x : x = qn,\ q \in \mathbb{Z}\}$$

кратных числа n. Это множество является подгруппой. Оно замкнуто относительно групповой операции: $q_1 n + q_2 n = (q_1 + q_2)n \in n\mathbb{Z}$; нейтральный элемент относительно сложения, то есть 0, принадлежит этому множеству: $0 = 0 \cdot n$; если число a является кратным n, то и противоположное число также кратно n.

Конструкция построения аддитивной группы вычетов по модулю n с помощью подгруппы кратных n имеет важное для теории групп обобщение (факторгруппы), которое подробно разбирается ниже. □

Пример 2.24. Рассмотрим те перестановки из симметрической группы S_n, которые оставляют на месте число 1. Проверим, что они образуют подгруппу.

Если $\pi(1) = 1$ и $\sigma(1) = 1$, то $\pi \circ \sigma(1) = \pi(1) = 1$, то есть выполнено свойство 1 из определения.

Тождественная перестановка оставляет на месте все числа, в том числе и 1. Значит, выполнено и свойство 2.

Наконец, используя биективность перестановки, заключаем, что $\pi(1) = 1$ влечёт $\pi^{-1}(1) = 1$ (в общем случае $\pi(a) = b$ влечёт $\pi^{-1}(b) = a$). □

Мы задали подгруппу явным указанием свойств. Оказывается, что эти свойства равносильны требованию, чтобы H являлась группой относительно групповой операции G.

Лемма 2.25. *Подмножество H является подгруппой тогда и только тогда, когда ограничение H является группой относительно ограничения групповой операции на G.*

Доказательство. В одну сторону мы уже обсудили доказательство. Теперь в другую сторону.

Пусть H является группой относительно ограничения групповой операции на G. Это означает, в частности, что групповая операция с элементами из H даёт результат, который также принадлежит H, то есть первое свойство подгруппы в определении 2.22.

Пусть e_H — нейтральный элемент в группе на множестве H, а e_G — нейтральный элемент в группе на множестве G. Тогда для любого элемента $h \in H$ выполняются равенства $e_H h = h = e_G h$ и из свойства сокращения получаем $e_H = e_G$. То есть, выполняется второе свойство из определения 2.22.

Проверим последнее свойство. Пусть g — обратный к h в группе на множестве H, а h^{-1} — обратный к h в группе G. Получаем равенства $gh = e = h^{-1}h$, откуда по свойству сократимости получаем, что $g = h^{-1}$, это и есть третье свойство из определения 2.22. □

Укажем ещё один критерий подгруппы, который иногда бывает удобнее исходного определения 2.22.

Теорема 2.26. *Множество H является подгруппой группы G тогда и только тогда, когда для любых $a, b \in H$ выполнено $ab^{-1} \in H$.*

Доказательство. Если H — подгруппа, то для любых $a, b \in H$ из свойств 3 и 1 определения 2.22 подгруппы следует, что $ab^{-1} \in H$.

Теперь докажем необходимость. Пусть для любых $a, b \in H$ выполнено $ab^{-1} \in H$. Возьмём любой элемент $a \in H$. Тогда $e = a \cdot a^{-1} \in H$ (свойство 2). Выбрав пару $e, a \in H$, убеждаемся, что $ea^{-1} = a^{-1} \in H$ (свойство 3). Поскольку $(b^{-1})^{-1} = b$, то $ab = a(b^{-1})^{-1}$ и свойство 1 также выполнено. □

Применение этого критерия упрощает доказательство простого факта о пересечении подгрупп.

Лемма 2.27. *Пересечение подгрупп — подгруппа.*

Доказательство. Рассмотрим две подгруппы H_1, H_2 некоторой группы G.

Пусть $a, b \in H_1 \cap H_2$. Тогда $ab^{-1} \in H_1$, $ab^{-1} \in H_2$. Значит, $ab^{-1} \in H_1 \cap H_2$. Из теоремы 2.26 следует, что $H_1 \cap H_2$ — подгруппа. □

Пример 2.28 (подгруппа, порождённая множеством)**.** Рассмотрим некоторое подмножество $S \subseteq G$ элементов подгруппы и пересечение всех групп, содержащих это множество. В силу леммы 2.27 это пересечение само по себе является подгруппой. Обозначим такую подгруппу $\langle S \rangle$. Она содержит S по построению и содержится в любой подгруппе, которая содержит множество элементов S. Другими словами, это минимальная (по включению) подгруппа, содержащая S.

Для подгруппы $\langle S \rangle$, $S \neq \varnothing$, есть альтернативное определение (то есть равносильное свойство). А именно, рассмотрим все возможные произведения

$$g_1 \cdot g_2 \cdot \ldots \cdot g_n,$$

в которых $g_i \in S$ или $g_i^{-1} \in S$. Эти произведения в точности совпадают с подгруппой $\langle S \rangle$.

Действительно, в силу свойств 1 и 3 определения подгруппы все такие произведения обязаны входить в любую подгруппу, содержащую S. Поэтому они лежат в пересечении таких подгрупп, которое мы и обозначили через $\langle S \rangle$. С другой стороны, они уже образуют подгруппу. Замкнутость относительно произведения очевидна: попросту припишем одно произведение к другому. Среди этих произведений обязательно есть нейтральный элемент группы, так как $e = g \cdot g^{-1}$ для любого $g \in S$.

Наконец, из леммы 2.15 по индукции получаем

$$(g_1 \cdot g_2 \cdot \ldots \cdot g_n)^{-1} = g_n^{-1} \cdot g_{n-1}^{-1} \cdot \ldots \cdot g_2^{-1} \cdot g_1^{-1},$$

откуда следует свойство 3 из определения подгруппы.

Используя понятие подгруппы, порождённой множеством элементов, можно компактно сформулировать утверждения из раздела 1.5:

$$S_n = \langle \{(i\ j) : 1 \leqslant i < j \leqslant n\} \rangle \quad \text{теорема 1.76,}$$
$$S_n = \langle \{\big(i\ (i+1)\big) : 1 \leqslant i < n\} \rangle \quad \text{лемма 1.77.}$$ □

В общем случае подмножество, которое лишь замкнуто относительно групповой операции (свойство 1 в определении 2.22), не обязано быть подгруппой.

Контрольный вопрос 2.29. Проверьте, что множество $\mathbb{N}$ неотрицательных целых чисел замкнуто относительно сложения, но не является подгруппой группы $(\mathbb{Z}, +)$ целых чисел с операцией сложения.

Однако для конечных групп ситуация упрощается.

Лемма 2.30. *Пусть G — конечная группа. Множество $H \subseteq G$ является подгруппой группы G тогда и только тогда, когда H замкнуто относительно групповой операции.*

Доказательство. В одну сторону утверждение очевидно, так как замкнутость относительно групповой операции входит в определение подгруппы.

В другую сторону: в силу теоремы 2.26 достаточно доказать, что если H замкнуто относительно групповой операции, то оно замкнуто и относительно обращения.

Пусть $h \in H$. Рассмотрим последовательность положительных степеней h:

$$h^1 = h;\ h^2 = h \cdot h;\ \ldots\ h^{k+1} = h \cdot h^k.$$

Поскольку H замкнуто относительно групповой операции, все члены этой последовательности лежат в H.

Это бесконечная последовательность элементов конечного множества. Поэтому какие-то члены в ней одинаковы, $h^s = h^t$, $s \neq t$. Считаем не ограничивая общности, что $s > t$.

Умножим равенство $h^s = h^t$ на $(h^t)^{-1} = (h^{-1})^t$ (см. лемму 2.20). Получаем $h^{s-t} = e$. Если $s - t = 1$, то $h = e$, то есть совпадает со своим обратным. Если $s - t > 1$, то $h^{s-t-1} \cdot h = h^{s-t} = e$, а с другой стороны h^{s-t-1} принадлежит последовательности положительных степеней h, которая целиком лежит в H. □

Мы привели несколько примеров подгрупп. Сейчас рассмотрим конструкции, которые дают подгруппы абелевых групп, но необязательно дают подгруппы произвольных групп.

Пример 2.31 (подгруппа квадратов)**.** Множеством квадратов группы G назовём множество $G^{(2)}$, состоящее в точности из тех элементов группы, которые имеют вид g^2, $g \in G$.

Пусть A — абелева группа. Тогда $A^{(2)}$ — подгруппа A. Действительно, $a^2b^2 = (ab)(ab) = (ab)^2$ в силу коммутативности умножения в абелевой группе, $e = e^2$, а $(g^2)^{-1} = g^{-1} \cdot g^{-1}$ (проверьте последнее свойство прямым вычислением).

В общем случае множество квадратов не образует группу. Рассмотрим основной для нас пример неабелевой группы — группу перестановок. Квадраты перестановок обладают особыми свойствами и незамкнуты относительно умножения. Рассмотрим пример квадратов в группе S_6 перестановок 6 элементов. Перестановки $(12)(34)$ и (456) являются квадратами, как показывают равенства

$$(12)(34) = (1324) \circ (1324),$$
$$(456) = (654) \circ (654).$$

Их произведение равно, как нетрудно вычислить,

$$(12)(34) \circ (456) = (12)(3456)$$

и не является квадратом никакой перестановки. Как это проверить, не перебирая все возможные квадраты перестановок из 6 элементов?

Поскольку любая перестановка разлагается в произведение непересекающихся циклов, рассмотрим квадраты циклов. Квадрат цикла нечётной длины является циклом той же длины:

$$\big(1\ 2\ \ldots\ (2t+1)\big) \circ \big(1\ 2\ \ldots\ (2t+1)\big) = \big(1\ 3\ 5\ \ldots\ (2t+1)\ 2\ 4\ \ldots\ 2t\big).$$

А квадрат цикла чётной длины является произведением двух циклов половинной длины:

$$\big(1\ 2\ \ldots\ 2t\big) \circ \big(1\ 2\ \ldots\ 2t\big) = \big(1\ 3\ 5\ \ldots\ (2t-1)\big)\big(2\ 4\ \ldots\ 2t\big).$$

Поэтому в цикловом разложении квадрата перестановки циклы чётной длины должны встречаться парами, так как они могут появиться только из квадрата цикла удвоенной длины.

В перестановке (12)(3456) есть один цикл длины 2 и один цикл длины 4. Поэтому она не является квадратом никакой перестановки. □

С каждой подгруппой связано разбиение группы на множества, называемые *смежными классами*. Мы уже использовали частный случай этой конструкции в разделе 1.3, когда определяли вычеты по модулю n. Вычеты — это и есть смежные классы по подгруппе $n\mathbb{Z}$ кратных числа n, как мы сейчас увидим.

Определение 2.32. Пусть $H < G$ — подгруппа группы G, а x — некоторый элемент группы G.

Левый смежный класс по подгруппе H с представителем x — это множество

$$xH = \{y : y = xh,\ h \in H\}.$$

Аналогично, *правый смежный класс* по подгруппе H с представителем x — это множество

$$Hx = \{y : y = hx,\ h \in H\}.$$

Ясно, что всегда $x \in xH$ и $x \in Hx$. Для абелевых групп разницы между левыми и правыми смежными классами нет. В этом случае говорят просто о смежном классе (так же говорят и в неабелевом случае, когда выбор левого или правого смежного класса ясен из контекста).

В этом определении использовано полезное обозначение, которое распространяет произведение на подмножества элементов группы. Пусть A, B — подмножества элементов группы G. Тогда «произведение» AB подмножества A на подмножество B — это подмножество элементов группы, состоящее из всех попарных произведений ab, $a \in A$, $b \in B$. При умножении одноэлементного подмножества $\{x\}$ фигурные скобки для краткости записи опускаются.

Контрольный вопрос 2.33. Проверьте ассоциативность умножения множеств и дистрибутивность относительно операции объединения множеств: $(AB)C = A(BC)$, $A(B \cup C) = AB \cup AC$, $(A \cup B)C = AC \cup BC$ для любых подмножеств A, B и C.

Проверьте, что для любой подгруппы $H^2 = H$.

Почему G^2 не совпадает в общем случае с множеством квадратов группы?

Пример 2.34. Рассмотрим группу $\mathbb{R}^2$, которая состоит из пар действительных чисел с операцией покомпонентного сложения:

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$$

(используем здесь аддитивную запись). В этой группе есть подгруппа R, состоящая из пар $(x, 0)$, $x \in \mathbb{R}$ (проверку свойств подгруппы оставляем читателю в качестве полезного упражнения).

Смежными классами будут множества $R_a = \{(x, y) : y = a,\ x \in \mathbb{R}\}$. Это очевидно, так как пары с совпадающей второй компонентой отличаются на элемент из R:

$$(x_1, a) = (x_2, a) + (x_1 - x_2, 0).$$ □

Пример 2.35. Рассмотрим некоммутативную группу S_3 и в ней подгруппу $H = \langle(12)\rangle$ (используем сокращённую цикловую запись перестановок). Запишем смежные классы по этой подгруппе

$$\begin{aligned} H: &\quad () \quad (12) \\ (23)H: &\quad (23) \quad (132) \\ (13)H: &\quad (13) \quad (123) \end{aligned}$$ □

Следующий факт о смежных классах принципиально важен, хотя и очень просто доказывается.

Теорема 2.36 (теорема о смежных классах)**.** *Смежные классы xH и yH либо не пересекаются, либо совпадают.*

Доказательство. Предположим, что у двух смежных классов нашёлся общий элемент z. Тогда $z = xh_i = yh_j$. Отсюда получаем $x = y(h_j h_i^{-1}) \in yH$, $y = x(h_i h_j^{-1}) \in \in xH$, то есть представитель одного класса принадлежит другому.

Пусть $w \in xH$, то есть $w = xh'$. Но $x = yh_x$, значит, $w = y(h_x h') \in yH$. Таким образом, $xH \subseteq yH$. Совершенно аналогично доказывается включение в противоположную сторону. Значит, $xH = yH$. Мы доказали, что смежные классы совпадают, если у них есть хотя бы один общий элемент. □

Поскольку сам элемент g группы G принадлежит смежному классу gH по подгруппе H с представителем g, то вся группа G разбивается на объединение смежных классов по H (в данном случае левых, хотя то же самое верно и для правых смежных классов):

$$G = \bigsqcup_i g_i H.$$

На словах это означает, что каждый элемент группы входит в точности в один (левый) смежный класс по подгруппе. В разделе 1.2.3 мы доказали, что разбиения множества на непересекающиеся подмножества взаимно однозначно соответствуют отношениям эквивалентности на множестве (лемма 1.20). Так что теорему о смежных классах можно переформулировать равносильным образом так: для любой подгруппы отношение «x принадлежит смежному классу по подгруппе с представителем y» является отношением эквивалентности (хотя в такой формулировке оно даже не выглядит симметричным). Это свойство смежных классов понадобится нам далее при построении факторгрупп.

Контрольный вопрос 2.37. Докажите, что сама подгруппа H является смежным классом по подгруппе H. С каким представителем?

Заметим, что правые смежные классы задают в общем случае другое разбиение и другое отношение эквивалентности.

Аналогично критерию подгруппы 2.26 можно сформулировать критерии принадлежности левым и правым смежным классам. Обратите внимание, что формулировки несимметричны, хотя само отношение симметрично как всякое отношение эквивалентности.

Лемма 2.38. *Элементы x, y группы G принадлежат одному левому смежному классу по подгруппе H тогда и только тогда, когда $y^{-1}x$ принадлежит H.*

Элементы x, y группы G принадлежат одному правому смежному классу по подгруппе H тогда и только тогда, когда xy^{-1} принадлежит H.

Доказательство. Принадлежность x, y одному левому смежному классу равносильна тому, что $x \in yH$, то есть $x = yh$ для некоторого $h \in H$. Это и означает, что $y^{-1}x \in H$ (домножим равенство слева на y^{-1}).

Аналогично для правых смежных классов: условие $x \in Hy$ равносильно тому, что $x = hy$ для некоторого $h \in H$, что в свою очередь равносильно $xy^{-1} \in H$ (теперь умножаем равенство на y^{-1} справа). □

Рассмотрим пример, который позволяет понять разницу в разбиениях на левые и правые смежные классы. Опять используем группу перестановок в качестве образцовой неабелевой группы.

Пример 2.39. Пусть G — множество перестановок из S_n, которые чётные числа переводят в чётные. Это множество замкнуто относительно композиции, так что по теореме 2.30 оно является подгруппой S_n.

Применим критерий 2.38 для принадлежности двух перестановок одному левому смежному классу по подгруппе G. Он говорит, что $\pi \in \sigma G$ равносильно тому, что $\sigma^{-1} \circ \pi \in G$, то есть композиция $\sigma^{-1} \circ \pi$ сохраняет множество чётных чисел. Для перестановок это можно сформулировать так: образы множества чётных чисел для перестановок π и σ одинаковы. В таком случае перестановка переводит чётное число в элемент образа чётных чисел, обозначим этот образ X. Обратная к σ перестановка отправляет X в множество чётных чисел. Обратное утверждение также легко доказывается.

Более кратко и формально это рассуждение можно записать так. Пусть E — множество чётных чисел. Тогда $\sigma^{-1} \circ \pi \in G$ равносильному тому, что $E = (\sigma^{-1} \circ \pi)E$, что в свою очередь равносильно $\sigma(E) = \pi(E)$ (для равносильности существенна биективность перестановок).

Аналогично для правых смежных классов. Но теперь условие $\pi \circ \sigma^{-1} \in G$ равносильно тому, что $E = (\pi \circ \sigma^{-1})E$, что равносильно $\pi^{-1}E = \sigma^{-1}E$. На словах это означает, что совпадают прообразы множества чётных чисел для перестановок π и σ.

Пользуясь этим критерием, легко проверить, что перестановки (2 3 5) и (2 5) принадлежат одному правому смежному классу по G, так как прообраз множества чётных чисел в обоих случаях 5 и чётные числа, отличные от 2.

Но эти перестановки не принадлежат одному левому смежному классу по G: в первом случае образ чётных чисел — это 3 и чётные числа, отличные от 2, а во втором — 5 и чётные числа, отличные от 2. □

2.4 Теорема Лагранжа. Приложения

Разбиение на смежные классы является «однородным»: все смежные классы имеют одинаковую мощность (в случае конечных групп одинаковое количество элементов).

Лемма 2.40. *Пусть $H < G$ — подгруппа группы G, а x — некоторый элемент группы G. Тогда отображение «сдвига» $f\colon h \mapsto xh$ задаёт биекцию между подгруппой H и смежным классом с представителем H.*

Доказательство. Для любого $h \in H$ образ $f(h) = xh$ принадлежит xH по определению. Опять-таки, по определению это отображение сюръективно. Инъективность следует из закона сокращения в группе: из равенства $xh_1 = xh_2$ следует равенство $h_1 = h_2$. □

Следствие 2.41. *Если в подгруппе H конечное количество элементов, то $|H| = |xH|$ для любого x.*

Количество смежных классов группы G по подгруппе H называется *индексом подгруппы* и обозначается через $(G : H)$. (Если смежных классов по H бесконечно много, то H называется подгруппой бесконечного индекса.)

Теорема 2.42 (теорема Лагранжа)**.** *Пусть H — подгруппа конечной группы G. Тогда порядок H является делителем порядка G и, более того,*

$$|G| = (G : H) \cdot |H|.$$

Доказательство. Фактически уже всё доказано. Теорема 2.36 говорит, что группа G разбивается на $(G : H)$ смежных классов по подгруппе H, а следствие 2.41 говорит, что в каждом смежном классе ровно $|H|$ элементов. Поэтому общее количество элементов в группе, то есть порядок группы $|G|$, равно произведению индекса $(G : H)$ на порядок подгруппы $|H|$. □

Теорема Лагранжа накладывает теоретико-числовые ограничения на существование подгрупп заданного порядка в группе: порядок подгрупп обязательно является делителем порядка группы.

Получаем, в частности, такое следствие. В любой группе есть так называемые *несобственные* подгруппы: единичная подгруппа $\{e\}$ и вся группа. Остальные подгруппы называются *собственными*.

Следствие 2.43. *Группа простого порядка не имеет несобственных подгрупп.*

Доказательство. Простое число не имеет делителей, отличных от 1 и самого этого числа. Значит, порядок подгруппы равен либо 1 (единичная подгруппа, так как единичный элемент лежит во всякой подгруппе), либо совпадает с порядком всей группы. □

Однако условие из теоремы Лагранжа не является достаточным. В общем случае есть и другие препятствия к существованию подгрупп.

Пример 2.44. Докажем, что в группе S_5 нет подгруппы порядка $40 \mid 120 = 5! = |S_5|$.

Индекс такой подгруппы обязан равняться 3 в силу теоремы Лагранжа. Значит, если $H < S_5$ и $|H| = 40$, то группа S_5 разбивается ровно на три различных смежных класса по H. Обозначим их H, aH и bH.

Пусть π — цикл длины 5. Какая-то степень π должна попасть в H, ведь группа конечная. Пусть k — наименьший положительный показатель степени, для которого $\pi^k \in H$. Заметим, что $1 \leqslant k < 5$, так как какие-то две степени p_1, p_2 меньше 5 лежат в одном классе смежности (их ведь всего $3 < 4$). Но тогда из леммы 2.38 получаем, что $\pi^{p_1 - p_2} \in H$.

Поскольку 5 — простое число, k взаимно просто с 5. Поэтому по теореме 1.48 для некоторого ℓ выполняется сравнение $k\ell \equiv 1 \pmod 5$, то есть $k\ell = 1 + 5q$. Поэтому перестановка $(\pi^k)^\ell = \pi^{k\ell} = \pi^{1+5q} = \pi$ принадлежит H. (Последнее равенство выполняется из-за того, что n-я степень цикла сдвигает числа по циклу на n, так что $(1 + 5q)$-я степень сдвигает числа так же, как первая.)

Итак, мы доказали, что все циклы длины 5 принадлежат подгруппе H. Любой цикл длины 3 можно представить как произведение циклов длины 5, как видно из примера

$$(1\ 2\ 3) = (1\ 5\ 4\ 3\ 2) \circ (4\ 5\ 1\ 3\ 2).$$

Поэтому все циклы длины 3 также принадлежат H.

Теперь займёмся подсчётами. Всего в группе S_5 есть $5!/5 = 24$ циклов длины 5 и $\binom{5}{3} \cdot 2 = 20$ циклов длины 3. Но $24 + 20 > 40$, что противоречит сделанному предположению, что порядок группы H равен 40. □

Пример 2.45 (порядки групп многогранников)**.** В разделе 1.6.2 мы нашли порядок группы диэдра D_n довольно замысловатым рассуждением (можно было бы и проще). Но это рассуждение является фактически применением теоремы Лагранжа, чем и ценно для нашего изложения.

Действительно, мы нашли подгруппу H_a симметрий правильного n-угольника, которые оставляют на месте вершину a. Порядок этой группы равен 2.

Далее мы нашли количество смежных классов по этой подгруппе. Оно совпадает с количеством вершин. Действительно, любая симметрия t правильного n-угольника переводит вершину a в некоторую вершину a'. Поэтому $t = r_{aa'} \circ t'$, где t' — симметрия, оставляющая вершину a на месте, а $r_{aa'}$ — поворот, переводящий вершину a в некоторую вершину a'. Это значит, что смежных классов не больше, чем вершин.

Все такие повороты лежат в разных левых смежных классах по подгруппе H_a: если $a' \neq a''$, то $r_{aa'}^{-1} \circ r_{aa''} = r_{a'a} \circ r_{aa''}$ не оставляет a на месте, поэтому по лемме 2.38 эти повороты лежат в разных смежных классах. Это значит, что количество смежных классов совпадает с числом вершин.

Теперь осталось применить теорему Лагранжа: $|D_n| = |H_a| \cdot (D_n : H_a) = 2 \cdot n = 2n$.

Точно то же рассуждение применимо к группам многогранников. Если H_v — подгруппа вращений многогранника, оставляющая вершину v на месте, то количество

смежных классов по этой подгруппе равно количеству вершин (для любой пары вершин существует вращение, которое переводит первую вершину во вторую).

Порядок подгруппы вращений многогранника, сохраняющих заданную вершину, найти легко. Он совпадает с количеством рёбер, исходящих из этой вершины. Применяя теорему Лагранжа, получаем такие порядки для групп многогранников:

группа тетраэдра: порядок H_v равен 3, всего вершин 4, значит, порядок группы равен $3 \cdot 4 = 12$;

группа куба: порядок H_v равен 3, всего вершин 8, значит, порядок группы равен $3 \cdot 8 = 24$;

группа октаэдра: порядок H_v равен 4, всего вершин 6, значит, порядок группы равен $4 \cdot 6 = 24$;

группа додекаэдра: порядок H_v равен 3, всего вершин 20, значит, порядок группы равен $3 \cdot 20 = 60$;

группа икосаэдра: порядок H_v равен 5, всего вершин 12, значит, порядок группы равен $5 \cdot 12 = 60$.

Обратите внимание на то, что порядки групп куба и октаэдра, как и групп додекаэдра и икосаэдра совпадают. Это не случайно. Как будет видно дальше, это по сути дела одни и те же группы. Наше рассуждение выделяет в этих группах разные подгруппы, поэтому вычисления оказываются различными. □

Теорема Лагранжа имеет многочисленные приложения в алгебре и элементарной теории чисел. Многие из них основаны на частном случае этой теоремы, когда подгруппа H порождена одним элементом, $H = \langle h \rangle$. В этом случае порядок $\langle h \rangle$ называется также *порядком элемента* h, общее обозначение $\operatorname{ord} h$. Более традиционное определение для порядка элемента получается из следующего утверждения.

Утверждение 2.46. *Порождённая элементом h подгруппа совпадает с множеством целых степеней h. Все различные степени имеют показатели* $0, 1, \ldots, \operatorname{ord} h - 1$.

Доказательство. Первое утверждение очевидно из доказанной выше леммы 2.20 о свойствах степеней.

Докажем второе утверждение. Пусть h^i — последовательность неотрицательных степеней h. Рассмотрим первое повторение в ней $h^s = h^t$, $s > t$. Тогда, как уже проверялось выше, $h^{s-t} = e$ и $h^i \neq e$ при $i < s - t$.

Обозначим $s - t = d$ и докажем, что любая степень h^k равна какой-то степени h^i, $i < d$. Действительно, разделим k на d с остатком: $k = qd + i$, $0 \geqslant i < d$. Тогда, используя лемму о свойствах степеней, получаем равенства

$$h^k = h^{qd+i} = h^{qd} \cdot h^i = (h^d)^q \cdot h^i = e^q \cdot h^i = h^i,$$

что и требовалось.

С другой стороны, $h^i \neq h^j$ при $i \neq j$, $0 \leqslant i < d$, $0 \leqslant j < d$, так как при построении d мы выбирали первое повторение в последовательности степеней.

Отсюда следует, что $d = \operatorname{ord} h$ и второе утверждение полностью доказано. □

Это утверждение показывает, что данное выше определение порядка элемента совпадает с традиционным.

Определение 2.47. Порядок $\operatorname{ord}_G h$ элемента h группы G — это наименьшее среди тех положительных чисел d, для которых $h^d = e$. Если таких чисел нет, порядок считается бесконечным (у группы $\langle h \rangle$ в этом случае также бесконечный порядок).

По теореме Лагранжа порядок элемента (как и порядок любой подгруппы) делит порядок группы. Приведём несколько следствий из этого утверждения. Первое выполняется для любых конечных групп.

Следствие 2.48. *Пусть x — элемент группы G из n элементов. Тогда $x^n = e$.*

Доказательство. Действительно, порядок элемента $\operatorname{ord} x$ является делителем порядка группы, то есть $n = q \operatorname{ord} x$. Из леммы 2.20 о свойствах степеней получаем $x^n = x^{q \operatorname{ord} x} = \left(x^{\operatorname{ord} x}\right)^q = e^q = e$. □

Аналогично примеру 2.44, обращение этого утверждения неверно: не всякий делитель порядка группы является порядком некоторого её элемента. Мы приведём пример в группе перестановок, когда научимся находить порядки перестановок, см. раздел 2.7.

Если пересказать следствие 2.48 для мультипликативной группы вычетов по простому и произвольному модулю, получим две теоремы элементарной теории чисел с громкими названиями.

Следствие 2.49 (малая теорема Ферма)**.** *Если p — простое число, то для любого $a \not\equiv 0 \pmod p$ выполняется сравнение $a^{p-1} = 1 \pmod p$.*

Малая теорема Ферма совпадает со следствием 2.48 для мультипликативной группы вычетов по простому модулю, так как все ненулевые остатки по модулю простого числа взаимно просты с этим простым числом. Таким образом, порядок группы Z_p^*, p — простое, равен $p - 1$.

Теорема 1.48 позволяет сформулировать малую теорему Ферма более симметрично. Такая формулировка пригодится нам позже, во второй части курса.

Следствие 2.50 (малая теорема Ферма для всех вычетов)**.** *Если p — простое число, то для любого a выполняется сравнение $a^p = a \pmod p$.*

Доказательство. Сравнение очевидно выполнено для тех a, которые кратны p (обе части сравнения делятся на p, то есть равны 0 по модулю p).

Для остальных значений a это сравнение равносильно приведённому выше в следствии 2.49 (оно получается умножением на a по модулю p). □

Теперь рассмотрим случай составного модуля. Как уже говорилось, порядок мультипликативной группы вычетов по модулю n равен функции Эйлера $\varphi(n)$, то есть количеству остатков, взаимно простых с n. Из следствия 2.48 получаем такой результат.

Следствие 2.51 (теорема Эйлера)**.** *Пусть a взаимно просто с n. Тогда $a^{\varphi(n)} \equiv \equiv 1 \pmod{n}$.*

Малая теорема Ферма и теорема Эйлера упрощают вычисления по модулю, в которых встречаются большие степени. Приведём простые примеры (ниже мы научимся вычислять функцию Эйлера для больших чисел и список подобных примеров будет расширен).

Пример 2.52. Найдём остаток при делении 19^{91} на 31. Поскольку 31 — простое число, малая теорема Ферма даёт сравнение $19^{30} \equiv 1 \pmod{31}$. Но тогда

$$19^{91} = 19^{90} \cdot 19 = (19^{30})^3 \cdot 19 \equiv 1 \cdot 19 = 19 \pmod{31}.$$

Ответ: 19. □

Пример 2.53. Найдём остаток при делении 19^{81} на 32. Чтобы применить теорему Эйлера, нужно знать значение $\varphi(32)$. В данном случае несложно понять, что все остатки, которые не взаимно просты с 32, нечётны, причём любой нечётный остаток взаимно прост с 32. Значит, $\varphi(32) = 16$. Поскольку 19 взаимно просто с 32, из теоремы Эйлера получаем $19^{16} \equiv 1 \pmod{32}$. Дальнейшие вычисления аналогичны предыдущему примеру

$$19^{81} = 19^{80} \cdot 19 = (19^{16})^5 \cdot 19 \equiv 1 \cdot 19 = 19 \pmod{32}.$$ □

Пример 2.54. Попробуем вычислить остаток при делении 20^{17} на 32. Тут возникает новая трудность: 20 не взаимно просто с 32. Однако в данном случае эта трудность очень легко преодолевается: $20 = 2^2 \cdot 5$, поэтому уже 20^3 делится на $(2^2)^3 = 2^6 = 64$. Значит, искомый остаток равен 0. □

2.5 Основная теорема арифметики для целых чисел

По определению порядок вычета $[a]$ в группе Z_n равен наименьшему положительному числу k, для которого

$$\underbrace{[a] + [a] + \cdots + [a]}_{k \text{ раз}} \equiv [ka] \equiv 0 \pmod{n}.$$

Этот порядок легко выражается через наибольший общий делитель a и n.

Лемма 2.55. $\operatorname{ord}_{Z_n}(a) = \dfrac{n}{\text{НОД}(a, n)}$.

Доказательство. В разделе 1.3 при анализе вычетов по модулю n мы ввели множество S_a тех вычетов, которые кратны a. В терминах подгрупп это как раз подгруппа $\langle a \rangle$ аддитивной группы вычетов по модулю n.

Мы нашли в разделе 1.3 все вычеты, входящие в эту подгруппу. Это те остатки по модулю n, которые кратны НОД(a, n). Их ровно $n/$ НОД(a, n) штук. Это и есть порядок группы $\langle a \rangle$, который совпадает с порядком вычета a в силу утверждения 2.46. □

Отсюда получаем интересные для теории чисел следствия.

Лемма 2.56. *Для любых двух положительных целых чисел x, y выполняется равенство*

$$\text{НОК}(x, y) = \frac{xy}{\text{НОД}(x, y)} .$$

Доказательство. Из леммы 2.55 следует, что при всех $k < y/\,\text{НОД}(x, y)$ число kx не делится на y (это равносильно тому, что $k[x] \neq 0 \pmod{y}$). С другой стороны, $x \cdot y/\,\text{НОД}(x, y)$ делится на y. □

При решении уравнений в числах (целых, рациональных, действительных или комплексных) часто используется такое утверждение: если произведение нескольких чисел равно 0, то одно из чисел равно 0. Для арифметики по модулю это не всегда верно. Скажем, $2 \cdot 2 \equiv 0 \pmod 4$, но $2 \neq 0 \pmod 4$.

Оказывается, эта трудность возникает только при составных модулях.

Лемма 2.57. *Пусть p — простое число. Тогда из сравнения $xy \equiv 0 \pmod p$ следует $x \equiv 0 \pmod p$ или $y \equiv 0 \pmod p$.*

Доказательство. Пусть $x \neq 0 \pmod p$. Тогда НОД$(x, p) = 1$ и из леммы 2.55 получаем, что $\text{ord}_{Z_p} x = p$ и для любого $0 < r < p$ выполняется $xr \neq 0 \pmod p$. Разделим y на p с остатком: $y = qp + r$. Получаем

$$0 \equiv xy = x(qp + r) \equiv xr \pmod p,$$

то есть $r = 0$. □

Следствие 2.58. *Пусть p — простое число. Тогда из сравнения*

$$x_1 x_2 \cdot \ldots \cdot x_n \equiv 0 \pmod p$$

следует $x_i \equiv 0 \pmod p$ для некоторого $1 \leqslant i \leqslant n$.

Доказательство. Индукция по количеству множителей, для базы и индуктивного шага используется лемма 2.57. □

Лемма 2.57 позволяет анализировать уравнения (сравнения) по простому модулю привычным из элементарной алгебры способом.

Пример 2.59. Сколько решений у сравнения $x^2 \equiv 1 \pmod p$, где $p > 2$ — простое число? (Напомним, что решением сравнения мы называем класс вычетов, поэтому количество решений конечно.)

Два решения видны сразу: $+1$ и -1. А других нет. Действительно, если $x^2 \equiv 1 \pmod p$, то $x^2 - 1 = (x - 1)(x + 1) \equiv 0 \pmod p$. Но раз произведение делится на p, то один из сомножителей делится на p по лемме 2.57. Поэтому $x \equiv 1 \pmod p$ или $x \equiv -1 \pmod p$. □

Контрольный вопрос 2.60. Почему для $p = 2$ получается другой ответ?

Из этого наблюдения можно вывести доказательство известного в теории чисел сравнения.

Пример 2.61. Докажем, что $(p-1)! \equiv -1 \pmod p$ для любого простого числа p. Случай $p = 2$ проверяется в уме.

Если $p > 2$, то в группе Z_p^* есть ровно один элемент порядка 2 (-1) — это переформулировка утверждения о количестве решений квадратичного сравнения из примера 2.59.

Поэтому все элементы этой группы за исключением ± 1 разбиваются на пары взаимно обратных $a_i b_i \equiv 1 \pmod p$, $a_i \neq b_i$. После перегруппировки сомножителей получаем

$$(p-1)! = 1 \cdot 2 \cdot \ldots \cdot (p-1) = 1 \cdot (a_1 \cdot b_1) \cdot \ldots \cdot (a_t \cdot b_t) \cdot (p-1) \equiv p-1 \equiv -1 \pmod p.$$

□

Замечание 2.62. На самом деле верно и обратное: если $(p-1)! \equiv -1 \pmod p$, то p — простое (докажите!). Эти два утверждения вместе составляют *теорему Вильсона*. Она даёт критерий простоты числа. Однако для практических вычислений этот критерий неудобен из-за трудности вычисления факториала.

Для составных модулей подсчёт количества решений у сравнения $x^2 \equiv 1 \pmod p$ сложнее. Их может быть больше двух.

Пример 2.63. Сколько решений у сравнения $x^2 \equiv 1 \pmod 8$? Легко проверить, что их четыре: квадрат любого нечётного числа по модулю 8 равен 1. □

Теперь выведем из леммы 2.57 важную теорему теории чисел.

Эта теорема говорит об однозначности разложения положительного целого числа на простые множители. Утверждение кажется самоочевидным. Однако в более сложных случаях оно может и не выполняться, как мы выясним дальше. Для целых чисел доказательство получается из леммы 2.57 несложной индукцией.

Мы уже неявно использовали эту теорему в примере 2.53, когда находили остатки, взаимно простые с 32. Более подробное рассуждение, которое проведено в этом примере звучит так: так как $32 = 2^5$, а 2 — простое число, то все его делители — это степени 2 (вот здесь и нужна однозначность разложения на простые); поэтому остаток не взаимно прост с 32 тогда и только тогда, когда он чётен.

Замечание 2.64. Умножение чисел коммутативно. Поэтому в строгом смысле однозначности разложения на простые нет, например

$$15 = 3 \cdot 5 = 5 \cdot 3.$$

Имеется в виду однозначность с точностью до перестановок простых множителей. Сгруппировав все множители p в множитель p^a и расположив простые множители

в порядке возрастания, любое разложение на простые множители можно записать в виде

$$p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_s^{a_s}, \quad p_1 < p_2 < \cdots < p_s. \tag{2.3}$$

И вот такая запись (последовательность пар (p_i, a_i), в которой простые числа p_i образуют возрастающую последовательность простых чисел) уже однозначно определена для любого положительного числа $n > 1$.

Теорема 2.65 (основная теорема арифметики)**.** *Для любого целого $n > 1$ существует разложение вида* (2.3) *и это разложение однозначно определено.*

Доказательство. Существование доказывается индукцией по величине n. База индукции $n = 2 = 2^1$.

Индуктивный переход: пусть существование разложения на простые доказано для всех $k < n$. Если n простое, то $n = n^1$ — искомое разложение на простые множители. Если у n есть нетривиальный делитель $1 < d < n$, то возьмём разложения на простые для чисел $d < n$, $n/d < n$ и перемножим их.

Доказательство единственности будем проводить индукцией по количеству простых множителей в разложении, то есть сумме показателей $a_1 + \cdots + a_s$ в выражении (2.3).

База индукции: $n = p$ — простое число. Если $p = q_1 q_2$ для некоторых целых положительных чисел q_1, q_2, то по определению простого числа $q_1 = p$ или $q_2 = p$.

Шаг индукции. Предположим, что единственность доказана для всех чисел, которые разлагаются в произведение $< N$ простых множителей. Рассмотрим два разложения

$$p_1 \cdot p_2 \cdot \ldots \cdot p_N = q_1 \cdot q_2 \cdot \ldots \cdot q_L \tag{2.4}$$

на простые множители, в одном из которых ровно N простых множителей.

Правая часть равенства (2.4) делится на p_N. По следствию 2.58 одно из чисел q_i делится на p_N. Так как все они простые, то это число попросту равно p_N.

Сокращая равенство (2.4) на p_N, получаем равенство двух разложений на простые, в котором одно из разложений содержит меньше N множителей. Из предположения индукции получаем, что эти разложения одинаковы. А значит, одинаковы и разложения в равенстве (2.4), которые получаются из них домножением на один и тот же простой множитель p_N. □

Основная теорема арифметики позволяет по-другому выразить отношение делимости целых чисел. Пусть

$$n = p_1^{a_1} \cdot p_2^{a_2} \cdot \ldots \cdot p_s^{a_s}, \quad m = p_1^{b_1} \cdot p_2^{b_2} \cdot \ldots \cdot p_s^{b_s}, \qquad 0 \leqslant a_i,\ 0 \leqslant b_i. \tag{2.5}$$

Здесь для удобства сравнения мы допускаем нулевые показатели степени. Так как $p^0 = 1$, нулевой показатель означает, что число не делится на p.

Утверждение 2.66. *$n \mid m$ тогда и только тогда, когда в разложениях* (2.5) *выполняются неравенства $a_i \leqslant b_i$ для всех $1 \leqslant i \leqslant s$.*

Доказательство. В одну сторону очевидно: если $a_i \leqslant b_i$, то число

$$q = p_1^{b_1-a_1} \cdot p_2^{b_2-a_2} \cdot \ldots \cdot p_s^{b_s-a_s}$$

целое и $m = qn$.

В обратную сторону: пусть $m = qn$. Запишем разложения на простые для q и n. Для любого простого множителя p показатель a, с которым p входит в разложение m на простые множители, равен сумме аналогичных показателей для q и для n. Поэтому он не меньше, чем показатель в разложении n. □

Такое разбиение делимости на сравнения показателей по различным простым множителям часто оказывается удобным.

Лемма 2.67. *Если выполняются разложения* (2.5)*, то*

$$\text{НОД}(n,m) = \prod_{i=1}^{s} p_i^{\min(a_i,b_i)}, \qquad \text{НОК}(n,m) = \prod_{i=1}^{s} p_i^{\max(a_i,b_i)}.$$

Доказательство. Доказательство легко следует из утверждения 2.66.

Ясно, что выражения в лемме являются, соответственно, общим делителем и общим кратным чисел n и m. С другой стороны, в любой общий делитель чисел n и m простой множитель p_i входит с показателем, не превышающим a_i и b_i, — иное противоречит утверждению 2.66. Аналогично для любого общего кратного: показатель, с которым входит p_i, не меньше a_i и b_i. □

Основная теорема арифметики позволяет быстро доказывать иррациональность многих чисел. Рассмотрим классический пример $\sqrt{2}$. Обычно в доказательстве иррациональности этого числа используют метод бесконечного спуска. Но прямое применение основной теоремы арифметики работает столь же хорошо.

Пример 2.68. Докажем, что $\sqrt{2}$ иррационально. Рассуждаем от противного. Пусть $\sqrt{2} = m/n$. Тогда $m^2 = 2n^2$. Разложим левую и правую части этого равенства по степеням простых чисел. Показатель двойки в левой части равенства чётный (удвоенный показатель двойки в разложении m), а в правой части равенства — нечётный (удвоенный показатель двойки в разложении n плюс ещё одна единица из-за множителя два). Но эти показатели должны совпадать в силу основной теоремы арифметики. Получили противоречие. Значит, $\sqrt{2}$ иррациональное число. □

Замечание 2.69. Аналогично без труда доказывается иррациональность $\sqrt{3}$, $\sqrt{5}$, $\sqrt[3]{2}$, $\sqrt[3]{4}$ и других чисел такого же вида.

Основная теорема арифметики позволяет продолжить анализ числа решений у квадратичных сравнений.

Пример 2.70 (продолжение примера 2.59)**.** Сколько решений у сравнения

$$x^2 \equiv 1 \pmod{p^n},$$

где $p > 2$ — простое число?

Докажем, что их два, как и для простого модуля. Если $x^2 \equiv 1 \pmod{p^n}$, то $(x-1)(x+1) \equiv 0 \pmod{p^n}$. Значит, сумма показателей степеней, с которыми p входит в разложения $x-1$ и $x+1$ по крайней мере n.

Поскольку $p > 2$, то делится на p может лишь одно из этих чисел $x-1$, $x+1$. Поэтому либо $x \equiv 1 \pmod{p^n}$, либо $x \equiv -1 \pmod{p^n}$. □

Для случая степеней двойки получается другой ответ, как уже было видно в примере 2.63. Нужно уточнить анализ предыдущего примера, так как различающиеся на 2 числа вполне могут быть одновременно чётными.

Пример 2.71 (продолжение примера 2.63)**.** Сколько решений у сравнения

$$x^2 \equiv 1 \pmod{2^n}?$$

Для $n = 1$ решение одно, для $n = 2$ решений 2 (± 1). Докажем, что при $n \geqslant 3$ решений всегда 4.

Если $x^2 \equiv 1 \pmod{2^n}$, то $(x-1)(x+1) \equiv 0 \pmod{2^n}$, то есть

$$x - 1 = 2^k r,\ r \text{ — нечётное, и } x + 1 = 2^\ell s,\ s \text{ — нечётное, а } k + \ell \geqslant n.$$

Поскольку $2^\ell s - 2^k r = 2$, хотя бы один из показателей k, ℓ равен 1. Рассмотрим оба случая.

$\ell = 1$. Тогда $s - 2^{k-1} r = 1$, $k \geqslant n-1$. В этом случае

$$x = 2s - 1 = 2^k r + 1.$$

Остатки при делении $2^{n-1} r + 1$ на 2^n принимают два возможных значения 1 для чётных r ($k \geqslant n$) и $2^{n-1} + 1$ для нечётных r ($k = n-1$). Получаем два решения сравнения $x = 1$ и $x = 2^{n-1} + 1$ для этого случая.

Аналогично рассматриваем второй случай $k = 1$. Теперь

$$x = 2r + 1 = 2^\ell s - 1, \qquad \ell \geqslant n - 1.$$

Остатки при делении $2^{n-1} s - 1$ на 2^n принимают два возможных значения 1 для чётных s ($\ell \geqslant n$) и $2^{n-1} - 1$ для нечётных s ($k = n-1$). Получаем два решения сравнения $x = -1$ и $x = 2^{n-1} - 1$ для этого случая.

Всего получается четыре решения. □

Контрольный вопрос 2.72. Почему этот анализ не проходит для модулей 2, 4?

2.6 Оценка количества простых чисел

У основной теоремы арифметики необъятное количество приложений. Мы отдельно остановимся на одном из важнейших. Как уже говорилось выше (замечание 1.57), количество простых чисел, не превосходящих x, обозначаемое через $\pi(x)$, имеет точную асимптотическую оценку

$$\pi(x) \sim \frac{x}{\ln x}\ .$$

Доказательство этой оценки довольно трудное. Намного проще доказать более грубые асимптотические оценки — с точностью до мультипликативной константы. Впервые эти оценки были получены Чебышёвым, поэтому мы называем их оценками Чебышёва.

Теорема 2.73 (оценка Чебышёва)**.** *Существуют такие числа C_1 и C_2, что для всех достаточно больших x выполняются неравенства*

$$C_1 \frac{x}{\ln x} \leqslant \pi(x) \leqslant C_2 \frac{x}{\ln x} \ .$$

Доказательства этих оценок хорошо известны. Мы следуем изложению в [16, с. 25–26].

Перед доказательством этой теоремы докажем пару теоретико-числовых фактов, которые интересны и сами по себе.

Начнём с того, что найдём разложение факториалов на простые множители.

Лемма 2.74. *Пусть p — простое число. Наибольшая степень p, которая делит $n!$, равна*

$$\left\lfloor \frac{n}{p} \right\rfloor + \left\lfloor \frac{n}{p^2} \right\rfloor + \cdots + \left\lfloor \frac{n}{p^k} \right\rfloor + \ldots$$

Хотя формула содержит бесконечное количество слагаемых, для любого n лишь конечное количество слагаемых отлично от нуля: $n < p^k$ для достаточно больших k, и тогда $\lfloor n/p^k \rfloor = 0$.

Доказательство. Как известно, $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$. Каждый множитель kp, $1 \leqslant k \leqslant \lfloor n/p \rfloor$ делится на p. Поэтому степень p в разложении $n!$ по степеням простых не меньше $\lfloor n/p \rfloor$.

Однако каждый множитель kp^2, $1 \leqslant k \leqslant \lfloor n/p^2 \rfloor$ делится на p^2. Это даёт прибавку к степени $\lfloor n/p^2 \rfloor$.

Продолжая это рассуждение, получаем формулу из условия леммы. □

Формулу из леммы 2.74 можно записать в конечном виде, воспользовавшись представлением n в p-ичной системе счисления. А именно, пусть

$$n = n_t p^t + n^{t-1} + \cdots + n_1 p + n_0 = \overline{(n_t n_{t-1} \ldots n_1 n_0)}_p, \quad 0 \leqslant n_i < p.$$

Тогда наибольшая степень p в разложении $n!$ по степеням простых равна

$$\begin{aligned} &\overline{(n_t n_{t-1} \ldots n_2 n_1)}_p + \\ &\overline{(n_t n_{t-1} \ldots n_1)}_p + \\ &\ldots + \\ &\overline{(n_t n_{t-1})}_p + \\ &\overline{(n_t)}_p, \end{aligned} \tag{2.6}$$

так как k-е сверху слагаемое равно $\lfloor n/p^k \rfloor$ (запись в p-ичной системе счисления).

Напомним выражение для биномиального коэффициента через факториалы

$$\binom{a+b}{a} = \frac{(a+b)!}{a!b!}\ .$$

Наша следующая цель — найти наибольшую степень p, на которую делится биномиальный коэффициент. Это вариант теоремы Люка, которая выражает через цифры p-ичных представлений остаток при делении биномиального коэффициента на p.

Теорема 2.75 (теорема Люка)**.** *Пусть p — простое число. Наибольшая степень p, которая делит $\binom{a+b}{a}$, равна количеству переносов при сложении чисел a и b, записанных в p-ичной системе счисления.*

Доказательство. Обозначим $c = a + b$. По лемме 2.74 искомая степень равна

$$\begin{pmatrix} \overline{(c_t c_{t-1} \dots c_2 c_1)}_p + \\ \overline{(c_t c_{t-1} \dots c_1)}_p + \\ \dots + \\ \overline{(c_t c_{t-1})}_p + \\ \overline{(c_t)}_p \end{pmatrix} - \begin{pmatrix} \overline{(a_t a_{t-1} \dots a_2 a_1)}_p + \\ \overline{(a_t a_{t-1} \dots a_1)}_p + \\ \dots + \\ \overline{(a_t a_{t-1})}_p + \\ \overline{(a_t)}_p \end{pmatrix} - \begin{pmatrix} \overline{(b_t b_{t-1} \dots b_2 b_1)}_p + \\ \overline{(b_t b_{t-1} \dots b_1)}_p + \\ \dots + \\ \overline{(b_t b_{t-1})}_p + \\ \overline{(b_t)}_p \end{pmatrix} \tag{2.7}$$

Сравним числа в первой строке формулы (2.7), то есть

$$\overline{(a_t a_{t-1} \dots a_2 a_1)}_p + \overline{(b_t b_{t-1} \dots b_2 b_1)}_p \quad \text{и} \quad \overline{(c_t c_{t-1} \dots c_2 c_1)}_p$$

Если при сложении a и b в p-ичной системе счисления переноса в первый разряд не было, эти числа равны. Если перенос был, то второе на единицу больше первого (при сложении в следующий разряд переносится не больше 1).

Аналогично сравниваются и остальные слагаемые в (2.7): в k-й строчке получается 0, если не было переноса в k-й разряд при сложении a и b в p-ичной системе счисления; и получается 1, если перенос был.

Таким образом, значение формулы (2.7) равно общему количеству переносов при сложении a и b в p-ичной системе счисления. □

Теперь мы готовы к доказательству оценок Чебышёва.

Доказательство теоремы 2.73. Будем доказывать оценки для чётных чисел вида $2n$. Этого достаточно, так как $\pi(2n+1) - \pi(2n)$ не больше 1.

Пусть p^k — наибольшая степень простого числа p, которая делит $\binom{2n}{n}$. По теореме Люка число k равно количеству переносов при сложении числа n с числом n в p-ичной системе счисления. Докажем, что $p^k \leqslant 2n$.

В самом деле, если $n = \overline{(n_t n_{t-1} \dots n_2 n_1)}_p$, $n_t > 0$, количество переносов не больше $t+1$ (всего $t+1$ разряд используется). Если количество переносов меньше $t+1$, то $p^k \leqslant n$. Если же количество переносов равно $t+1$, то докажем, что число n не меньше $p^{t+1}/2$. В этом случае $n_0 + n_0 \geqslant p$, а $n_i + n_i + 1 \geqslant p$. Поэтому

$$2n \geqslant (p-1)p^t + (p-1)p^{t-2} + \dots + p = p^{t+1} = p^k.$$

Формула бинома даёт равенство

$$(1+1)^{2n} = \binom{2n}{0} + \binom{2n}{1} + \cdots + \binom{2n}{n} + \cdots + \binom{2n}{2n-1} + \binom{2n}{2n},$$

причём самый большой биномиальный коэффициент — средний.

Упражнение 2.76. Докажите это последнее утверждение.

Таким образом, получаем оценку

$$\binom{2n}{n} \geqslant \frac{2^{2n}}{2n+1}\,.$$

Но, как мы выяснили, вклад в биномиальный коэффициент каждой степени простого не превосходит $2n$. Поэтому

$$(2n)^{\pi(2n)} \geqslant \binom{2n}{n} \geqslant \frac{2^{2n}}{2n+1},$$

из которой следует нижняя оценка Чебышёва при любом $C_1 < \ln 2$.

Для доказательства верхней оценки Чебышёва оценим средний биномиальный коэффициент сверху как 2^{2n} (см. формулу бинома выше). С другой стороны заметим, что средний биномиальный коэффициент делится на любое простое p, для которого $n < p < 2n$ (поскольку такое простое делит $(2n)!$, но не делит $n!$). Поэтому

$$n^{\pi(2n)-\pi(n)} < \prod_{n<p<2n} p < \binom{2n}{n} < 2^{2n},$$

то есть

$$\pi(2n) - \pi(n) < \frac{2n}{\log_2 n}\,. \tag{2.8}$$

Складывая несколько неравенств вида (2.8) для $2n = 2^t, 2^{t-1}, \ldots, 2^2$, получаем

$$\pi(2^t) - \pi(2) < \frac{2^t}{t-1} + \frac{2^{t-1}}{t-2} + \cdots + \frac{4}{1}\,.$$

Так как при $t \geqslant 4$ выполняется

$$\frac{2^{t-1}}{t-2} \leqslant \frac{3}{4} \cdot \frac{2^t}{t-1},$$

оценивая сумму геометрической прогрессией, получаем оценку

$$\pi(2^t) < 3 + 4 \cdot \frac{2^t}{t-1}\,.$$

Чтобы продолжить эту оценку на произвольное x, возьмём наименьшую степень 2, не меньшую x:

$$2^{t-1} < x \leqslant 2^t.$$

Функция $\pi(x)$ неубывающая, поэтому

$$\pi(x) \leqslant \pi(2^t) < 3 + 4 \cdot \frac{2^t}{t-1} < 3 + 8 \cdot \frac{x}{\log_2 x - 1}$$

и верхняя оценка в теореме выполняется при всех $C_2 > 8 \ln 2$. □

2.7 Порядки перестановок

В этом разделе рассмотрим группу перестановок S_n и научимся определять порядки перестановок. Общее определение порядка элемента в группе для перестановок пересказывается так: порядок перестановки π — это такое наименьшее положительное k, что перестановка

$$\pi^k = \underbrace{\pi \circ \pi \circ \cdots \circ \pi}_{k \text{ композиций}}$$

является тождественной (тождественная перестановка — это нейтральный элемент в группе S_n).

Начнём с вычисления порядка цикла.

Утверждение 2.77. *Порядок цикла* $(a_1\ a_2\ \ldots\ a_\ell)$ *равен* ℓ.

Доказательство. Обозначим $\pi = (a_1\ a_2\ \ldots\ a_\ell)$. Если $\pi^k = \mathrm{id}$, то число a_1 переводится перестановкой π^k в a_1. Это возможно лишь когда k кратно ℓ. Наименьшее положительное число, кратное ℓ, это как раз ℓ.

Рассуждение остаётся справедливым и для всех остальных чисел цикла. Поэтому $\pi^\ell = \mathrm{id}$ и это в точности порядок цикла. □

Теперь уже несложно найти порядок произвольной перестановки, воспользовавшись цикловым разложением.

Лемма 2.78. *Пусть цикловое разложение перестановки* π *состоит из циклов длин* $\ell_1, \ell_2, \ldots, \ell_s$. *Тогда порядок перестановки* π *равен* $\mathrm{НОК}(\ell_1, \ell_2, \ldots, \ell_s)$.

Доказательство. Если $\pi^k = \mathrm{id}$, то в каждом цикле все числа остаются на своих местах. Для цикла длины ℓ_1 это возможно при k кратном ℓ_1; для цикла длины ℓ_2 — при k кратном ℓ_2 и т.д. Поэтому k должно быть общим кратным длин циклов. Значит, порядок перестановки — наименьшее общее кратное длин циклов. □

Пример 2.79. Порядок перестановки $(1\ 2)(3\ 4\ 5)(6\ 7\ 8\ 9)$ равен $\mathrm{НОК}(2, 3, 4) = 12$. □

Теперь уже нетрудно привести пример, который показывает, что обращение следствия 2.48 из теоремы Лагранжа для порядков элементов неверно.

Пример 2.80. Рассмотрим группу S_6 перестановок 6 элементов, в ней $6! = 720 = 16 \cdot 45$ элементов. Поэтому $8 \mid 6! = |S_6|$.

Докажем однако, что в группе S_6 нет перестановок порядка 8. Действительно, чтобы наименьшее общее кратное набора чисел равнялось 8, в этом наборе должно быть число 8: максимальная степень 2, на которую делится НОК совпадает с максимальной степенью 2, на которую делится хотя бы одно число из набора.

Но цикла длины 8 в перестановке 6 чисел быть не может, так как числа в цикле не повторяются. □

Пример 2.81. Тем не менее, подгруппа порядка 8 в группе S_6 существует (из предыдущего следует, что эта подгруппа не циклическая).

Рассмотрим три транспозиции $\alpha_1 = (1\ 2)$, $\alpha_2 = (3\ 4)$, $\alpha_3 = (5\ 6)$. Так как эти циклы не пересекаются, перестановки α_i попарно коммутируют. Обозначим $G = \langle \alpha_1, \alpha_2, \alpha_3 \rangle$. Докажем, что $|G| = 8$.

В примере 2.28 мы доказали, что группа G состоит из всех возможных произведений α_i и их обратных. Поскольку порядки α_i равны 2, они совпадают со своими обратными. А так как они коммутируют, то после перестановки сомножителей и сокращения пар одинаковых сомножителей получаем, что любое произведение α_i равно произведению вида $\alpha_1^{k_1}\alpha_2^{k_2}\alpha_3^{k_3}$, где $k_i \in \{0, 1\}$. Все такие произведения различны.

Действительно, $\alpha_1^{k_1}\alpha_2^{k_2}\alpha_3^{k_3} = \mathrm{id}$ тогда и только тогда, когда $k_1 = k_2 = k_3 = 0$. Равенство $\alpha_1^{k_1}\alpha_2^{k_2}\alpha_3^{k_3} = \alpha_1^{t_1}\alpha_2^{t_2}\alpha_3^{t_3}$ равносильно равенству $\alpha_1^{k_1-t_1}\alpha_2^{k_2-t_2}\alpha_3^{k_3-t_3} = \mathrm{id}$ (второе получается из первого умножением обеих частей равенства на перестановку $\alpha_1^{-t_1}\alpha_2^{-t_2}\alpha_3^{-t_3}$). Поэтому равенство произведений возможно лишь при совпадении всех показателей в произведениях. □

Замечание 2.82. Из теоремы Силова (см., например, учебник Э. Б. Винберга [5]) следует, что в S_6 есть даже подгруппа порядка 16. Найти такую подгруппу сложнее. Подумав над конструкцией из предыдущего примера, можно понять, что достаточно построить подгруппу порядка 8 в S_4 (то есть, в группе перестановок элементов $\{1, 2, 3, 4\}$): тогда перестановки этой подгруппы и транспозиция (5 6) порождают подгруппу порядка 16.

Подгруппу порядка 8 в группе S_4 искать подбором довольно трудно. Ниже конструкция такой группы появится как следствие более общих рассуждений (см. пример 6.8).

3 Изоморфизмы групп

3.1 Определение и примеры

Если интересоваться только свойствами групповой операции, то многие группы оказываются «одинаковыми», общепринятый термин — изоморфными.

Определение 3.1. *Изоморфизм групп* $(G, *)$ и $(G', \circ)$ — это отображение $\varphi\colon G \to G'$, которое (1) биективно; (2) сохраняет операцию, то есть для любых элементов a, b группы G выполняется равенство

$$\varphi(a * b) = \varphi(a) \circ \varphi(b)$$

(образ произведения равен произведению образов).

Группы называются *изоморфными*, если между ними существует изоморфизм. Обозначение $G \cong G'$.

Если группы изоморфны, то с алгебраической точки зрения между ними нет различий: любое свойство группы, которое можно выразить, используя групповую операцию, выполняется или не выполняется в обеих группах одновременно. У нас уже появлялись примеры изоморфных групп, хотя мы пока не обращали на это внимания.

Пример 3.2. Аддитивная группа вычетов Z_n изоморфна группе U_n корней из единицы. Напомним, что элементы U_n — это комплексные корни уравнения $z^n = 1$. Как мы уже выяснили в разделе 1.2.5, их ровно n штук (лемма 1.28), то есть столько же, сколько вычетов по модулю n.

Операция в U_n — умножение комплексных чисел. Корни из единицы замкнуты относительно умножения, поэтому по лемме 2.30 они образуют подгруппу мультипликативной группы $\mathbb{C}^*$ (ненулевые комплексные числа с операцией умножения).

В доказательстве леммы 1.28 мы также нашли явное представление для корней n-й степени из единицы: все они имеют модуль 1, а их аргументы имеют вид $2\pi k/n$, $k \in \mathbb{Z}$.

Отсюда следует изоморфизм между U_n и Z_n:

$$\varphi\colon (\cos\frac{2\pi k}{n} + i\sin\frac{2\pi k}{n}) \mapsto [k],$$

то есть корню из единицы с аргументом $2\pi k/n$ соответствует вычет по модулю n, который содержит остаток при делении k на n.

Это соответствие взаимно однозначно. Осталось проверить второе свойство изоморфизма — сохранение операции. Оно легко следует из геометрической интерпретации умножения комплексных чисел. При умножении аргументы складываются, причём комплексные числа равны, если они равны по модулю, а их аргументы различаются на целое кратное 2π.

Важный частный случай этого изоморфизма $n = 2$. В этом случае группы состоят из двух элементов и изоморфизм имеет вид

$$+1 \mapsto 0, \qquad -1 \mapsto 1.$$

Эта группа из двух элементов часто возникает в комбинаторике и теоретической информатике. Чтобы подчеркнуть связь с булевой логикой, элементы этих групп обычно называются булевыми значениями (аддитивный или мультипликативный комплект, в зависимости от выбора U_2 или Z_2),

В современном анализе булевых функций мультипликативный комплект используется очень часто, когда свойства функций выражаются с помощью преобразования Фурье. □

Совпадение всех алгебраических свойств изоморфных групп имеет много конкретных следствий. Приведём часть из них в следующем утверждении.

Утверждение 3.3. *Пусть $\varphi\colon G \to G'$ — изоморфизм групп $(G, *)$ и $(G', \circ)$. Тогда*

1. *$\varphi(e)$ — нейтральный элемент группы G' (изоморфизм сохраняет единицу);*
2. *$\varphi(a^{-1}) = \varphi(a)^{-1}$ (образ обратного элемента — обратный элемент к образу);*
3. *обратное отображение φ^{-1} является изоморфизмом;*
4. *композиция изоморфизмов является изоморфизмом.*

Проверка всех этих свойств рутинная, приведём её для полноты изложения.

Доказательство. П. 1: для любого $a \in G$ имеем $a * e = e * a = a$. Поскольку изоморфизм сохраняет операцию, получаем равенства

$$\varphi(a) \circ \varphi(e) = \varphi(a * e) = \varphi(a) = \varphi(e * a) = \varphi(e) \circ \varphi(a),$$

которые означают, что $e' = \varphi(e)$ является нейтральным элементом в группе G'.

П. 2: равенство

$$\varphi(a) \circ \varphi(a^{-1}) = \varphi(a * a^{-1}) = \varphi(e) = e',$$

означает, что обратный к образу a элемент есть образ обратного: $\varphi(a)^{-1} = \varphi(a^{-1})$.

П. 3: обратное к биекции отображения также является биекцией. Сохранение операции проверяется прямым вычислением:

$$\begin{aligned}\varphi^{-1}(a \circ b) &= \varphi^{-1}\big(\varphi(\varphi^{-1}(a)) \circ \varphi(\varphi^{-1}(b))\big) = \\ &= \varphi^{-1}\big(\varphi(\varphi^{-1}(a) * \varphi^{-1}(b))\big) = \big(\varphi^{-1}\varphi\big)(\varphi^{-1}(a) * \varphi^{-1}(b)) = \varphi^{-1}(a) * \varphi^{-1}(b).\end{aligned}$$

П. 4: пусть $\varphi\colon G_1 \to G_2$ и $\psi\colon G_2 \to G_3$ — изоморфизмы. Композиция $\psi \circ \varphi$ биекций является биекцией. Сохранение операции проверяется прямым вычислением

$$\psi(\varphi(a \cdot b)) = \psi(\varphi(a) \cdot \varphi(b)) = \psi(\varphi(a)) \cdot \psi(\varphi(b)).$$

Здесь для простоты записи операции во всех трёх группах обозначены одинаково. □

Отсюда получаем, что у конечных изоморфных групп одинаковы порядки (так как биекции существуют только между множествами с одинаковым количеством элементов), одинаково количество элементов заданного порядка (так как изоморфизм сохраняет порядок элемента). Список таких свойств легко продолжить.

Контрольный вопрос 3.4. Докажите, что у конечных изоморфных групп одинаково количество подгрупп заданного порядка.

Пример 3.5. Рассмотрим две подгруппы группы перестановок S_6. Первая подгруппа G_1 описана в примере 2.81. Она порождена транспозициями $\alpha_1 = (1\ 2)$, $\alpha_2 = (3\ 4)$, $\alpha_3 = (5\ 6)$. Её порядок, как уже проверено в том примере, равен 8.

Вторая подгруппа G_2 порождена перестановками $\beta_1 = (1\ 2\ 3\ 4)$ и $\beta_2 = (5\ 6)$. Её порядок, как нетрудно видеть, также равен 8. Аналогично анализу в примере 2.81 проверяется, что элементами этой подгруппы являются перестановки вида $\beta_1^{k_1}\beta_2^{k_2}$, где $k_1 \in \{0, 1, 2, 3\}$, $k_2 \in \{0, 1\}$.

Порядки этих групп одинаковы. Поэтому какие-то биекции между элементами этих групп существуют. Однако эти группы неизоморфны, то есть любая биекция нарушает свойство «образ произведения равен произведению образов» хотя бы для одной пары элементов. Прямая проверка требует перебора всех 8! биекций, причём для каждой ещё нужно найти пару, нарушающую свойство изоморфизма.

Однако доказательство неизоморфности этих групп намного проще. В первой группе все элементы имеют порядок 2 или 1. А во второй есть элемент порядка 4: это перестановка β_1. Поскольку изоморфизм сохраняет порядок элемента, эти группы неизоморфны. □

Приведём ещё несколько примеров изоморфизма групп.

Пример 3.6. Изоморфны ли группы $(\mathbb{R}, +)$ действительных чисел по сложению и $(\mathbb{R}_+, \cdot)$ положительных действительных чисел по умножению?

Контрольный вопрос 3.7. Проверьте, что $(\mathbb{R}_+, \cdot)$ действительно группа.

Ответ: да, эти группы изоморфны. Изоморфизм задаётся, например, отображением

$$\exp\colon x \mapsto e^x.$$

Обратное отображение — это логарифмическая функция, как известно из анализа. Сохранение операции — это одно из основных свойств показательной функции $\exp(x + y) = \exp(x)\exp(y)$. □

Мы ещё не раз увидим разницу между алгебраическими свойствами числовых систем. Проявляется она и в этом случае.

Пример 3.8. Изоморфны ли группы $(\mathbb{Q}, +)$ рациональных чисел по сложению и $(\mathbb{Q}_+, \cdot)$ положительных рациональных чисел по умножению?

Замечание 3.9. Попытки перенести показательную функцию на множество рациональных чисел к успеху не приводят.

Число e — иррациональное и даже трансцендентное. Доказательство иррациональности не так уж сложно, но относится скорее к анализу, чем к алгебре, мы его пропустим.

Попытка изменить основание показательной функции также не приведёт к успеху. Скажем, если взять основание 2, то проблема будет с иррациональностью $\sqrt{2}$ (см. пример 2.68).

Прямой анализ общего случая (произвольное основание) представляется довольно трудным делом, но отрицательный ответ сохраняется, как мы сейчас увидим.

Для бесконечных групп перебрать все возможные биекции вообще невозможно. Поэтому, если не удаётся построить изоморфизм, нужно искать какое-то алгебраическое свойство, различающее эти группы.

В данном случае нужное свойство легко обнаружить. В группе $(\mathbb{Q}, +)$ из любого числа извлекается «корень n-й степени». Поскольку в этой группе используется аддитивная запись, это означает, что уравнение $nx = a$ имеет решения при любом $a \in \mathbb{Q}$, $n \in \mathbb{Z}_+$.

Для группы $(\mathbb{Q}_+, \cdot)$ это не так. Уравнение $x^2 = 2$ решений в положительных рациональных числах не имеет.

Предположим, что существует изоморфизм $\varphi\colon (\mathbb{Q}, +) \to (\mathbb{Q}_+, \cdot)$ между этими группами. Обозначим $a = \varphi^{-1}(2)$, а $b = a/2$. Тогда $\varphi(b)$ является корнем уравнения $x^2 = 2$, что легко проверить, используя сохранение операции:

$$\varphi(b)\varphi(b) = \varphi(b + b) = \varphi(a) = \varphi(\varphi^{-1}(2)) = 2. \qquad \square$$

Пример 3.10. Рассмотрим пример изоморфизма неабелевых групп. Докажем, что $D_3 \cong S_3$. Каждый элемент группы симметрий треугольника переводит треугольник в себя. Значит, вершины треугольника переходят в вершины. Пронумеруем вершины треугольника числами 1, 2, 3 и сопоставим элементу $g \in D_3$ перестановку $v(g) \in S_3$ чисел, которая задается перестановкой соответствующих вершин треугольника. Например, $v(e) = ()$ — каждая вершина остаётся на месте. Из построения ясно, что композиции элементов D_3 соответствует композиция соответствующих перестановок. С другой стороны, образы трёх точек однозначно определяют движение плоскости. Поэтому разным элементам D_3 соответствуют разные перестановки. Поскольку $|D_3| = |S_3| = 6$, указанное выше отображение $v\colon D_3 \to S_3$ — изоморфизм. $\square$

Пример 3.11. Среди правильных многогранников есть двойственные: куб двойственен октаэдру, а додекаэдр — икосаэдру. В частности, центры граней куба являются вершинами октаэдра, а центры граней додекаэдра — вершинами икосаэдра. Поскольку центры граней при вращении переходят в центры граней, получаем из этого наблюдения изоморфизм группы куба и группы октаэдра, а также группы додекаэдра и группы икосаэдра.

Именно в этом смысле есть только три разных группы многогранников. Группа октаэдра с алгебраической точки зрения совпадает с группой куба, а группа икосаэдра — с группой додекаэдра. $\square$

Пример 3.12. Пусть $(G, \cdot)$ — группа. Определим на том же множестве другую операцию $*$ по правилу $x * y = y \cdot x$. Можно проверить (сделайте это упражнение), что эта операция также задаёт структуру группы. Будем называть эту группу «транспонированной».

Как соотносятся между собой группы $(G, \cdot)$ и $(G, *)$? Если группа $(G, \cdot)$ абелева, то эти группы попросту совпадают. Для неабелевой группы $(G, \cdot)$ порядок сомножителей в произведении существенен, в общем случае $x \cdot y \neq y \cdot x$.

Тем не менее кажется очевидным, что никакой существенной разницы между этими группами нет. Например, если бы мы определили умножение перестановок в другом порядке (сначала применяется левая, а потом — правая), большинство фактов о перестановках, которые мы доказывали и будем доказывать (порядки перестановок, чётность перестановок и т.п.), останутся верными и для «транспонированной» группы, а в некоторых произойдут несущественные изменения (левые смежные классы станут правыми и наоборот).

Это не случайно: группа изоморфна своей транспонированной. Изоморфизм задаётся отображением $g \mapsto g^{-1}$. Действительно, это отображение взаимно однозначно (обратный к обратному к элементу g — это сам g). Сохранение операции означает выполнение уже известного нам тождества в группе:

$$(g \cdot h)^{-1} = h^{-1} \cdot g^{-1} = h^{-1} * g^{-1}.$$

Таким образом, если по какой-то причине необходимо рассматривать групповую операцию в обратном порядке записи, то свойства полученной группы точно такие же, как исходной. Однако изоморфизм задаётся нетождественным отображением, что бывает существенно в конкретных вычислениях. □

Определение 3.13. Изоморфизм группы с самой собой называется *автоморфизмом*.

Тривиальный пример автоморфизма — тождественное отображение, однако автоморфизмов групп может быть гораздо больше.

Пример 3.14. Любая перестановка π из группы перестановок S_n задаёт отображение $\sigma \mapsto \sigma^\pi = \pi \circ \sigma \circ \pi^{-1}$ группы S_n на себя. Проверим, что это отображение является изоморфизмом.

Биективность. Обратное отображение задаётся формулой $\sigma \mapsto \pi^{-1} \circ \sigma \circ \pi$, что проверяется прямым вычислением

$$\pi^{-1} \circ (\pi \circ \sigma \circ \pi^{-1}) \circ \pi = (\pi^{-1} \circ \pi) \circ \sigma \circ (\pi^{-1} \circ \pi) = \sigma.$$

Сохранение операции также проверяется прямым вычислением

$$(\sigma_1 \circ \sigma_2)^\pi = \pi \circ (\sigma_1 \circ \sigma_2) \circ \pi^{-1} = \pi \circ \sigma_1 \circ \pi^{-1} \circ \pi \circ \sigma_2 \circ \pi^{-1} = \sigma_1^\pi \circ \sigma_2^\pi.$$

Этот автоморфизм группы S_n называется *сопряжением* посредством перестановки π. Ниже мы подробнее рассмотрим свойства сопряжения в общем случае. □

Автоморфизмы — это «симметрии» групп. Как и симметрии многогранников, да и любых других математических объектов, автоморфизмы образуют группу относительно композиции.

Из утверждения 3.3 получаем следствие.

Утверждение 3.15. *Автоморфизмы любой группы G образуют относительно композиции группу, которая называется* группой автоморфизмов *группы G.*

Группу автоморфизмов группы G будем обозначать $\operatorname{Aut} G$.

Из уже разобранных примеров групп проще всего описать группу автоморфизмов аддитивной группы вычетов $\mathbb{Z}_n$.

Лемма 3.16. $\operatorname{Aut} Z_n = Z_n^*$.

Доказательство. Рассмотрим автоморфизм $\varphi\colon Z_n \to Z_n$. Обозначим $\varphi(1) = a$. Тогда свойство сохранения операции позволяет выразить образ любого элемента группы через a:

$$\varphi(k) = \varphi(\underbrace{1+1+\cdots+1}_{k \text{ раз}}) = \underbrace{\varphi(1)+\cdots+\varphi(1)}_{k \text{ раз}} = ka.$$

Таким образом, все возможные автоморфизмы группы Z_n имеют вид $x \mapsto ax$. Однако не все a дают автоморфизм группы Z_n. Ясно, что для существования обратного отображения у 1 должен быть прообраз, то есть у вычета $[a]$ должен быть обратный. Это требование равносильно тому, что сравнение $ax \equiv 1 \pmod n$ имеет решение. Как известно из теоремы 1.48, решение такого сравнения существует тогда и только тогда, когда a взаимно просто с n.

Значит, чтобы отображение $x \mapsto ax$ было автоморфизмом аддитивной группы вычетов, необходима обратимость вычета $[a]$, то есть взаимная простота a и n. Проверим, что это условие является также и достаточным. Пусть $ab \equiv 1 \pmod n$. Тогда обратное к $x \mapsto ax$ отображение задаётся формулой $x \mapsto bx$, так как $bax \equiv x \pmod n$ для любого x.

Сохранение групповой операции в данном случае — это свойство дистрибутивности $a(x+y) = ax + ay$.

Осталось заметить, что композиция отображений $\varphi(x) = ax$ и $\psi(x) = bx$ имеет вид $\psi(\varphi(x)) = (ba)x$. То есть композициям автоморфизмов соответствует умножение по модулю n коэффициентов, задающих эти автоморфизмы. □

Конструкцию из леммы 3.16 можно обобщить на произвольную абелеву группу. Будем, как и в доказательстве леммы 3.16, записывать групповую операцию аддитивно (как сложение чисел). Тогда отображение $\varphi\colon x \mapsto ax$ — это возведение в степень. Это отображение всегда сохраняет операцию:

$$\varphi(x+y) = \underbrace{(x+y)+\cdots+(x+y)}_{a \text{ раз}} = \underbrace{x+\cdots+x}_{a \text{ раз}} + \underbrace{y+\cdots+y}_{a \text{ раз}} = \varphi(x)+\varphi(y)$$

(во втором равенстве использована коммутативность групповой операции).

Контрольный вопрос 3.17. Как будет выглядеть это вычисление, если записывать групповую операцию мультипликативно (как умножение чисел)?

Но, как и в доказательстве леммы 3.16, такое отображение не обязательно биективно. Для биективности достаточна взаимна простота a и порядка группы.

Утверждение 3.18. *Пусть G — абелева группа порядка n, групповая операция записывается аддитивно. Если* $\text{НОД}(a, n) = 1$*, то отображение $\varphi\colon x \mapsto ax$ биективно и является автоморфизмом группы G.*

Доказательство. Докажем инъективность отображения и воспользуемся тем, что инъективное отображение конечного множества в себя обязательно биективно.

Предположим $ax_1 = ax_2$. Это равносильно $a(x_1 - x_2) = 0$ (напомним, что $x - y$ в аддитивной записи — это то же самое, что xy^{-1} в мультипликативной). Обозначим $z = x_1 - x_2$. Поскольку $az = 0$, порядок z обязан делить a. По теореме Лагранжа порядок z также делит порядок группы G. Поскольку a и порядок группы G взаимно просты, то единственное возможное значение порядка z равно 1. То есть $z = 0$ (нейтральному элементу группы), так как нейтральный элемент в группе — это единственный элемент порядка 1. Но если $z = x_1 - x_2 = 0$, то $x_1 = x_2$, что и означает инъективность отображения φ. □

Замечание 3.19. На самом деле, взаимная простота a и порядка абелевой группы также необходима для биективности отображения $\varphi\colon x \mapsto ax$, то есть является критерием того, что возведение в степень a — автоморфизм группы. Мы докажем необходимость этого условия позже, когда найдём подходящую характеризацию всех конечных абелевых групп. (См. ниже лемму 5.23.) Но для этого потребуются дополнительные соображения и конструкции, которые мы обсуждаем ниже.

3.2 Циклические группы

До сих пор мы определяли группы «явно», указывая множество и операцию на нём. Есть, однако, и другие способы задавать группы. В этом разделе мы разберём важный класс групп, которые имеют абстрактное описание, не привязанное к конкретному множеству. Окажется, впрочем, что эти группы уже появлялись у нас в виде явных конструкций.

Определение 3.20. Группа называется *циклической*, если она порождена одним элементом: $G = \langle a \rangle$, $a \in G$. Любой элемент с таким свойством называется *порождающим* циклической группы.

Замечание 3.21. Для конечных групп порядок элемента a по определению равен порядку группы $\langle a \rangle$, порождённой этим элементом. Поэтому определение циклической группы в случае конечных групп пересказывается равносильным образом так: группа циклическая, если в ней есть элемент, порядок которого совпадает с порядком группы.

Определение циклической группы фиксирует лишь свойства группы и ничего не говорит о том, на каком множестве эта группа реализована. Оказывается, с точностью до изоморфизма все циклические группы легко классифицировать. Более того, многие свойства циклических групп мы уже получили при анализе этих конкретных представлений. В этом разделе мы в основном пересказываем общими словами результаты, которые были получены ранее.

Теорема 3.22. *С точностью до изоморфизма есть ровно одна циклическая группа бесконечного порядка и она изоморфна* $(\mathbb{Z}, +)$.

Для каждого целого положительного n также есть ровно одна (с точностью до изоморфизма) циклическая группа порядка n и она изоморфна Z_n.

Доказательство. Легко убедиться, что названные в теоремы группы циклические: каждая порождена элементом 1 (суммами $+1$ и -1 получается любое целое число, значит, и любой вычет по любому модулю).

В обратную сторону: пусть группа $G = \langle a \rangle$ циклическая. Это означает (см. анализ в примере 2.28), что любой элемент такой группы представляется в виде произведения a и a^{-1}. Так как степени a и степени a^{-1} коммутируют, это произведение упрощается до степени a^k, $k \in \mathbb{Z}$.

Возможны два случая: (1) порядок a бесконечный; (2) порядок a конечный, обозначим его n.

В первом случае все степени a различны. Построим изоморфизм G с $(\mathbb{Z}, +)$ по правилу $\varphi\colon a^k \mapsto k$. Взаимная однозначность следует из того, что все степени a различны. Сохранение операции — это частный случай леммы 2.20: $a^k \cdot a^\ell = a^{k+\ell}$, что равносильно $\varphi(a^k \cdot a^\ell) = k + \ell$ (в группе-образе операция обозначается аддитивно).

Во втором случае всё аналогично. Только теперь $a^k = a^\ell$ равносильно тому, что $k \equiv \ell \pmod{n}$ (утверждение 2.46). Поэтому сложение показателей происходит по модулю n. □

Из этой теоремы, замечания 3.21 и теоремы Лагранжа следует исчерпывающее описание всех групп простого порядка (с точностью до изоморфизма).

Утверждение 3.23. *Любая группа простого порядка циклическая. Для любого простого p с точностью до изоморфизма есть ровно группа порядка p.*

Контрольный вопрос 3.24. Докажите эти утверждения.

Циклическую группу n-го порядка мы будем обозначать через C_n. Конечно, это обозначение дублирует обозначение Z_n, как следует из теоремы о циклических группах 3.22. Однако удобно не привязываться к конкретному воплощению группы. Например, уже потому, что порождающих элементов в циклической группе много и все они равноправны с алгебраической точки зрения, а конкретная реализация Z_n выделяет один из порождающих (число 1) из всех прочих.

Утверждение 3.25. *У группы C_n есть $\varphi(n) = |Z_n^*|$ порождающих.*

Доказательство. Как следует из замечания 3.21 и теоремы 3.22, количество порождающих в C_n равно количеству элементов порядка n в аддитивной группе вычетов Z_n. Это количество легко найти из доказательства теоремы об обратных вычетах (теорема 1.48).

Действительно, порядок a равен n тогда и только тогда, когда множество кратных вычета a совпадает со всем множеством вычетов, что равносильно $\text{НОД}(a, n) = 1$. Но количество остатков, взаимно простых с n, в точности равно $\varphi(n)$. □

Мы уже знаем, как определять порядок элемента a в аддитивной группе вычетов Z_n. По лемме 2.55 он равен $n/\,\text{НОД}(a, n)$. На более абстрактном языке циклических групп можно сделать добавление к этому утверждению.

Лемма 3.26. *Для каждого делителя k порядка n циклической группы C_n существует элемент порядка k. Всякая подгруппа циклической группы циклическая. Подгруппа порядка $k \mid n$ в циклической группе C_n единственна.*

Доказательство. Первое утверждение очевидно выполняется для остатка n/k аддитивной группы вычетов Z_n.

Пусть $H < Z_n$ — некоторая подгруппа группы Z_n. Обозначим h минимальный положительный остаток среди элементов подгруппы H. Докажем, что любой остаток $x \in H$ кратен h. Разделим x на h с остатком: $x = qh+r$. Тогда $r = x-qh \in H$, так как подгруппа замкнута относительно групповой операции и обращения (в данном случае это сложение и взятие противоположного вычета). По определению деления с остатком $0 \leqslant r < h$. Так как h — минимальный остаток в подгруппе H, то $r = 0$ и x делится на h. Таким образом, $H = \langle h \rangle$ — циклическая и её порядок равен $n/\operatorname{НОД}(h, n)$ по лемме 2.55.

Из этого рассуждения следует также единственность подгруппы данного порядка: ведь мы показали, что подгруппа однозначно определяется наименьшим положительным остатком, который в неё входит.

Поскольку любая циклическая группа порядка n изоморфна Z_n, те же утверждения верны и для любой циклической группы. □

Из этой леммы получается интересное теоретико-числовое следствие.

Следствие 3.27 (формула суммирования для функции Эйлера)**.** $\sum\limits_{d \mid n} \varphi(d) = n.$

Доказательство. Разложим элементы циклической группы порядка n на кучки соответственно величине порядка элемента. Для каждого делителя $d \mid n$ элементы порядка d порождают (единственную) циклическую подгруппу порядка d, как говорит лемма 3.26. По утверждению 3.25 в кучке, отвечающей порядку d, ровно $\varphi(d)$ элементов.

Суммирование размеров кучек по всем возможным значениям d даёт порядок группы n (ведь каждый элемент группы попадает ровно в одну кучку). □

Для циклических групп легко находится количество решений уравнения $x^k = e$.

Утверждение 3.28. *Количество решений уравнения $x^k = e$ в циклической группе C_n порядка n равно* $\operatorname{НОД}(n, k)$.

Доказательство. В любой абелевой группе решения уравнения $x^k = e$ образуют подгруппу: $(xy)^k = x^k y^k$ в силу коммутативности умножения, то есть решения уравнения замкнуты относительно групповой операции; $\left(x^{-1}\right)^k = x^{-k} = \left(x^k\right)^{-1}$ (лемма 2.20), то есть решения уравнения замкнуты и относительно взятия обратного.

По лемме 3.26 эта подгруппа циклическая. Будем рассматривать C_n как аддитивную группу вычетов Z_n. Для аддитивной группы вычетов мы обозначаем групповую операцию как сложение, поэтому уравнение приобретает вид $kx = 0$. Поскольку

$kn/\operatorname{НОД}(n,k) \equiv 0 \pmod{n}$, остаток $n/\operatorname{НОД}(n,k)$ является решением уравнения, как и все его кратные, которых ровно $\operatorname{НОД}(n,k)$ штук.

Осталось проверить, что других решений нет. Если kx делится на n, то это число является общим кратным k и n. Поэтому оно кратно $\operatorname{НОК}(k,n) = kn/\operatorname{НОД}(n,k)$ (лемма 2.56). Но тогда x кратно $n/\operatorname{НОД}(n,k)$. □

Как проверить, что некоторая группа является циклической? Для конечных групп самый простой способ состоит в том, чтобы найти элемент, порядок которого совпадает с порядком группы. Для бесконечных групп такой способ не проходит. Перед тем, как привести контрпример, докажем аналог леммы 3.26 для бесконечной циклической группы.

Во всякой группе есть подгруппа порядка 1, состоящая из одного нейтрального элемента. Будем называть такую подгруппу тривиальной, а остальные подгруппы — нетривиальными.

Лемма 3.29. *Всякая нетривиальная подгруппа бесконечной циклической группы бесконечная циклическая и имеет конечный индекс. Подгруппа индекса k в бесконечной циклической группе единственна.*

Доказательство. Аналогично лемме о конечных циклических группах.

Пусть $H < \mathbb{Z}$ — некоторая подгруппа группы $\mathbb{Z}$ целых чисел по сложению (это и есть единственная с точностью до изоморфизма бесконечная циклическая группа, см. теорему 3.22).

Обозначим h минимальное положительное число среди элементов подгруппы H. Докажем, что любое число $x \in H$ кратно h. Разделим x на h с остатком: $x = qh + r$. Тогда $r = x - qh \in H$, так как подгруппа замкнута относительно групповой операции и обращения (в данном случае это сложение и взятие противоположного вычета). По определению деления с остатком $0 \leqslant r < h$. Так как h — минимальное число в подгруппе H, то $r = 0$ и x делится на h. Таким образом, $H = \langle h \rangle$ — циклическая. Её индекс конечен и равен h (количество вычетов по модулю h).

Из этого рассуждения следует также единственность подгруппы данного индекса: мы показали, что подгруппа однозначно определяется наименьшим положительным числом, которое в неё входит. □

Пример 3.30. Рассмотрим группу $(\mathbb{Z}^2, +)$, которая состоит из пар целых чисел, с операцией покомпонентного сложения: $(x_1, y_1) + (x_2, y_2) = (x_1 + x_2, y_1 + y_2)$. Проверка аксиом группы в этом случае очень простая (сделайте её!).

В этой группе все элементы кроме нейтрального имеют бесконечный порядок, однако она не является циклической. Действительно, в этой группе есть нетривиальная подгруппа $H = \langle (1,0) \rangle$ бесконечного индекса: если $a \neq b$, то пары $(0,a)$ и $(0,b)$ принадлежат разным смежным классам по подгруппе H. По лемме 2.38 достаточно проверить, что $(0, a-b) \notin H$. Это очевидно, так как у любого кратного $(1,0)$ во второй компоненте стоит 0. □

Лемма 3.29 имеет интересное следствие для наибольших общих делителей.

Рассмотрим подгруппу $G = \langle a_1, a_2, \ldots, a_t \rangle < \mathbb{Z}$, порождённую числами a_1, a_2, $\ldots$, $a_t \in \mathbb{Z}$. Она состоит из целочисленных линейных комбинаций чисел a_i:

$$x \in G \quad \Leftrightarrow \quad x = \sum_i x_i a_i,\ x_i \in \mathbb{Z}.$$

Если подгруппа G нетривиальная (хотя бы одно из a_i отлично от нуля), то по лемме 3.29 она бесконечная циклическая, $G = \langle h \rangle$, $h > 0$. По сделанному выше замечанию порождающий выражается как целочисленная линейная комбинация a_i:

$$h = \sum_i h_i a_i. \tag{3.1}$$

Легко проверить, что порождающий h является НОД чисел a_i. Действительно, любое a_i является кратным h. Если d — общий делитель a_i, то из (3.1) следует, что d является также делителем h.

Получаем следующий результат.

Теорема 3.31. *НОД чисел a_1, a_2, $\ldots$, a_t является целочисленной линейной комбинацией этих чисел.*

Эта теорема имеет полезные обобщения. Мы рассмотрим их во второй части, при изучении колец.

Приведём ещё одно полезное следствие из описания подгрупп конечной группы. Рассмотрим подгруппу $H < G$ и некоторый элемент $g \in G$. Подгруппа $\langle g \rangle$, порождённая g, пересекается с H по подгруппе K (лемма 2.27). Это подгруппа циклической группы $\langle g \rangle$ и потому также циклическая. Её порядок d является делителем как порядка g (обозначим его n), так и порядка H (теорема Лагранжа для групп H и $\langle g \rangle$).

Смежные классы по H, в которых лежат степени g, это в точности $[g^0] = [e] = H$, $[g]$, g^2, $\ldots$, $[g^{n/d-1}]$ (эти степени g лежат в разных смежных классах по подгруппе K).

Получаем такую лемму, которая часто оказывается полезной.

Лемма 3.32. *Пусть порядки подгруппы $H < G$ и элемента $g \in G \setminus H$ взаимно просты или порядок g — простое число (которое может быть делителем порядка H). Тогда все степени g лежат в разных смежных классах по подгруппе H.*

Доказательство. К предыдущему рассуждению осталось добавить, что в условиях леммы $d = 1$: это общий делитель порядка подгруппы и порядка элемента, причём он не равен порядку элемента, так как $g \in G \setminus H$. □

3.3 Прямые произведения групп

В примере 3.30 неявно появилась важная общая конструкция групп: прямое произведение. Опишем её в общем случае.

Определение 3.33. *Прямое произведение* групп G и H обозначается $G \times H$ и состоит из всех возможных пар (g, h), $g \in G$, $h \in H$. Операция в $G \times H$ — это покомпонентное выполнение операций в G и H:

$$(g_1, h_1) \cdot (g_2, h_2) = (g_1 \cdot g_2, h_1 \cdot h_2).$$

Замечание 3.34. Если группы абелевы мы также называем прямое произведение *прямой суммой* (поскольку часто групповая операция в абелевом случае записывается как сложение). Обозначение в этом случае $G \oplus H$.

Утверждение 3.35. *$G \times H$ является группой.*

Доказательство. Проверка всех аксиом группы выполняется покомпонентно. При проверке ассоциативности

$$\big((g_1,h_1)\cdot(g_2,h_2)\big)\cdot(g_3,h_3) = (g_1\cdot g_2, h_1\cdot h_2)\cdot(g_3,h_3) = \big((g_1\cdot g_2)\cdot g_3, (h_1\cdot h_2)\cdot h_3\big) \overset{*}{=}$$
$$= \big(g_1\cdot(g_2\cdot g_3), h_1\cdot(h_2\cdot h_3)\big) = (g_1,h_1)\cdot(g_2\cdot g_3, h_2\cdot h_3) =$$
$$= (g_1,h_1)\cdot\big((g_2,h_2)\cdot(g_3,h_3)\big)$$

равенство, отмеченное $*$, выполняется в силу ассоциативности умножения в группах G и H, остальные равенства получаются непосредственно из определения.

Единицей $G \times H$ является пара (e_G, e_H), где e_G — единица группы G, а e_H — единица группы H:

$$(g,h)\cdot(e_G,e_H) = (g\cdot e_G, h\cdot e_H) = (g,h).$$

Обратный к паре (g,h) — это элемент (g^{-1},h^{-1}):

$$(g,h)\cdot(g^{-1},h^{-1}) = (e_G,e_H).$$ □

Порядок сомножителей в прямом произведении не существенен.

Утверждение 3.36. *$G \times H \cong H \times G$.*

Доказательство. Изоморфизм задается отображением, меняющим компоненты местами: $\varphi\colon (g,h) \mapsto (h,g)$. Взаимная однозначность и сохранение групповой операции очевидны. □

Контрольный вопрос 3.37. Докажите, что $(G \times H) \times K \cong G \times (H \times K)$. (Формально это разные группы. Из каких элементов они состоят?)

В силу утверждения вопроса 3.37 корректно (с точностью до изоморфизма) определено прямое произведение нескольких групп $G_1 \times G_2 \times \cdots \times G_n$. По-другому произведение нескольких групп можно определить как множество всех возможных последовательностей $(g_1, g_2, \ldots, g_n)$, $g_i \in G_i$, с групповой операцией покомпонентного произведения

$$(g_1,g_2,\ldots,g_n)\cdot(h_1,h_2,\ldots,h_n) = (g_1\cdot h_1, g_2\cdot h_2, \ldots, g_n\cdot h_n).$$

Если все сомножители одинаковы, то такое произведение обозначается G^n, как в примере 3.30 («прямая степень» группы).

Контрольный вопрос 3.38. Докажите, что группа $G_1 \times G_2 \times \cdots \times G_n$ абелева тогда и только тогда, когда все группы-сомножители абелевы.

Из комбинаторного правила произведения получаем такое следствие.

Утверждение 3.39. *Если G_i — конечные группы, то*

$$|G_1 \times G_2 \times \cdots \times G_n| = |G_1| \cdot |G_2| \cdot \ldots \cdot |G_n|$$

(в правой части обычное произведение чисел).

Следующее утверждение выглядит очевидно, однако для полноты приведём его доказательство.

Утверждение 3.40. *Если $G_i \cong H_i$, то $G_1 \times G_2 \times \cdots \times G_n \cong H_1 \times H_2 \times \cdots \times H_n$.*

Доказательство. Если $\varphi_i \colon G_i \to H_i$ — изоморфизмы сомножителей, то изоморфизм прямых произведений задаётся правилом

$$\varphi \colon (g_1, g_2, \ldots, g_n) \mapsto \big(\varphi_1(g_1), \varphi_2(g_2), \ldots, \varphi_n(g_n)\big). \tag{3.2}$$

Это биекция, обратное отображение задаётся формулой

$$\varphi^{-1} \colon (h_1, h_2, \ldots, h_n) \mapsto \big(\varphi_1^{-1}(h_1), \varphi_2^{-1}(h_2), \ldots, \varphi_n^{-1}(h_n)\big).$$

Из определений прямого произведения и изоморфизма сразу следует, что отображение (3.2) сохраняет групповую операцию. □

Пример 3.41 (булев куб)**.** Группа C_2^n изоморфна группе Z_2^n, которую можно описать как множество двоичных наборов $(a_1, \ldots, a_n)$ длины n, $a_i \in \{0, 1\}$, с операцией покомпонентного сложения по модулю 2. □

Пример 3.42. Группа $\mathbb{Z}^n$ состоит из последовательностей целых чисел длины n с операцией покомпонентного сложения.

Группы $\mathbb{Z}^n$ и $\mathbb{Z}^k$ при $n \neq k$ неизоморфны. Хотя это утверждение выглядит очевидно, доказать его намного сложнее, чем неизоморфизм групп $\mathbb{Z}^1$ и $\mathbb{Z}^2$ (пример 3.30). Ниже мы приведём доказательство, для этого потребуются новые средства, которые мы постепенно введём в игру. □

Замечание 3.43. Соблазнительно доказать неизоморфизм $\mathbb{Z}^n$ и $\mathbb{Z}^k$ при $n \neq k$ индукцией. Для шага индукции хотелось бы использовать такое утверждение: если $G \times H_1 \cong G \times H_2$, то $H_1 \cong H_2$ («закон сокращения» для прямых произведений, подумайте как закончить рассуждение по индукции с его помощью).

Однако этот закон сокращения неверен. Рассмотрим группу $\mathbb{Z}^\infty$ бесконечных последовательностей целых чисел с операцией покомпонентного сложения (проверка свойств группы аналогична предыдущим случаям). Проверим, что $\mathbb{Z} \times \mathbb{Z}^\infty \cong \mathbb{Z}^\infty \cong \mathbb{Z}^2 \times \mathbb{Z}^\infty$.

Сопоставим бесконечной последовательности $a_0, a_1, \ldots, a_n, \ldots$ пару из числа a_0 и бесконечной последовательности $a_1, \ldots, a_n, \ldots$. Это соответствие имеет обратное: пара из числа b и последовательности $b_0, b_1, \ldots, b_n, \ldots$ соответствует ровно одной последовательности $b, b_0, b_1, \ldots, b_n, \ldots$.

Получили биекцию $\varphi \colon \mathbb{Z}^\infty \to \mathbb{Z} \times \mathbb{Z}^\infty$. Эта биекция сохраняет операцию, поскольку сложение последовательностей осуществляется покомпонентно и в прямом произведении групповая операция выполняется покомпонентно.

Итак, построен изоморфизм $\mathbb{Z}^\infty \cong \mathbb{Z} \times \mathbb{Z}^\infty$. Изоморфизм $\mathbb{Z}^\infty \cong \mathbb{Z}^2 \times \mathbb{Z}^\infty$ устроен аналогично: последовательности $a_0, a_1, a_2, \ldots, a_n, \ldots$ он сопоставляет пару пару из пары чисел (a_0, a_1) и бесконечной последовательности $a_2, \ldots, a_n, \ldots$. Проверка свойств изоморфизма аналогична предыдущему.

Как мы уже видели в примере 3.30, $\mathbb{Z}$ не изоморфна $\mathbb{Z}^2$. Поэтому закон сокращения для прямых произведений не выполняется.

Для доказательства неизоморфизма произведений конечных групп зачастую полезно рассмотреть возможные порядки элементов в прямом произведении групп. Общее утверждение формулируется и доказывается аналогично лемме 2.78.

Лемма 3.44. *Порядок элемента* $(g_1, g_2, \ldots, g_n) \in G_1 \times G_2 \times \cdots \times G_n$ *равен* НОК *порядков элементов* g_i.

Доказательство. Равенство $(g_1, g_2, \ldots, g_n)^k = e$ равносильно тому, что в каждом сомножителе $g_i^k = e$ (здесь и далее для краткости символ e обозначает нейтральные элементы в разных группах).

Но если $g_i^k = e$, то k является кратным порядка элемента g_i. Таким образом, $(g_1, g_2, \ldots, g_n)^k = e$ тогда и только тогда, когда k является общим кратным порядков элементов. □

Пример 3.45. Найдём порядок элемента $(10, 15)$ в прямом произведении циклических групп $Z_{18} \times Z_{24}$. Используя лемму 2.55 (формула для порядка элемента в циклической группе) и лемму 3.44, получаем

$$\operatorname{ord}_{Z_{18}}(10) = \frac{18}{\text{НОД}(10, 18)} = 9, \quad \operatorname{ord}_{Z_{24}}(15) = \frac{24}{\text{НОД}(15, 24)} = 8,$$
$$\operatorname{ord}_{Z_{18} \times Z_{24}}(10, 15) = \text{НОК}(9, 8) = 72.$$

Таким образом, искомый порядок равен 72. □

Пример 3.46. Докажем, что группы $C_9 \times C_{16}$ и $C_{12} \times C_{12}$ неизоморфны. (Хотя порядки их равны: $9 \cdot 16 = 144 = 12 \cdot 12$.)

Поскольку 9 и 16 взаимно просты, в группе $C_9 \times C_{16}$ есть элемент порядка 144 (это пара, составленная из порождающих групп C_9 и C_{16}). Однако в группе $C_{12} \times C_{12}$ порядки всех элементов не превосходят 12 (порядок любого элемента делит порядок группы, значит, 12 всегда является общим кратным порядков любой пары элементов из C_{12}).

Так как изоморфизм сохраняет порядки элементов, делаем вывод, что группы неизоморфны. □

Аналогия между леммами 2.78 и 3.44 имеет сравнительно простое объяснение. Рассмотрим подгруппу

$$H = \langle \pi_1, \pi_2, \ldots, \pi_k \rangle < S_n$$

группы перестановок, порождённую циклами π_i, носители которых не пересекаются (то есть π_1 переставляет числа из множества S_1, а остальные оставляет на месте; π_2 переставляет числа из множества S_2, а остальные оставляет на месте и т.д.; при

этом множества S_1, S_2, …, S_k попарно не пересекаются). Обозначим через ℓ_i длину цикла π_i.

Утверждение 3.47. $H \cong C_{\ell_1} \times C_{\ell_2} \times \cdots \times C_{\ell_k}$.

Доказательство. Аналогично анализу, проведённому в примере 2.81.

Непересекающиеся циклы коммутируют, $\pi_i\pi_j = \pi_j\pi_i$. Поэтому после перестановки сомножителей и сокращения пар одинаковых сомножителей любое произведение π_i и их обратных приводится к виду $\pi_1^{a_1}\pi_2^{a_2}\cdot\ldots\cdot\pi_k^{a_k}$. Так как порядок цикла π_i равен его длине ℓ_i (утверждение 2.77), то достаточно рассматривать только $0 \leqslant a_i < \ell_i$.

Все такие произведения различны. Равенство $\pi_1^{a_1}\pi_2^{a_2}\cdot\ldots\cdot\pi_k^{a_k} = \mathrm{id}$ равносильно тому, что каждое a_i кратно ℓ_i. А так как равенство $\pi_1^{a_1}\pi_2^{a_2}\cdot\ldots\cdot\pi_k^{a_k} = \pi_1^{b_1}\pi_2^{b_2}\cdot\ldots\cdot\pi_k^{b_k}$ равносильно равенству $\pi_1^{a_1-b_1}\pi_2^{a_2-b_2}\cdot\ldots\cdot\pi_k^{a_k-b_k} = \mathrm{id}$, равенство произведений при условиях $0 \leqslant a_i < \ell_i$ на показатели степеней возможно лишь при совпадении всех показателей в произведениях.

Из коммутативности непересекающихся циклов сразу следует равенство

$$\left(\pi_1^{a_1}\pi_2^{a_2}\cdot\ldots\cdot\pi_k^{a_k}\right)\circ\left(\pi_1^{b_1}\pi_2^{b_2}\cdot\ldots\cdot\pi_k^{b_k}\right) = \pi_1^{a_1+b_1}\pi_2^{a_2+b_2}\cdot\ldots\cdot\pi_k^{a_k+b_k}.$$

Сделанные наблюдения означают, что отображение

$$\varphi\colon \pi_1^{a_1}\pi_2^{a_2}\cdot\ldots\cdot\pi_k^{a_k} \mapsto (a_1, a_2\ldots, a_k)$$

является изоморфизмом группы H и группы $Z_{\ell_1} \times Z_{\ell_2} \times \cdots \times Z_{\ell_k}$. □

Группы автоморфизмов прямых произведений циклических групп устроены более сложно, чем группы автоморфизмов циклических групп. Пока ограничимся лишь тем, что определим порядок одной из таких групп.

Пример 3.48. Найдём порядок группы автоморфизмов $Z_5 \times Z_5$.

Эта группа порождается двумя векторами $(1, 0)$ и $(0, 1)$. Поэтому её автоморфизм однозначно задаётся образами этих векторов. Зафиксируем образы:

$$(1, 0) \mapsto (a, b),$$
$$(0, 1) \mapsto (c, d).$$

Любой паре векторов (a, b) и (c, d) можно сопоставить отображение группы $Z_5 \times Z_5$ в себя, сохраняющее операцию (в таком случае говорят об *эндоморфизме* группы):

$$(x, y) \mapsto x(a, b) + y(c, d) = (xa + yc, xb + yd). \tag{3.3}$$

Сохранение операции легко проверить (образ суммы равен сумме образов для отображения, задаваемого линейными формулами).

Осталось выяснить, когда такое отображение будет взаимно однозначным. Ясно, что если $(c, d) = k(a, b)$, то отображение не взаимно однозначно: все образы лежат в подгруппе $\langle(a, b)\rangle$.

В остальных случаях отображение взаимно однозначно. Это легко проверяется средствами линейной алгебры, но приведём прямое доказательство.

Пусть $(c,d) \notin \langle (a,b) \rangle$, в частности $(c,d) \neq (0,0)$. Тогда порядок (c,d) равен 5, так как 5 — простое число. По лемме 3.32 любое его ненулевое кратное не лежит в подгруппе $\langle (a,b) \rangle$, порядок которой также равен 5.

Итак, в каждом смежном классе по подгруппе $\langle (a,b) \rangle$ лежит ровно одно кратное (c,d). Это означает, что любой элемент $Z_5 \times Z_5$ однозначно раскладывается в сумму кратных (a,b) и (c,d), то есть отображение (3.3) взаимно однозначно.

Теперь остались несложные комбинаторные подсчёты. Чтобы отображение (3.3) было автоморфизмом группы $Z_5 \times Z_5$ необходимо и достаточно, чтобы $(a,b) \neq \neq (0,0)$ (24 варианта выбора), а $(c,d) \notin \langle (a,b) \rangle$ $(25 - 5 = 20$ вариантов выбора). По комбинаторному правилу произведения получаем, что порядок $\mathrm{Aut}(Z_5 \times Z_5)$ равен $24 \cdot 20 = 480$. □

3.4 Китайская теорема об остатках для целых чисел

Сформулируем обобщение наблюдения, которое уже встретилось в примере 3.46.

Теорема 3.49 (китайская теорема об остатках)**.** *Если p, q — взаимно простые числа, то $C_{pq} \cong C_p \times C_q$.*

Доказательство. Достаточно указать в группе $C_p \times C_q$ элемент порядка pq. Это пара (a,b), где a — порождающий C_p, а b — порождающий C_q. Так как p и q взаимно просты, то по лемме 3.44 порядок пары $(a,b) \in C_p \times C_q$ равен pq. □

Условие взаимной простоты порядков групп-сомножителей существенно для изоморфизма, см. пример 3.46.

Эта теорема позволяет устанавливать изоморфизмы прямых произведений циклических групп.

Пример 3.50. Изоморфны ли группы $C_{28} \times C_{26}$ и $C_{52} \times C_{14}$? Разложим эти группы в прямые произведения меньших, используя теорему 3.49:

$$C_{28} \times C_{26} \cong C_4 \times C_7 \times C_2 \times C_{13},$$
$$C_{52} \times C_{14} \cong C_4 \times C_{13} \times C_2 \times C_7.$$

Получившиеся разложения в прямое произведение одинаковы с точностью до перестановки сомножителей. Поэтому группы изоморфны (см. выше утверждения 3.36 и 3.37). □

Мы сформулировали китайскую теорему об остатках алгебраически (и позже дадим ещё более общую формулировку). В теории чисел под китайской теоремой об остатках понимают обычно более конкретное утверждение.

Возьмём группу $Z_p \times Z_q$, где p и q взаимно просты. Поскольку 1 является порождающим элементом для любой Z_n, то из доказательства теоремы 3.49 следует, что порядок пары $(1,1)$ равен pq. Другими словами, если брать поочерёдно пары $(k \bmod p, k \bmod q)$ при $k = 0, 1, \ldots, pq-1$, то получим все возможные пары остатков по модулям p и q, причём ровно по разу.

Сформулируем это наблюдение в виде следствия.

Следствие 3.51. *Если p, q — взаимно простые числа, то у системы сравнений*

$$x \equiv a \pmod{p},$$
$$x \equiv b \pmod{q}$$

есть ровно одно решение x_0 в диапазоне от 0 до $pq-1$. А все возможные целые решения такой системы сравнений образуют класс вычетов с остатком x_0 по модулю pq.

Применив индукцию, нетрудно обобщить это следствие на системы сравнений по нескольким модулям, если они попарно взаимно просты.

Пример 3.52. Найдём наименьшее положительное целое число x такое, что

$$x \equiv 4 \pmod{5};$$
$$x \equiv 5 \pmod{6};$$
$$x \equiv 6 \pmod{7}.$$

Для начала заметим, что -1 является решением этой системы сравнений. Поэтому все целые решения — это класс вычетов $[-1]_{5\cdot6\cdot7}$. Наименьшее положительное число в этом классе равно $5 \cdot 6 \cdot 7 - 1 = 209$. □

Функция Эйлера $\varphi(n)$ равна порядку мультипликативной группы вычетов по модулю n, что то же самое, количеству обратимых вычетов по модулю n, или количеству натуральных чисел, меньших n и взаимно простых с n.

Теорема 3.49 позволяет установить важное свойство функции Эйлера. Тут мы используем ещё одну характеризацию функции Эйлера, даваемую утверждением 3.25: это количество порождающих циклической группы порядка n.

Теорема 3.53. *Функция $\varphi(n)$ мультипликативна, то есть если* НОД$(n, m) = 1$, *то* $\varphi(nm) = \varphi(n)\varphi(m)$.

Доказательство. Найдём количество порождающих в группе $C_n \times C_m$. По теореме 3.49 и утверждению 3.25 оно равно $\varphi(nm)$.

С другой стороны, порядок пары $(x, y) \in C_n \times C_m$ равен nm тогда и только тогда, когда порядок x в C_n равен n, а порядок y в C_m равен m. Первых элементов ровно $\varphi(n)$ штук, вторых — $\varphi(m)$ штук. Комбинаторное правило произведения говорит, что всего таких пар (то есть порождающих $C_n \times C_m$) ровно $\varphi(n)\varphi(m)$ штук. □

Обратите внимание, что свойство мультипликативности функции Эйлера выполняется не для всех пар сомножителей, а только для взаимно простых.

Например, легко видеть, что $\varphi(p^n) = p^n - p^{n-1}$ для любого простого p (не взаимно просты с p^n только остатки, которые делятся на p, а их в p раз меньше общего количества остатков). Свойство мультипликативности для произведений степеней простых нарушается. Например,

$$\varphi\big(2^2 \cdot 2^2\big) = 2^4 - 2^3 = 8 \neq 4 = \big(2^2 - 2\big)\big(2^2 - 2\big) = \varphi\big(2^2\big)\varphi\big(2^2\big).$$

Основная теорема арифметики говорит, что каждое целое число раскладывается в произведение степеней различных простых чисел. Отсюда получаем полезную формулу для функции Эйлера.

Лемма 3.54. *Если* $n = p_1^{a_1} \cdot p_2^{a_2} \cdot \dots \cdot p_s^{a_s}$, *то*

$$\varphi(n) = (p_1^{a_1} - p_1^{a_1-1})(p_2^{a_2} - p_1^{a_2-1}) \dots (p_s^{a_s} - p_s^{a_s-1}) = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_s}\right)\dots\left(1 - \frac{1}{p_s}\right).$$

Доказательство. Первое равенство непосредственно следует из мультипликативности функции Эйлера и формулы для $\varphi(p^n)$, где p — простое.

Второе равенство получается несложным преобразованием

$$\begin{aligned}(p_1^{a_1} - p_1^{a_1-1})(p_2^{a_2} - p_1^{a_2-1}) \dots (p_s^{a_s} - p_s^{a_s-1}) = \\ = p_1^{a_1}(1 - p_1^{-1})p_2^{a_2}(1 - p_1^{-1}) \dots p_s^{a_s}(1 - p_s^{-1}) = \\ = n\left(1 - \frac{1}{p_1}\right)\left(1 - \frac{1}{p_s}\right)\dots\left(1 - \frac{1}{p_s}\right),\end{aligned}$$

что и требовалось. □

Пример 3.55. Докажем, что сравнение $x^{11} \equiv 1 \pmod{10^6}$ имеет единственное решение (в вычетах, разумеется, как всегда при подсчёте количества решений сравнений).

Одно решение очевидно: $x = 1$. Осталось показать, что в мультипликативной группе вычетов $Z_{10^6}^*$ нет элемента порядка 11.

Найдём $\varphi(10^6) = 10^6 \cdot \frac{1}{2} \cdot \frac{4}{5}$ по формуле из леммы 3.54. Простые делители этого числа — 2 и 5. Так как порядок элемента делит порядок группы (в данном случае $\varphi(10^6)$), заключаем, что элементов порядка 11 в этой группе нет. □

Китайская теорема позволяет найти количество решений сравнения

$$x^2 \equiv 1 \pmod{n}$$

для любого n. Выше мы нашли решения этих сравнений для степеней простых (примеры 2.70 и 2.71). Произвольное n нужно разложить по степеням простых и воспользоваться следующим утверждением.

Утверждение 3.56. *Пусть* p, q *— взаимно простые числа. Если количество решений сравнения* $x^2 \equiv 1 \pmod{p}$ *равно* N, *а количество решений сравнения* $x^2 \equiv 1 \pmod{q}$ *равно* L, *то количество решений сравнения* $x^2 \equiv 1 \pmod{pq}$ *равно* NL.

Доказательство. Число y даёт остаток 1 при делении на pq тогда и только тогда, когда оно даёт остаток 1 при делении на p и при делении на q.

Поэтому любому решению сравнения $x^2 \equiv 1 \pmod{pq}$ отвечает пара ($x \bmod p, x \bmod q$) решений сравнений $x^2 \equiv 1 \pmod{p}$ и $x^2 \equiv 1 \pmod{q}$. В другую сторону китайская теорема гарантирует, что любой паре решений сравнений $x^2 \equiv 1 \pmod{p}$ и $x^2 \equiv 1 \pmod{q}$ отвечает какое-то решение сравнения $x^2 \equiv 1 \pmod{pq}$. □

Пример 3.57. Сколько решений у сравнения $x^2 \equiv 1 \pmod{360}$?

Разложим 360 на произведение степеней простых: $360 = 5 \cdot 8 \cdot 9$. Решений сравнения $x^2 \equiv 1 \pmod 5$ два, сравнения $x^2 \equiv 1 \pmod 8$ четыре, а $x^2 \equiv 1 \pmod 8$ снова два.

Итого получаем $2 \cdot 4 \cdot 2 = 16$ решений. □

В заключение этого раздела заметим, что для мультипликативных групп вычетов есть изоморфизм, аналогичный китайской теореме об остатках.

Теорема 3.58. *Если p, q — взаимно простые числа, то $Z^*_{pq} \cong Z^*_p \times Z^*_q$.*

Доказательство. В доказательстве свойства мультипликативности функции Эйлера мы установили, что остатков, взаимно простых с pq, столько же, сколько пар остатков, взаимно простых с p и q соответственно (это ведь в точности порождающие соответствующих аддитивных групп). Рассмотрим взаимно однозначное отображение

$$k \mapsto (k \bmod p, k \bmod q), \qquad 0 \leqslant k < pq.$$

Оно является изоморфизмом аддитивных групп вычетов, поэтому переводит взаимно простые с pq остатки в пары остатков, взаимно простых с p и q соответственно.

Покажем, что такая биекция задаёт изоморфизм мультипликативных групп вычетов. Действительно, как мы уже проверяли, взятие остатка согласовано с произведением

$$(k\ell) \bmod p = \big((k \bmod p) \cdot (\ell \bmod p)\big) \bmod p,$$
$$(k\ell) \bmod q = \big((k \bmod q) \cdot (\ell \bmod q)\big) \bmod q.$$

Но это и означает, что образ произведения $k\ell$ элементов группы Z^*_{pq} совпадает с произведением пар $(k \bmod p, k \bmod q)$ и $(\ell \bmod p, \ell \bmod q)$ в группе $Z^*_p \times Z^*_q$. □

4 Гомоморфизмы групп и факторгруппы

4.1 Определение, примеры и основные свойства гомоморфизмов групп

Определение изоморфизма групп указывает на два свойства: биективность и сохранение операции. Если оставить только второе, получаем важнейшее для теории групп понятие гомоморфизма. (В более общих алгебраических ситуациях используется слово «морфизм», но мы без него в этом вводном курсе обойдёмся.)

Определение 4.1. Отображение $\varphi\colon G \to G'$, где $(G, *)$, $(G', \circ)$ — две группы, называется *гомоморфизмом*, если $\varphi(a * b) = \varphi(a) \circ \varphi(b)$.

Поскольку от гомоморфизма не требуется взаимной однозначности, алгебраические свойства групп G и G' могут сильно различаться.

Фактически мы уже встречали примеры гомоморфизмов, только не называли их этим словом.

Пример 4.2. Основное для элементарной теории чисел отображение $\mathbb{Z} \to Z_n$ целых чисел в аддитивную группу вычетов по модулю n является гомоморфизмом, как мы уже проверяли выше (утверждение 1.43).

Как мы видели, свойства группы вычетов отчасти похожи на свойства группы целых чисел по сложению. □

Пример 4.3. Для любой группы G и её элемента $g \in G$ отображение $\mathbb{Z} \to G$, задаваемое правилом

$$k \mapsto g^k,$$

является гомоморфизмом. Сохранение операции очевидно (формально нужно использовать лемму 2.20).

Если группа G содержит элемент g порядка n, то отображение $k \mapsto g^k$ задаёт гомоморфизм циклической группы Z_n в группу G. Для корректности такого определения (однозначности отображения) существенно, что порядок элемента совпадает с порядком циклической группы.

Такой гомоморфизм мало что говорит о свойствах группы G, она может сколь угодно сильно отличаться от циклической группы. □

Укажем ещё несколько простых свойств гомоморфизмов групп, которые далее будут постоянно использоваться.

Утверждение 4.4. *Композиция гомоморфизмов — гомоморфизм.*

Доказательство. Обозначая для краткости групповую операцию одинаково во всех рассматриваемых группах, проверим сохранение операции композицией гомоморфизмов $\varphi\colon G_1 \to G_2$ и $\psi\colon G_2 \to G_3$:

$$(\psi \circ \varphi)(gh) = \psi(\varphi(gh)) = \psi(\varphi(g) \cdot \varphi(h)) = \psi(\varphi(g)) \cdot \psi(\varphi(h)) = (\psi \circ \varphi)(g) \cdot (\psi \circ \varphi)(h)$$

(здесь $\circ$ обозначает композицию отображений). □

Пример 4.5. Из линейной алгебры известно, что определитель произведения матриц равен произведению определителей: $\det(AB) = \det(A)\det(B)$. Поэтому отображение

$$\det\colon GL(\mathbb{R}, n) \to \mathbb{R}^*, \quad \det\colon A \mapsto \det(A)$$

является гомоморфизмом группы $GL(\mathbb{R}, n)$ невырожденных матриц порядка n с действительными элементами в мультипликативную группу действительных чисел, отличных от 0. □

Утверждение 4.6. *Пусть $\varphi\colon G \to G'$ — гомоморфизм групп $(G, *)$ и $(G', \circ)$. Тогда*

1. *$\varphi(e)$ — нейтральный элемент группы G' (изоморфизм сохраняет единицу);*
2. *$\varphi(a^{-1}) = \varphi(a)^{-1}$ (образ обратного элемента — обратный элемент к образу).*

Доказательства буквально повторяют доказательства тех же свойств для изоморфизмов (свойства 1, 2 в утверждении 3.3; прочитав доказательства, легко убедиться, что в них не используется лишь свойство сохранения операции).

Определение 4.7 (ядро и образ)**.** Пусть $\varphi\colon G \to G'$ — гомоморфизм групп. Тогда *образ* $\operatorname{Im}\varphi = \varphi(G)$ состоит из тех элементов $g' \in G'$, для которых есть прообраз (такой элемент $g \in G$, что $\varphi(g) = g'$). Это общее определение для любого отображения.

Ядром $\operatorname{Ker}\varphi$ гомоморфизма называется множество $\{g \in G : \varphi(g) = e_{G'}\}$. Другими словами, элемент принадлежит ядру, если он отображается гомоморфизмом в нейтральный элемент.

Контрольный вопрос 4.8. Найдите образ и ядро для примеров гомоморфизмов, которые приведены выше.

Утверждение 4.9. *Ядро и образ являются подгруппами.*

Доказательство. Пусть $\varphi\colon G \to G'$ — гомоморфизм групп, обозначим $H = \varphi(G)$.

Единица e' группы G' принадлежит H и образ H замкнут относительно взятия обратного (утверждение 4.6). Если $\varphi(x), \varphi(y) \in H$, то $\varphi(x) \circ \varphi(y) = \varphi(x * y) \in H$, то есть образ замкнут относительно групповой операции. Значит, это подгруппа.

Обозначим $K = \operatorname{Ker}\varphi$. Если $x, y \in K$, то есть $\varphi(x) = \varphi(y) = e'$, то $\varphi(xy) = \varphi(x)\varphi(y) = e'e' = e'$. Поэтому $xy \in K$ и ядро замкнуто относительно групповой операции. Нейтральный элемент e группы G ядро содержит в силу утверждения 4.6. Замкнутость относительно взятия обратного проверяется прямым вычислением. Пусть $\varphi(x) = e'$. Тогда

$$\varphi(x^{-1}) = \varphi(x^{-1})e' = \varphi(x^{-1})\varphi(x) = \varphi(x^{-1}x) = \varphi(e) = e',$$

в третьем равенстве использовано сохранение операции при гомоморфизме, а в последнем ещё раз использовано утверждение 4.6. □

Не только образ всей группы является подгруппой, но и образ любой подгруппы является подгруппой. Действительно, ограничение гомоморфизма на подгруппу является гомоморфизмом (подгруппа замкнута относительно групповой операции). Поэтому достаточно применить утверждение 4.9 к этому ограничению.

Полный прообраз подгруппы при гомоморфизме также является подгруппой.

Утверждение 4.10. *Пусть $\varphi\colon G \to G'$ — гомоморфизм групп, H — подгруппа G'. Тогда $\varphi^{-1}(H)$ является подгруппой G.*

Доказательство. Применим критерий подгруппы, заданный в теореме 2.26: множество элементов группы является подгруппой, тогда и только тогда, когда оно замкнуто относительно операции xy^{-1}.

Пусть $x, y \in \varphi^{-1}(H)$, то есть $\varphi(x) \in H$ и $\varphi(y) \in H$. Образ xy^{-1} в силу сохранения операции при гомоморфизме имеет вид $\varphi(x) \cdot \varphi(y^{-1})$. Поскольку $\varphi(y^{-1}) = \varphi(y)^{-1}$, то $\varphi(xy^{-1}) = \varphi(x) \cdot \varphi(y)^{-1}$ принадлежит подгруппе H, то есть $xy^{-1} \in \varphi^{-1}(H)$, что и требовалось проверить. □

Гомоморфизмов из группы в группу гораздо меньше, чем произвольных отображений. Мы уже установили, что (непустой) прообраз единицы — ядро гомоморфизма — является подгруппой. Это утверждение можно уточнить: множество элементов, имеющих одинаковый образ, является смежным классом по ядру.

Утверждение 4.11. *Пусть $\varphi\colon G \to G'$ — гомоморфизм групп. Тогда элементы (левого или правого) смежного класса по $\operatorname{Ker}\varphi$ переходят в один и тот же элемент группы G'. И обратно: если $g' \in G'$, то $\varphi^{-1}(g')$ является смежным классом (левым и правым) по ядру.*

Доказательство. Пусть $g = hk$, $k \in \operatorname{Ker}\varphi$. Тогда $\varphi(g) = \varphi(hk) = \varphi(h)\varphi(k) = \varphi(h)e_{G'} = \varphi(h)$. Поэтому образ смежного класса состоит из одного элемента.

В обратную сторону точно так же: пусть $\varphi(x_1) = \varphi(x_2)$. Тогда

$$\varphi(x_1^{-1}x_2) = \varphi(x_1^{-1})\varphi(x_2) = \varphi(x_1)^{-1}\varphi(x_2) = \varphi(x_2)^{-1}\varphi(x_2) = e_{G'},$$

то есть $x_1^{-1}x_2 \in \operatorname{Ker}\varphi$. Из леммы 2.38 заключаем, что x_1 и x_2 принадлежат одному смежному классу по ядру.

Доказательства для правых смежных классов аналогичны. □

Из этого утверждения немедленно выводятся два важных следствия (леммы 4.12 и 4.16 ниже).

Лемма 4.12. *У ядра любого гомоморфизма левые классы смежности совпадают с правыми.*

Доказательство. Для элемента $g \in G$ и левый, и правый классы смежности совпадают с прообразом элемента $\varphi(g) \in G$. □

Определение 4.13. Если для любого $x \in G$ левый смежный класс по подгруппе H совпадает с правым, $xH = Hx$, то такая подгруппа называется *нормальной*.

Для нормальных подгрупп используется специальное обозначение $H \triangleleft G$.

В этих терминах лемма 4.12 формулируется так: ядро гомоморфизма является нормальной подгруппой.

Приведём простые примеры нормальных подгрупп. Сама группа и её единичная подгруппа являются нормальными: в первом случае смежный класс один, а во

втором каждый элемент группы является и левым, и правым смежным классом по единичной подгруппе. Вот чуть более сложный пример.

Пример 4.14. Любая подгруппа индекса 2 нормальная. Действительно, в ней есть ровно два левых смежных класса: H и $G \setminus H$ и ровно два правых смежных класса: тех же самых. □

Далеко не все группы нормальны и это накладывает ограничения на возможные гомоморфизмы.

Пример 4.15. Существует ли гомоморфизм $S_9 \to H$, ядром которого является подгруппа G, состоящая из перестановок, которые элемент 1 переводят в себя?

Ответ: не существует. Левые смежные классы по подгруппе G состоят из тех перестановок, которые переводят 1 в одно и то же число. Правые — из перестановок, у которых прообразы 1 одинаковы. (Мы проводили аналогичный анализ, основанный на критерии леммы 2.38, выше в примере 2.39.)

Эти классы не совпадают: левый смежный класс по подгруппе F, содержащий перестановку (1 2), содержит также и перестановку (1 2 3). Однако эти две перестановки лежат в разных правых смежных классах (прообразы 1 различаются).

Поэтому подгруппа G не является нормальной и потому не является ядром какого-либо гомоморфизма. □

Лемма 4.16. *Для гомоморфизма конечных групп $\varphi\colon G \to G'$ выполняется равенство*

$$|G| = |\operatorname{Ker}\varphi| \cdot |\operatorname{Im}\varphi|.$$

Доказательство. Так как количество элементов в образе группы в точности равно индексу ядра (количеству смежных классов), достаточно применить теорему Лагранжа. □

4.1.1 Квадратичные вычеты и невычеты

Квадратичным вычетом будем называть такой обратимый вычет a по простому модулю p, который является точным квадратом: $x^2 \equiv a \pmod p$ для некоторого x. Остальные обратимые вычеты будем называть квадратичными невычетами.

Например, 2 является квадратичным вычетов по модулю 23, поскольку $5^2 \equiv 2 \pmod{23}$. А вот 5 — квадратичный невычет по модулю 23. Как это проверить? Конечно, можно составить таблицу всех квадратов по модулю 23 и убедиться, что 5 в неё не входит. Есть, однако, более удобный способ.

Мы уже проверяли выше, что отображение $x \mapsto x^2$ является гомоморфизмом мультипликативной группы вычетов (как и любой абелевой группы с коммутативной групповой операцией). Квадратичные вычеты — образ Z_p^* при этом гомоморфизме. Ядро этого гомоморфизма — это решения сравнения $x^2 \equiv 1 \pmod p$. Мы находили эти решения в примере 2.59. Их ровно два: ± 1. Поэтому из леммы 4.16 получаем, что порядок группы квадратов равен $|Z_p^*|/2 = (p-1)/2$.

По теореме Лагранжа это означает, что для любого квадратичного вычета выполняется сравнение $a^{(p-1)/2} \equiv 1 \pmod p$. Получаем необходимое условие того, что

a является квадратичным вычетом. (Позже мы докажем, что это условие является и достаточным, но пока мы доказали лишь необходимость.)

Вычет 5 не удовлетворяет этому условию, как легко проверить вычислением:

$$5^{(23-1)/2} = 5^{11} = 25^5 \cdot 5 \equiv 2^5 \cdot 5 \equiv 9 \cdot 5 \equiv -1 \pmod{23}.$$

Поэтому это квадратичный невычет.

Обратите внимание, что в ответе получился вычет -1. Это неудивительно, так как квадрат $a^{(p-1)/2}$ по малой теореме Ферма равен 1. А все решения сравнения $x^2 \equiv 1 \pmod{p}$ — это ± 1.

4.1.2 Чётность перестановок

Чтобы почувствовать, насколько сильны ограничения на существование гомоморфизма, рассмотрим следующий важный пример.

Найдём гомоморфизмы $\sigma\colon S_n \to C_2$ из группы перестановок n элементов в циклическую группу порядка 2. Нам будет удобно представлять эту циклическую группу как $\{\pm 1, \cdot\}$, то есть подгруппу группы ненулевых действительных чисел по умножению.

Один гомоморфизм очевиден: он переводит все перестановки в $+1$. Какие ещё есть?

Вспомним, что транспозиции порождают S_n (теорема 1.76). Поэтому если все транспозиции лежат в ядре σ, то и любая перестановка лежит в ядре σ.

Из коммутативности C_2 можно вывести и другой факт: если хотя бы одна транспозиция лежит в ядре, то и все транспозиции лежат в ядре. Для этого посмотрим на равенство

$$(k\ \ell) = (i\ k)(j\ \ell) \circ (i\ j) \circ (i\ k)(j\ \ell),$$

которое можно проверить прямым вычислением (сделайте это вычисление!), а можно получить «из общих соображений», которые мы обсудим позже (сопряжённость перестановок).

Применим к этому равенству гомоморфизм σ:

$$\begin{aligned}\sigma\big((k\ \ell)\big) &= \sigma\big((i\ k)(j\ \ell) \circ (i\ j) \circ (i\ k)(j\ \ell)\big) = \\ &= \sigma\big((i\ k)(j\ \ell)\big)\sigma\big((i\ j)\big)\sigma\big((i\ k)(j\ \ell)\big) = \sigma\big((i\ k)(j\ \ell)\big)^2\sigma\big((i\ j)\big) = \\ &= \sigma\big((i\ j)\big)\end{aligned}$$

(в последнем равенстве мы использовали тождество $x^2 = 1$ для группы C_2). Это означает, что образы всех транспозиций при гомоморфизме σ одинаковы.

Итак, если гомоморфизм σ нетривиальный (есть прообраз у -1), то образы всех транспозиций равны -1. Чтобы найти образ произвольной перестановки, нужно представить её как композицию N транспозиций и отобразить в $(-1)^N$. Это означает, что нетривиальный гомоморфизм $\sigma\colon S_n \to C_2$ единственный.

Однако мы не закончили анализ. Пока неясно, существует ли этот нетривиальный гомоморфизм. Могло бы так случиться, что одна и та же перестановка раскладывалась и в композицию чётного числа транспозиций, и в композицию нечётного

числа транспозиций. Это дало бы два противоречивых условия на образ такой перестановки и означало бы, что нетривиального гомоморфизма нет.

Оказывается, что такого противоречия не возникает, и нетривиальный гомоморфизм $\sigma\colon S_n \to C_2$ существует. Он сопоставляет перестановке *знак*. Перестановки знака $+1$ называются чётными, а знака -1 называются нечётными.

Докажем, что чётность числа транспозиций в любом разложении данной перестановки в композицию транспозиций одинакова. Для этого мы применим следующую лемму.

Лемма 4.17. *Умножение на транспозицию изменяет количество циклов в перестановке на* ± 1.

Доказательство. Рассмотрим циклы перестановки $(i\ j) \circ \pi$. Ясно, что среди них будут все циклы перестановки π, которые не содержат чисел i, j (на числа из этих циклов обе перестановки действуют одинаково, так что и циклы не изменятся).

Далее нужно рассмотреть два случая.

I. Числа i и j лежат на разных циклах перестановки π. В этом случае докажем, что в перестановке $(i\ j) \circ \pi$ эти циклы объединяются. Действительно, обозначим их как

$$(i = i_0\ i_1\ \ldots\ i_a) \quad \text{и} \quad (j = j_0\ j_1\ \ldots\ j_b).$$

Легко видеть, что перестановка $(i\ j) \circ \pi$ переводит i_a в j, j — в j_1, …, j_{b-1} — в j_b, j_b — в i, i — в i_1, …, i_{a-1} — в i_a.

II. Числа i и j лежат на одном цикле перестановки π. В этом случае докажем, что в перестановке $(i\ j) \circ \pi$ этот цикл распадается на два цикла. Обозначим цикл как

$$(i = i_0\ i_1\ \ldots\ i_a\ j = j_0\ j_1\ \ldots\ j_b).$$

Тогда легко проверить, что перестановка $(i\ j) \circ \pi$ переводит i_a в i, i — в i_1, …, i_{a-1} — в i_a (получили один цикл), а j_b — в j, j — в j_1, …, j_{b-1} — в j_b (получили второй цикл).

Лемма об изменении числа циклов при умножении на транспозицию доказана. □

Из этой леммы следует, что количество циклов при умножении на чётное количество транспозиций изменяется на чётное число, а при умножении на нечётное количество транспозиций — на нечётное число. Поэтому невозможно представить какую-либо перестановку одновременно и как произведение чётного числа транспозиций, и как произведение нечётного числа транспозиций.

Определение 4.18. Подгруппа чётных перестановок (ядро гомоморфизма знака перестановки) обозначается A_n и называется *знакопеременной группой*.

Контрольный вопрос 4.19. Найдите порядок группы A_n.

Чётность перестановки легко определяется по её цикловому разложению. Для начала найдём чётность одного цикла. Мы уже раскладывали цикл в произведение транспозиций. Напомним формулу (1.22) из доказательства теоремы 1.76:

$$(1\ 2\ \ldots\ k) = (k\ 1) \circ ((k-1)\ 1) \circ \cdots \circ (3\ 1) \circ (2\ 1).$$

Из этой формулы видно, что цикл нечётной длины раскладывается в чётное число транспозиций (и потому это чётная перестановка), а цикл чётной положительной длины — в нечётное число транспозиций (и потому это нечётная перестановка). Цикл длины 0 (тождественная перестановка) раскладывается в 0 транспозиций и потому это чётная перестановка (нейтральный элемент лежит в ядре).

Поскольку чётность произведения перестановок равна сумме по модулю 2 чётностей множителей (сохранение операции при гомоморфизме знака), получаем общий результат.

Лемма 4.20. *Перестановка чётна тогда и только тогда, когда в её цикловом разложении количество циклов чётной длины чётно.*

Следствие 4.21. *Перестановка и обратная к ней имеют одну и ту же чётность.*

Доказательство. Длины циклов в цикловом разложении перестановки и обратной к ней одни и те же: чтобы получить обратную перестановку нужно переписать каждый цикл в обратном порядке. □

Группа всех перестановок порождается транспозициями. Группа чётных перестановок порождается циклами длины 3.

Теорема 4.22. *Любая чётная перестановка является произведением циклов длины* 3.

Доказательство. Представим чётную перестановку как произведение чётного количества транспозиций, разобьём транспозиции на пары соседних и заменим каждую такую пару на произведения циклов длины 3, пользуясь легко проверяемыми равенствами

$$(ik) \circ (ij) = (ijk),$$
$$(ij) \circ (kl) = (ilk) \circ (ijk).$$

Первое равенство используем, если у соседних транспозиций есть общий элемент, второе — если таких элементов нет. □

Проверка на чётность позволяет во многих случаях быстро получать отрицательные ответы на разные вопросы.

Пример 4.23. Порождают ли перестановки порядка 3 группу S_{33}?

Как мы знаем, порядок перестановки — НОК длин циклов в её цикловом разложении. Если НОК каких-то чисел равно 3, то среди этих чисел есть только 1 и 3. Но циклы длины 3 — чётные перестановки. Любое их произведение также является чётной перестановкой. Значит, перестановки порядка 3 порождают только подгруппу A_{33} (теорема 4.22). □

Пример 4.24. Сколько решений в группе S_n имеет уравнение

$$x \circ (1\ 2)(3\ 4\ 5) \circ x = (1\ 2\ 3\ 4\ 5)?$$

Заметим, что при любом x чётность левой части совпадает с чётностью перестановки (1 2)(3 4 5) (так чётность x будет посчитана дважды и даст чёт при любой чётности x). Перестановка (1 2)(3 4 5) нечётная, а перестановка (1 2 3 4 5). Поэтому равенство невозможно.

Ответ: 0 решений. □

4.2 Факторгруппы. Теорема о гомоморфизмах групп

Оказывается, любой сюръективный гомоморфизм групп однозначно (с точностью до симметрий) определяется своим ядром, а любая нормальная подгруппа является ядром некоторого гомоморфизма. В этом разделе мы разберём эти два важнейших факта о гомоморфизмах групп. Начнём со второго.

Пусть дана нормальная подгруппа $H \triangleleft G$ некоторой группы. Как построить гомоморфизм, ядром которого она является? Мы уже знаем, что все элементы H обязаны переходить в нейтральный элемент группы-образа, а все элементы из одного смежного класса по подгруппе H должны переходить в один элемент. Это подсказывает такой способ построения группы-образа: пусть её элементы — это смежные классы, а групповая операция на смежных классах согласована с групповой операцией в группе G.

Один пример такого рода мы знаем, именно так строится аддитивная группа вычетов по модулю n: классы вычетов — это смежные классы по подгруппе $n\mathbb{Z}$ целых чисел, кратных n, а сумма классов вычетов совпадает с суммой каких-нибудь чисел из этих классов. Такое же определение дословно переносится и на общий случай нормальной подгруппы некоторой группы.

Определение 4.25. Пусть $H \triangleleft G$ — нормальная подгруппа группы G.

Факторгруппа G/H состоит из смежных классов по H. Групповая операция в факторгруппе определяется как

$$(xH) \cdot (yH) = (xy)H.$$

С этим определением есть та же проблема, что и с определением сложения классов вычетов: оно может оказаться некорректным, если результирующий смежный класс зависит от выбора представителей в смежных классах-сомножителях.

Пример 4.26. Попробуем применить определение 4.25 к подгруппе $H = \langle (1\ 2) \rangle < S_3$.

Составим таблицу умножения элементов группы S_3 на смежные классы по подгруппе H (смежные классы по H разделены линиями для наглядности):

	H	$(2\ 3)H$	$(1\ 3)H$
()	H	$(2\ 3)H$	$(1\ 3)H$
(1 2)	H	$(1\ 3)H$	$(1\ 2)H$
(2 3)	$(2\ 3)H$	H	$(1\ 3)H$
(1 3 2)	$(2\ 3)H$	$(1\ 3)H$	H
(1 3)	$(1\ 3)H$	$(2\ 3)H$	H
(1 2 3)	$(1\ 3)H$	H	$(2\ 3)H$

Из таблицы видно, что результаты произведения смежного класса $(2\ 3)H$ на себя различаются в зависимости от выбора представителя смежного класса:

$$(2\ 3)\cdot(2\ 3)H = H, \qquad (1\ 3\ 2)\cdot(2\ 3)H = (1\ 3)H.$$

Поэтому определить групповую операцию на множестве смежных классов по этой подгруппе не получается. □

Однако для нормальных подгрупп всё хорошо и определение корректно.

Лемма 4.27. *Для нормальной подгруппы определение* 4.25 *корректно задаёт операцию на множестве смежных классов и относительно этой операции множество смежных классов является группой.*

Доказательство. Напомним, что из нормальности группы следует равенство $xH = Hx$ для любого $x \in G$.

Докажем, что операция на смежных классах определена корректно. Выберем представителей двух классов смежности

$$g_1' = g_1h_1, \qquad g_2' = g_2h_2.$$

Нужно доказать, что $g_1'g_2'$ принадлежит тому же смежному классу, что и g_1g_2 (это и означает, что результат операции — смежный класс — не зависит от выбора представителей смежных классов-сомножителей).

Проверяем:

$$g_1'g_2' = (g_1h_1)(g_2h_2) = g_1(h_1g_2)h_2 = g_1g_2h_1'h_2 \in (g_1g_2)H.$$

Последнее равенство выполняется для некоторого $h_1' \in H$ в силу равенства левых и правых смежных классов: $g_2H = Hg_2$.

Теперь проверим, что относительно этой операции смежные классы образуют группу.

Ассоциативность:

$$(g_1H)\cdot\big((g_2H)\cdot(g_3H)\big) = (g_1H)\cdot(g_2g_3)H = (g_1g_2g_3)H = \big((g_1H)\cdot(g_2H)\big)\cdot(g_3H).$$

Единичный элемент: это подгруппа H. Действительно, $H = eH$ и $(eH)\cdot(gH) = (eg)H = gH = (gH)(eH)$.

Обратный элемент к смежному классу gH — это $g^{-1}H$, так как $(gH)\cdot(g^{-1}H) = eH = H$. □

Из определения факторгруппы очевидно, что отображение

$$g \mapsto gH$$

является гомоморфизмом $G \to G/H$, причём ядро этого гомоморфизма равно H. Такой гомоморфизм называется *каноническим*.

Приведём несколько простых примеров факторгрупп.

Пример 4.28. Пусть G, H — некоторые группы. Множество $\{(x, y) : x = e_G, y \in H\}$, где e_G — нейтральный элемент группы G, является подгруппой прямого произведения $G \times H$.

Контрольный вопрос 4.29. Докажите это утверждение.

Будем обозначать эту подгруппу $e_G \times H$. Проверим, что она нормальная. Действительно,

$$(g,h)(e_G \times H) = \{(x,y) : x = g, y \in H\} = (e_G \times H)(g,h)$$

или словами: так как $hH = H = Hh$ для любого $h \in H$, левый смежный класс по $e_G \times H$ состоит из пар $(g,y) \in G \times H$, в которых фиксирован первый элемент, точно также устроены и правые смежные классы.

Докажем, что $(G \times H)/(e_G \times H) \cong G$. Это нетрудно вывести из предыдущего анализа: каждый смежный класс однозначно задаётся элементом $g \in G$ (представитель класса (g, e_H), где e_H — нейтральный элемент группы H), умножение этих смежных классов точно такое же, как в G: $(g_1, e_H) \cdot (g_2, e_H) = (g_1 g_2, e_H)$. □

Пример 4.30. Легко обобщить этот пример. Пусть K — нормальная подгруппа группы G. Тогда $K \times H$ — нормальная подгруппа группы $G \times H$. Действительно,

$$(g,h)(K \times H) = \{(x,y) : x = gK, y \in H\} = (K \times H)(g,h)$$

аналогично уже разобранному примеру.

В этом случае $(G \times H)/(K \times H) \cong G/K$, изоморфизм задаётся отображением

$$gK \mapsto (g, e_H)(K \times H).$$ □

Контрольный вопрос 4.31. Проверьте, что это отображение и впрямь изоморфизм (сохраняет операции и биективно).

Рассмотрим чуть более сложный пример, ограничившись частным случаем.

Пример 4.32. Есть ли у группы $C_6 \times C_9$ факторгруппа C_{18}?

Ответ: да.

Представим $C_6 \times C_9$ как $Z_6 \times Z_9$ и возьмём подгруппу $H = \langle (2,0) \rangle$. Она нормальная, так как $Z_6 \times Z_9$ абелева и её порядок равен 3. Докажем, что в $(Z_6 \times Z_9)/H$ есть элемент порядка 18, совпадающего с порядком такой факторгруппы. Это и означает, что факторгруппа циклическая порядка 18.

Искомый элемент факторгруппы — это смежный класс $[(3,1)]$. Порядок элемента $(3,1)$ в $Z_6 \times Z_9$ равен $\mathrm{НОК}(2,9) = 18$. Осталось проверить, что все положительные степени этого элемента, меньшие 18, не принадлежат подгруппе H. Чтобы $k(3,1) \in H$ необходимо, чтобы $k \equiv 0 \pmod 9$ (групповая операция в данном случае — сложение, в аддитивной записи степень записывается как кратное). Это оставляет единственный случай $k = 9$. Но $9(3,1) = (3,0) \notin H$ (так как 3 не является кратным 2 по модулю 6). Это и означает, что порядок элемента $[(3,1)]$ в факторгруппе $(Z_6 \times Z_9)/H$ равен 18. □

Пример 4.28 демонстрирует в наиболее буквальном виде суть слова «фактор» в термине «факторгруппа». Тут факторгруппа фактически является прямым множителем в группе (с точностью до изоморфизма, конечно). Но так дело обстоит

далеко не всегда, даже если группы абелевы. Напомним, что любая подгруппа абелевой группы нормальная — никакой разницы между левыми и правыми смежными классами нет.

Пример 4.33. Рассмотрим циклическую группу C_{p^2}, p — простое, порождающий элемент обозначим g. Подгруппа $H = \langle g^p \rangle$ имеет порядок p, поэтому $H \cong C_p$ (других групп простого порядка нет). Порядок группы G/H равен индексу H (количеству смежных классов), то есть также равен p. Значит, $G/H \cong C_p$. Но ясно, что C_{p^2} неизоморфна $C_p \times C_p$ (вторая группа не циклическая, порядки неединичных элементов в ней равны p). □

Однако в одном важном частном случае пример 4.28 оказывается исчерпывающим: когда порядок и индекс подгруппы абелевой группы взаимно просты.

Лемма 4.34. *Пусть H — такая подгруппа конечной абелевой группы G, что порядок H взаимно прост с её индексом. Тогда $G \cong H \oplus (G/H)$.*

Доказательство. Обозначим $q = |H|$, а через $G^{(q)}$ — подгруппу q-х степеней, то есть образ G при гомоморфизме $x \mapsto x^q$ возведения в степень q.

По утверждению 3.18 этот гомоморфизм является автоморфизмом G/H, так как порядок факторгруппы равен индексу H и по условию взаимно прост с q.

Докажем, что тогда $G^{(q)} \cap H = e$. Если $g^q \in H$, то $[g]^q = e_{G/H}$ (нейтральный элемент в факторгруппе — это сама подгруппа H, рассматриваемая как смежный класс по H). Отсюда следует, что $g \notin H$ влечёт $g^q \notin H$. Действительно, в терминах факторгруппы $[g]^q \neq e_{G/H}$, это выполняется, так как ядро гомоморфизма возведения в степень состоит только из единичного элемента, то есть подгруппы H.

Теперь проверим, что все элементы $G^{(q)}$ лежат в разных смежных классах по подгруппе H. Действительно, из равенства $g_1^q = g_2^q h$, $h \in H$, следует $\left(g_1 g_2^{-1}\right)^q = h$, а из предыдущего это означает, что $h = e$ и $g_1^q = g_2^q$.

Отсюда следует, во-первых, что каждый элемент группы G однозначно представляется в виде произведения $g^q h$, $g^q \in G^{(q)}$, $h \in H$. Поэтому корректно определена биекция

$$H \oplus G^{(q)} \mapsto G, \quad (h, g^q) \mapsto g^q h. \tag{4.1}$$

Как легко видеть, эта биекция сохраняет групповую операцию, поэтому является изоморфизмом.

Во-вторых, порядок $G^{(q)}$ равен $|G|/|H|$ (лемма 4.16), что совпадает с индексом H и порядком факторгруппы. Получаем взаимно однозначное соответствие между $G^{(q)}$ и G/H: элементу g^q сопоставляем смежный класс $g^q H$. Это соответствие является изоморфизмом, так как оно также сохраняет групповую операцию:

$$(g_1^q H) \cdot (g_2^q H) = (g_1^q g_2^q) H.$$

Из этих двух изоморфизмов получаем $G \cong (H \oplus G^{(q)}) \cong (H \oplus (G/H))$, что и требовалось. □

Для неабелевых групп лемма 4.34 не имеет места.

Пример 4.35. Рассмотрим диэдральную группу D_3. В ней есть циклическая подгруппа поворотов C_3 и индекс это подгруппы $|D_3|/|C_3| = 6/3 = 2$. Поэтому это нормальная подгруппа и факторгруппа $D_3/C_3 \cong C_2$ (других групп простого порядка нет). Хотя НОД$(2,3) = 1$, группы D_3 и $C_2 \oplus C_3$ неизоморфны: первая неабелева, а вторая абелева. □

Теперь сформулируем и докажем основную теорему о гомоморфизмах групп и факторгруппах.

Теорема 4.36 (теорема о гомоморфизмах)**.** *Пусть* $\varphi\colon G \to H$*. Тогда* $\operatorname{Im}\varphi \cong G/\operatorname{Ker}\varphi$*.*

Эта теорема говорит, что любой сюръективный гомоморфизм определяется ядром однозначно с точностью до изоморфизма. Действительно, применяя изоморфизм из теоремы, каждый сюръективный гомоморфизм φ представляется как композиция канонического гомоморфизма и некоторого изоморфизма факторгруппы по ядру φ и образа φ.

Доказательство теоремы 4.36. Построим отображение $\alpha\colon G/\operatorname{Ker}\varphi \to \operatorname{Im}\varphi$ по правилу

$$\alpha\colon (g \operatorname{Ker}\varphi) \mapsto \varphi(g).$$

Это правило корректно задаёт отображение, так как для всех элементов из смежного класса по ядру гомоморфизма φ образ при гомоморфизме φ один и тот же.

Биективность мы уже фактически проверили выше, в доказательстве утверждения 4.11. Проверим сохранение операции:

$$\alpha(g_1H \cdot g_2H) = \alpha((g_1g_2)H) = \varphi(g_1g_2) = \varphi(g_1)\varphi(g_2) = \alpha(g_1H)\alpha(g_2H).$$

Это означает, что построен изоморфизм $G/\operatorname{Ker}\varphi \to \operatorname{Im}\varphi$. □

Покажем как применяется теорема о гомоморфизмах к анализу существования сюръективных гомоморфизмов.

Пример 4.37. Существует ли сюръективный гомоморфизм группы $C_{36} \times C_{20}$ на группу C_8? Если существует, то по теореме 4.36 для некоторой подгруппы H факторгруппа $(C_{36} \times C_{20})/H$ изоморфна C_8. Но тогда вне этой подгруппы должен найтись элемент порядка, кратного 8. Это прообраз порождающего в C_8 при каноническом гомоморфизме. Действительно, в ядре гомоморфизма лежат в точности степени этого прообраза, кратные 8. Значит, и порядок кратен 8 (единичный элемент лежит в ядре).

Однако степень двойки, на которую делится порядок любого элемента из группы $C_{36} \times C_{20}$, не превосходит 4: это верно для обеих групп-сомножителей, значит, верно и для их прямого произведения (порядок элемента в прямом произведении равен НОК порядков компонент).

Поэтому сюръективного гомоморфизма группы $C_{36} \times C_{20}$ на группу C_8 не существует. □

Несюръективные гомоморфизмы устроены сложнее. Образ гомоморфизма может по-разному лежать в той группе, в которую действует гомоморфизм. Более точно это означает, что некоторые гомоморфизмы с одним и тем же ядром невозможно получить один из другого изоморфизмом групп, в которые действуют гомоморфизмы. Приведём пример.

Пример 4.38. Обозначим $G = Z_8 \times Z_{16}$ и рассмотрим два гомоморфизма $\varphi\colon Z_4 \to G$, $\psi\colon Z_4 \to G$. Группа Z_4 циклическая. Поэтому каждый такой гомоморфизм задаётся образом $[1] \in Z_4$, который может быть любым элементом в группе G, порядок которого является делителем 4.

Пусть $\varphi(1) = a = (2, 0)$, $\psi(1) = b = (4, 4)$ (проверьте, что порядки этих элементов группы G равны 4).

Ядро каждого из этих гомоморфизмов единичное, то есть это инъективные гомоморфизмы (ещё говорят, *вложения* группы Z_4). Но образы $\varphi(Z_4)$ и $\psi(Z_4)$ с алгебраической точки зрения различаются: нет автоморфизма группы G, переводящего один в другой.

Чтобы убедиться в этом, докажем неизоморфизм факторгрупп $G/\langle a\rangle$ и $G/\langle b\rangle$.

Первую из них легко выразить явно. Элементы (x, y), (x', y') группы G принадлежат одному классу смежности по подгруппе $\langle a\rangle$ тогда и только тогда, когда

$$(x, y) - (x', y') \in \langle a\rangle, \qquad \text{то есть } x - x' = 2k \pmod 8, \quad y - y' = 0k = 0 \pmod{16}.$$

Таким образом, класс смежности по подгруппе $\langle a\rangle$ однозначно задаётся парой (u, v), $u \in Z_2$, $v \in Z_{16}$. Сложение в факторгруппе $G/\langle a\rangle$ — это покомпонентное сложение элементов таких пар. Поэтому $G/\langle a\rangle \cong Z_2 \times Z_{16}$.

Со второй группой сложнее. Аналогичное рассуждение говорит, что элементы (x, y), (x', y') группы G принадлежат одному классу смежности по подгруппе $\langle b\rangle$ тогда и только тогда, когда

$$(x, y) - (x', y') \in \langle b\rangle, \qquad \text{то есть } x - x' = 4k \pmod 8, \quad y - y' = 4k = 0 \pmod{16}.$$

Однако предположение, что $G/\langle b\rangle \cong Z_4 \times Z_4$ как в первом случае, оказывается ложным. Хотя бы потому, что порядок факторгруппы равен $8 \cdot 16/4 = 32$, а не 16. Дело в том, что k в написанных выше условиях должно быть одним и тем же числом. В частности, пары $(3, 1)$ и $(7, -7)$ принадлежат разным смежным классам по подгруппе $\langle b\rangle$: не существует решения системы сравнений

$$-4 = 4k \pmod 8, \qquad 8 = 4k \pmod{16}.$$

Позже мы обсудим общие методы анализа структуры таких групп.

Однако убедиться в неизоморфизме групп $G/\langle a\rangle \cong Z_2 \times Z_{16}$ и $G/\langle b\rangle$ можно и проще. В первой группе есть элемент порядка 16, а во второй нет. Проверим, что восьмая степень любого элемента второй факторгруппы равна нейтральному элементу, то есть порядок любого элемента является делителем 8.

Пусть $[(x, y)] \in G/\langle b\rangle$ (квадратные скобки обозначают класс смежности с данным представителем). Операция у нас сложение, поэтому восьмая степень имеет вид $[(8x, 8y)]$. Докажем, что это элемент группы $\langle b\rangle$. Действительно, $(8x, 8y) = (0, 8y) =$

$= 2y(4,4)$ в группе G. Напомним, что сложение в первой компоненте происходит по модулю 8.

Итак, факторгруппы по подгруппам $G/\langle a\rangle$ и $G/\langle b\rangle$ неизоморфны. Поэтому, хотя сами подгруппы изоморфны, одна не переводится в другую автоморфизмом группы G. Это утверждение выглядит очевидным, но его нужно доказать, см. лемму ниже. □

Лемма 4.39. *Пусть $\alpha\colon G \to G$ — автоморфизм группы и $H \lhd G$. Тогда $\alpha(H)$ — нормальная и $G/H \cong G/\alpha(H)$.*

Доказательство. Первое утверждение вполне ясно: автоморфизм группы должен переводить подгруппу в подгруппу с теми же алгебраическими свойствами. Докажем это формально. Прежде всего проверим, что автоморфизм группы α переводит смежные классы по подгруппе H в смежные классы по подгруппе $\alpha(H)$. Для этого заметим, что $g\alpha(H) = \alpha(\alpha^{-1}(g)H)$ и аналогично для правого смежного класса: $(\alpha H)g = \alpha(H\alpha^{-1}(g))$. Поскольку левые и правые смежные классы по H совпадают, левые и правые смежные классы по $\alpha(H)$ также совпадают.

Теперь докажем второе утверждение. Построим отображение $\beta\colon G/H \to G/\alpha(H)$ по правилу

$$\beta\colon gH \mapsto \alpha(g)\alpha(H) = \alpha(gH).$$

Второе равенство объясняет корректность этого правила и взаимную однозначность отображения: автоморфизм (взаимно однозначное отображение) устанавливает взаимно однозначное соответствие между смежными классами и их образами.

Осталось проверить сохранение операции:

$$\begin{aligned}\beta((g_1H)(g_2H)) = \\ = \alpha(g_1g_2)\alpha(H) = \alpha(g_1)\alpha(g_2)\alpha(H)\alpha(H) = \alpha(g_1)\alpha(H)\alpha(g_2)\alpha(H) = \\ = \beta((g_1H))\beta((g_2H))\end{aligned}$$

(во второй строчке мы использовали, что группа $\alpha(H)$ нормальная и что $HH = H$ для любой группы). □

4.3 Сопряжённые элементы

Определение 4.40. Будем называть элемент b группы G *сопряжённым с* элементом a посредством элемента g, если $b = gag^{-1}$.

Два элемента группы называются *сопряжёнными*, если один сопряжён с другим посредством некоторого элемента g. Отношение сопряжённости будем обозначать $\sim$.

Свойство нормальности подгруппы переформулируется в терминах сопряжений: подгруппа H группы G нормальна тогда и только тогда, она не изменяется при сопряжении посредством любого элемента $g \in G$, то есть $gHg^{-1} = H$. Это условие очевидно равносильно совпадению левых и правых смежных классов $gH = Hg$ (умножение справа на g^{-1} превращает второе равенство в первое).

Лемма 4.41. *Отношение сопряжённости является отношением эквивалентности.*

Доказательство. Нужно проверить выполнение трёх свойств отношения эквивалентности для отношения сопряжённости.

Рефлексивность: $a \sim a$. Каждый элемент сопряжён сам с собой посредством нейтрального элемента e: $a = eae^{-1}$.

Симметричность: из $a \sim b$ следует $b \sim a$. Если b получается из a сопряжением посредством элемента g, то a получается из b сопряжением посредством g^{-1}:

$$b = gag^{-1} \quad \Leftrightarrow \quad a = g^{-1}bg$$

(умножение слева на g^{-1} и справа на g превращает первое равенство во второе).

Транзитивность: из $a \sim b$, $b \sim c$ следует $a \sim c$. Проверяется прямым вычислением: из $b = gag^{-1}$, $c = hbh^{-1}$ следует

$$c = hbh^{-1} = hgag^{-1}h^{-1} = (hg)a(hg)^{-1}. \tag{4.2}$$

□

Классом сопряжённости называется класс эквивалентности относительно отношения сопряжённости. Для абелевых групп классы сопряжённости тривиальные: каждый элемент сопряжён лишь самому себе, так как для любых x, y выполняется равенство $xyx^{-1} = xx^{-1}y = y$. Верно и обратное: если в группе есть нетривиальная пара сопряжённых элементов $x \neq y$, то x не коммутирует с тем g, сопряжение посредством которого переводит x в y. Другими словами, из неравенства $gxg^{-1} \neq x$ следует неравенство $gx \neq xg$ (умножение на g справа превращает первое неравенство во второе).

Контрольный вопрос 4.42. Подумайте, почему в этом рассуждении используется закон сокращения.

Поскольку для нормальной подгруппы $H \lhd G$ выполняется условие $gHg^{-1} = H$ при любом $g \in G$, она обязана состоят из объединения классов сопряжённости. Это даёт жёсткие ограничения на возможные нормальные группы в тех случаях, когда группа очень далека от абелевой (мало классов сопряжённости).

Основной для нас пример неабелевой группы — это группа перестановок S_n. Найдём для неё классы сопряжённости.

Неформально говоря, сопряжение посредством перестановки α — это «переименование» множества, которое переставляет перестановка. Число 1 называется теперь $\alpha(1)$, число $2 - \alpha(2)$, и т.д. Как изменяется при таком переименовании перестановка π? Перед общим рассуждением полезно рассмотреть пример.

Пример 4.43. Пусть переименование задаётся перестановкой

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 4 & 2 \end{pmatrix}.$$

Как изменится перестановка (1 2)(3 4)? Вместо 1 нужно написать 3, вместо $2 - 1$, вместо $3 - 4$, а вместо $4 - 2$. Получим перестановку (3 1)(4 2). □

В общем случае переименование α заменяет перестановку аналогично:

$$\text{из } \begin{pmatrix} 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \end{pmatrix} \quad \text{получается } \begin{pmatrix} \alpha(1) & \alpha(2) & \dots & \alpha(n) \\ \alpha(\pi_1) & \alpha(\pi_2) & \dots & \alpha(\pi_n) \end{pmatrix}. \tag{4.3}$$

Утверждение 4.44. *Перестановка, которая получается переименованием α из перестановки π, — это перестановка, сопряжённая π посредством α.*

Доказательство. Чтобы увидеть это утверждение наглядно, добавим ко второй таблице в (4.3) ещё две строки (первую таблицу в (4.3)):

$$\begin{pmatrix} \alpha(1) & \alpha(2) & \dots & \alpha(n) \\ 1 & 2 & \dots & n \\ \pi_1 & \pi_2 & \dots & \pi_n \\ \alpha(\pi_1) & \alpha(\pi_2) & \dots & \alpha(\pi_n) \end{pmatrix}.$$

Первые две строки — это таблица перестановки α^{-1}. Вторая и третья — таблица перестановки π. Третья и четвёртая — таблица перестановки α. Поэтому первая и четвёртая строки — это таблица перестановки $\alpha \circ \pi \circ \alpha^{-1}$, сопряжённой с π посредством α. □

Каждой перестановке сопоставим *цикловой тип* $(c_1, c_2, \dots, c_n)$, где c_i — количество циклов длины i в цикловом разложении. Другой способ задать цикловой тип — указать разбиение числа n в (неупорядоченную) сумму целых положительных слагаемых (длины циклов).

Лемма 4.45. *Перестановки сопряжены тогда и только тогда, когда их цикловые типы совпадают.*

Доказательство. Рассуждение полностью аналогично разобранному выше примеру. Поскольку сопряжённые перестановки получаются друг их друга переименованием, количество циклов данной длины не изменяется: при переименовании α цикл $(i_1\ i_2\ \dots\ i_\ell)$ переходит в цикл $(\alpha(i_1)\ \alpha(i_2)\ \dots\ \alpha(i_\ell))$.

В обратную сторону: пусть цикловые типы перестановок совпадают. Расположим циклы одинаковой длины друг под другом и получим правило переименования, которое переводит один цикл в другой. □

Пример 4.46. Найдём перестановку, сопряжение относительно которой переводит $(1\ 2\ 3)(4\ 5)$ в $(1\ 2)(3\ 4\ 5)$.

Нужно сопоставить элементы в циклах равной длины (и в том порядке, который задаётся циклом). Годятся перестановки

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 5 & 3 & 2 & 1 \end{pmatrix} \quad \text{или} \quad \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 4 & 5 & 1 & 2 \end{pmatrix},$$

но не годится, скажем,

$$\alpha = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 3 & 5 & 1 & 2 \end{pmatrix}.$$

(проверьте, что $\alpha \circ (1\ 2\ 3)(4\ 5) \circ \alpha^{-1} = (1\ 2)(4\ 5\ 3) \neq (1\ 2)(3\ 4\ 5)$). □

Итак, количество классов сопряжённости в группе перестановок S_n равно количеству разбиений числа n, которое обычно обозначается $p(n)$. Приведём таблицу первых значений этой функции:

n:	1	2	3	4	5	6	...
$p(n)$:	1	2	3	5	7	11	...

Из этой таблицы видно, что количество классов сопряжённости растёт гораздо медленнее общего количества перестановок в группе S_n. Асимптотически порядок роста $p(n) - 2^{O(\sqrt{n})}$, а порядок $n! - 2^{\Omega(n \log n)}$ (асимптотическое обозначение $\Omega(\cdot)$ аналогично обозначению $O(\cdot)$, только неравенство не сверху, а снизу).

Используя критерий сопряжённости перестановок (лемму 4.45) легко привести пример несопряжённых перестановок порядка 2. Например, перестановки (1 2) и (1 2)(3 4) в группе S_4. В некоторых случаях все перестановки одного порядка сопряжены.

Пример 4.47. Докажем, что все элементы порядка 42 сопряжены в группе S_{12}. Для этого найдём возможные цикловые типы перестановок порядка 42 в группе S_{12}.

Порядок перестановки равен НОК длин циклов. Поэтому цикл длины 7 обязан в такой перестановке быть (причём ровно один, так как $2 \cdot 7 = 14 > 12$). Аналогично, в перестановке порядка 42 нет цикла длины 6, так как $7 + 6 = 13 > 12$.

Поэтому в перестановке порядка 42 обязаны быть цикл длины 2 и цикл длины 3. Поскольку $2 + 3 + 7 = 12$, ничего больше быть не может.

Единственный возможный цикловой тип: $(2, 3, 7)$. Все перестановки такого типа сопряжены по лемме 4.45. □

Используя критерий сопряжённости перестановок, можно найти сильные ограничения на порядки нормальных подгрупп в S_n. Ограничимся примером.

Пример 4.48. Докажем, что единственная собственная (отличная от единичной и всей группы) нормальная подгруппа в S_5 — это знакопеременная группа A_5.

Всего есть 7 классов сопряжённости в S_5:

$5 = 5$	(циклы длины 5),	$\dfrac{5!}{5} = 24$ штуки;
$5 = 4 + 1$	(циклы длины 4),	$5 \cdot \dfrac{4!}{4} = 30$ штук;
$5 = 3 + 2$	(циклы длины 3 и 2),	$10 \cdot 2 = 20$ штук;
$5 = 3 + 1 + 1$	(циклы длины 3),	$10 \cdot 2 = 20$ штук;
$5 = 2 + 2 + 1$	(два цикла длины 2),	$5 \cdot 3 = 15$ штук;
$5 = 2 + 1 + 1 + 1$	(цикл длины 2),	$\binom{5}{2} = 10$ штук;
$5 = 1 + 1 + 1 + 1 + 1$	(тождественная),	1 штука.

(В вычислениях применяются простые комбинаторные соображения, проведите их самостоятельно.)

Поскольку нормальная группа состоит из объединения классов сопряжённости, её порядок должен быть суммой какого-то подмножества этих чисел, включать 1 (так как единичный элемент всегда входит в подгруппу) и делить 120 (порядок S_5) по теореме Лагранжа (а также отличаться от 120, поскольку нас интересуют собственные подгруппы). Одну такую сумму составить легко:

$$1 + 15 + 20 + 24 = 60,$$

это порядок $A_5 = 120/2 = 60$.

Других нормальных подгрупп порядка 60 в S_5 нет. Мы уже фактически проверили это, когда строили гомоморфизм знака перестановки (пример 4.1.2).

Ещё одна сумма, которая делит 120:

$$1 + 15 + 24 = 40.$$

Однако мы уже проверяли (пример 2.44), что подгрупп индекса 3 в группе S_5 нет.

Других сумм размеров классов сопряжённости, которые включают 1, делят 120, но не равны 120, 60 или 40, попросту нет. Это проверяется перебором, несколько утомительным, но прямолинейным.

Число 30 в такой набор входит не может, так как делители 120, которые не меньше 31, это 40, 60 и 120.

Пусть в набор входят 1 и 24. Тогда возможный делитель — это 30. Он отпадает, так как в оставшихся неиспользованными классах сопряжённости больше 5 элементов.

Пусть в набор входят 1 и 20. Единственный возможный делитель 120 в этом случае — опять 30. Он отпадает, так как в оставшихся классах сопряжённости больше 9 элементов.

Пусть в набор входят 1 и 15 или 10, но не входят 24, 20 и 30. Тогда делитель не получается: $1 + 10 + 15 \nmid 120$, $1 + 10 \nmid 120$, $1 + 15 \nmid 120$.

Разобрали все случаи. Других нормальных групп, кроме знакопеременной, нет по указанным арифметическим причинам. □

Задача 4.49. Предыдущим анализом мы исключили существование в S_5 нормальной подгруппы порядка 60, отличающейся от A_5. Докажите, что A_5 — единственная подгруппа порядка 60 в S_5.

Указание: проверьте, что любая подгруппа индекса 2 нормальная.

Группа, в которой нет нормальных подгрупп, кроме единичной и самой группы, называется *простой*. Поскольку ядро любого гомоморфизма является нормальной подгруппой, то сюръективных гомоморфизмов из простых групп крайне мало — всего два.

Самым очевидным примером простой группы является группа простого порядка. В ней вообще всего две подгруппы в силу теоремы Лагранжа.

Полная классификация простых групп — одно из самых выдающихся и технически сложных достижений математики XX века. Не претендуя даже на формулировку этой классификации, приведём один нетривиальный пример.

Группа перестановок S_n не является простой, так как содержит нормальную знакопеременную группу. Однако группа A_n при $n \geqslant 5$ уже является простой. Мы проверим это только для $n = 5$. (Заметим, что простота A_5 используется в некоторых результатах теоретической информатики, в других результатах приходится опираться на полную классификацию простых групп.)

Теорема 4.50. *Знакопеременная группа A_5 простая.*

Доказательство. Применим те же соображения, что в примере 4.48. Для этого нужно найти классы сопряжённости в A_5. Утверждение леммы 4.45 уже неприменимо буквально: в её доказательстве использовались для сопряжения все перестановки, а не только чётные. Ясно лишь, что классы сопряжённости в S_5 (быть может) распадаются на классы сопряжённости в A_5 (если даже в S_5 перестановки не сопряжены, то в A_5 они и подавно не сопряжены). Однако анализ, основанный на тех же идеях, вполне возможен.

Разберём по очереди все цикловые типы чётных перестановок.

$1+1+1+1+1$: это тождественная перестановка, которая, разумеется, образует класс сопряжённости.

$2+2+1$: пары циклов длины 2. Пусть один из этих циклов имеет вид $(i\ j)$. Если такая перестановка π сопряжена посредством какой-то перестановки α с перестановкой π', то она также сопряжена с π' и посредством перестановки $\alpha \circ (i\ j)$, чётность которой отличается от α. Действительно, перемена пары чисел в цикле длины 2 не изменяет этот цикл. Итак, все перестановки типа $2+2+1$ сопряжены какой-то чётной перестановкой. Значит, они сопряжены в A_5.

$3+1+1$: циклы длины 3. Пусть цикл π длины 3 не включает числа i, j. Если π сопряжён посредством какой-то перестановки α с перестановкой π', то он также сопряжён с π' и посредством перестановки $\alpha \circ (i\ j)$, чётность которой отличается от α. Действительно, перемена пары чисел вне цикла не изменяет цикл. Итак, все перестановки типа $3+1+1$ сопряжены какой-то чётной перестановкой. Значит, они сопряжены в A_5.

5: циклы длины 5. Тут возникает трудность. Перестановки, сопрягающие пару таких циклов

$$(i_1\ i_2\ i_3\ i_4\ i_5), \quad (j_1\ j_2\ j_3\ j_4\ j_5),$$

могут различаться только на цикл длины 5 (чтобы цикл переходил в цикл, должен сохраняться цикловой порядок на числах, а любая степень цикла длины 5 либо тождественная, либо цикл длины 5).

Поскольку цикл длины 5 является чётной перестановкой, то любая пара циклов длины 5 сопряжена либо только посредством чётных перестановок, либо только посредством нечётных. Это значит, что циклы длины 5 распадаются на два класса сопряжённости в группе A_5.

Итак, размеры классов сопряжённости в A_5 это 1, 12, 12, 15, 20. Суммы подмножеств этих чисел, которые делят 60 и включают 1, легко перечислить. Это 1 и $1+12+12+15+20$.

Проверим это перебором, аналогичным перебору в примере 4.48.

Пусть такая сумма содержит 1 и 20. Собственный делитель 60, не меньший 21, это 30. Но все оставшиеся классы сопряжённости содержат больше 9 элементов.

Пусть сумма содержит 1 и 15, не содержит 20. Все делители 60, которые не меньше 16, делятся на 10. Чтобы добавить к 1 и 15 числа из списка так, чтобы последняя цифра суммы равнялась 0, нужно добавить оба числа 12. Получаем 40, который не является делителем 60.

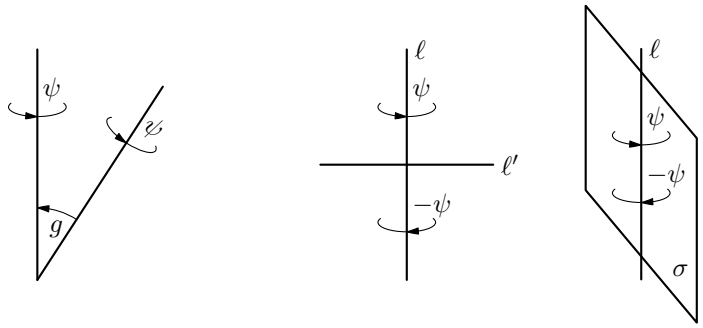
Пусть сумма содержит 1 и 12, не содержит 15 и 20. Вариантов всего два: $1+12 = 13 \nmid 60$ и $1 + 12 + 12 = 25 \nmid 60$.

Итак, возможные порядки нормальных подгрупп A_5 — это 1 и 60. То есть несобственных нормальных подгрупп нет и A_5 простая группа. □

Аналогично симметрической группе устроены сопряжения в группах движений. Из геометрических соображений ясно, что движения a и b сопряжены ($a = gbg^{-1}$), когда это одно и то же преобразование, выполненное в двух разных системах координат (преобразованием g^{-1} перешли в старую систему координат, применили b, вернулись преобразованием g в новую систему координат и получили преобразование a, которое в старой системе координат записывается так же, как b в новой). В частности, движение xax^{-1}, сопряжённое с поворотом a вокруг прямой ℓ, является поворотом на такой же угол вокруг той прямой ℓ_x, в которую переходит прямая ℓ при движении x ($\ell_x = x(\ell)$).

В силу этих рассуждений повороты на один и тот же угол вокруг двух разных осей сопряжены, если в группе есть преобразование g, переводящее одну ось в другую, как показано на рис. 21 (а). Если речь идёт о поворотах вокруг одной оси ℓ на одинаковые по абсолютной величине углы ψ, $-\psi$ (но в разные стороны), то сопрягающих преобразований g может быть всего два: поворот на 180° вокруг прямой ℓ', перпендикулярной ℓ, или отражение относительно плоскости σ, проходящей через прямую ℓ, рис. 21 (б).

Пример 4.51. Найдём сопряжённые элементы в группе диэдра D_n. Для этого удобно представлять симметрии правильного n-угольника как повороты в трёхмерном пространстве, в которое вложен этот n-угольник. Как мы уже выясняли ранее в



(а) (б)

Рис. 21: Сопряжения поворотов

примере 2.9, эта группа состоит из из поворотов на угол $\varphi = 2\pi/n$ вокруг прямой c, проходящей через центр многоугольника и перпендикулярной его плоскости, и поворотов на угол 180° относительно прямых, проходящих через вершину многоугольника и его центр (случай нечётного n), а в случае чётного n — ещё добавляются повороты на 180° относительно прямых, проходящих через центр многоугольника и середину одной из его сторон. Этим поворотам на 180° отвечают несобственные симметрии правильного n-угольника на плоскости (отражения относительно соответствующих прямых).

Заметим, что повороты на разные по абсолютной величине углы не сопряжены друг с другом. А повороты на одинаковые углы, но в противоположных направлениях, сопряжены в D_n, так как сопрягаются поворотом на угол 180° относительно прямой, проходящей через вершину многоугольника и его центр.

Повороты на угол 180° вокруг осей, лежащих в плоскости многоугольника, сопряжены тогда и только тогда, когда эти оси можно совместить собственной симметрией диэдра.

В случае нечётного n любые две такие оси можно совместить подходящим поворотом вокруг c. Поэтому все такие элементы группы диэдра образую один класс сопряжённости. Всего получаем получаем $k + 2$ классов сопряжённых элементов в D_{2k+1}: в одном только тождественное преобразование, в k — пары поворотов относительно c на противоположные углы, в последнем, $(k+2)$-м, — все повороты вокруг осей, лежащих в плоскости многоугольника.

В случае чётного n поворотом вокруг c можно совместить либо те оси, которые проходят через пару противоположных вершин, либо те, которые проходят через середины противоположных сторон. Поскольку вершины при симметрии многоугольника переходят в вершины, между собой эти оси совместить нельзя. Поэтому получаем $k + 3$ классов сопряжённых элементов в D_{2k}: в одном только тождественное преобразование, в $k - 1$ — пары поворотов относительно c на противоположные углы, ещё в одном — поворот на 180° относительно оси c (при такой чётности это симметрия, нужно ещё заметить, что угол 180° противоположен самому себе), два последних класса содержат по k поворотов на угол 180° вокруг осей, лежащих в плоскости многоугольника. □

4.4 Сопряжённость и внутренние автоморфизмы

Сопряжения имеют прямое отношение к автоморфизмам групп. В примере 3.14 был приведён важный пример автоморфизма группы перестановок. Он обобщается на произвольные группы.

Определение 4.52. Пусть g — некоторый фиксированный элемент группы G. Отображение $G \to G$, задаваемое сопряжением посредством g,

$$x \mapsto gxg^{-1}, \tag{4.4}$$

называется *внутренним автоморфизмом* группы G.

Автоморфизмы, не являющиеся внутренними, называются *внешними*.

Утверждение 4.53. *Отображение* (4.4) *является автоморфизмом группы.*

Доказательство. Повторим рассуждения из примера 3.14 в общем случае (см. также лемму 4.41).

Биективность. Обратное отображение задаётся формулой $x \mapsto g^{-1}xg$, что проверяется прямым вычислением

$$g^{-1}\big(gxg^{-1}\big)g = (g^{-1}g)x(g^{-1}g) = x.$$

Сохранение операции также проверяется прямым вычислением

$$g(xy)g^{-1} = g(xgg^{-1}y)g^{-1} = (gxg^{-1})(gyg^{-1}),$$

это и означает, что образ произведения равен произведению образов. □

Таким образом, нормальная подгруппа — это та подгруппа, которая сохраняется при внутренних автоморфизмах (однако может измениться при внешних). Сопряжённые элементы переводятся один в другой посредством автоморфизма группы, поэтому их алгебраические свойства одинаковы.

Контрольный вопрос 4.54. Докажите, что если x и y сопряжены, то их порядки одинаковы.

Утверждение 4.55. *Внутренние автоморфизмы образуют нормальную подгруппу всей группы автоморфизмов (групповая операция — композиция).*

Доказательство. Как было показано выше, обратное к сопряжению посредством g — это сопряжение посредством g^{-1}.

Тождественный автоморфизм представляется как сопряжение нейтральным элементом, $exe^{-1} = x$ для любого x.

Композиция сопряжений посредством элементов g и h — это сопряжение посредством их произведения:

$$g\big(hxh^{-1}\big)g^{-1} = (gh)x(gh)^{-1}.$$

Проверим, что подгруппа внутренних автоморфизмов нормальная. Рассмотрим автоморфизм $\alpha\colon G \to G$ группы G. Тогда сопряжение посредством этого автоморфизма некоторого внутреннего автоморфизма также является внутренним автоморфизмом:

$$\alpha\big(g\alpha^{-1}(x)g^{-1}\big) = \alpha(g)x\alpha(g^{-1}) = \alpha(g)x\alpha(g)^{-1}$$

(проследите за этой выкладкой внимательно и проверьте, что первое выражение и впрямь задаёт образ x при композиции обратного к α автоморфизма, сопряжения посредством g и автоморфизма α). □

Если группа G коммутативна, то единственный её внутренний автоморфизм — тождественное отображение. Для некоммутативных групп существуют нетривиальные внутренние автоморфизмы, как мы уже видели в примере 3.14.

Пример 4.56. В любом случае количество внутренних автоморфизмов конечной группы делит порядок группы. Вычисления в доказательстве утверждения 4.55 показывают, что отображение $G \to \operatorname{Aut} G$, которое элементу группы сопоставляет сопряжение этим элементом, является гомоморфизмом. Порядок образа при гомоморфизме, то есть порядок группы внутренних автоморфизмов, делит порядок группы. □

Пример 4.57. Интересно также и ядро гомоморфизма, описанного в предыдущем примере. Это *центр* группы G, то есть множество $Z(G)$ тех её элементов, которые коммутируют со всеми элементами группы:

$$Z(G) = \{x \in G : gx = xg \text{ для всех } g \in G\}.$$

Ясно, что сопряжение посредством элемента из центра является тождественным автоморфизмом группы. Поэтому центр группы — это как раз ядро гомоморфизма, переводящего элемент группы в сопряжение посредством этого элемента.

Если группа абелева, то $Z(G) = G$. □

Пример 4.58. Основной для нас пример неабелевой группы — это симметрическая группа S_n при $n \geqslant 3$ (легко видеть, что группы S_1 и S_2 абелевы).

Центр S_n при $n \geqslant 3$ состоит из только из тождественной перестановки. Действительно, пусть $\alpha \in S_n$ нетождественная. Тогда в ней есть цикл длины $\ell \geqslant 2$

$$(i_1\ i_2\ \ldots\ i_\ell).$$

Рассмотрим два случая.

Первый случай: $\ell = 2$. Обозначим через j какой-нибудь элемент, отличный от i_1, i_2 (напомним, что $n \geqslant 3$, так что такой элемент найдётся). Тогда сопряжение транспозиции $(i_1\ j)$ посредством α равно $(i_2\ \alpha(j))$, что не равно $(i_1\ j)$, так как $i_1 \neq i_2$ по построению и $i_1 \neq \alpha(j)$ из-за биективности α. Это означает, что сопряжение посредством α не является тождественным преобразованием (равносильно тому, что α не лежит в центре S_n).

Второй случай: $\ell \geqslant 3$. Тогда сопряжение транспозиции $(i_1\ i_\ell)$ посредством α равно $(i_2\ i_1)$, что не равно $(i_1\ i_\ell)$, так как $\ell \geqslant 3$. И в этом случае сопряжение посредством α не является тождественным преобразованием, то есть α не лежит в центре S_n. □

Факторгруппа $G/Z(G)$ изоморфна группе внутренних автоморфизмов. Оказывается, для неабелевых групп ($Z(G) \neq G$) эта группа не может быть циклической.

Теорема 4.59. *Если группа неабелева, её группа внутренних автоморфизмов не является циклической.*

Доказательство. Пусть $G/Z(G)$ циклическая. Обозначим её порождающий элемент $[g]$ (класс смежности, содержащий $g \in G$). Тогда любой элемент группы представляется как $g^i z$, $z \in Z(G)$. Докажем, что все такие элементы коммутируют, то есть группа абелева.

Прямое вычисление, использующее определение центра группы:

$$(g^i z') \cdot (g^k z'') = g^i (z' g^k) z'' = g^i g^k (z' z'') = g^{i+k} z'' z' = \cdots = (g^k z'') \cdot (g^i z').$$

(Заполните многоточия аналогично написанным равенствам.) □

Пример 4.60. Решим такую задачу. В группе порядка 720 нашлось 190 элементов, каждый из которых сам по себе образует класс сопряжённости. Докажите, что группа абелева.

Элементы, которые сами по себе образуют класс сопряжённости, и элементы центра — это одно и то же. Условие задачи говорит, что в центре не меньше 190 элементов. Но тогда его индекс не больше 3. Все группы порядка 2 и 3 — циклические, поэтому индекс центра никогда не равен 2 или 3 по предыдущей теореме. Остался единственный возможный вариант — 1. Поэтому центр совпадает со всей группой и группа абелева. □

5 Задание группы порождающими и соотношениями

Множество $S \subset G$ называется *порождающим множеством* для группы G, если $\langle S \rangle = G$, то есть все элементы группы представляются в виде произведений элементов из S и обратных к ним (дальше для краткости говорим о произведениях порождающих).

Некоторые произведения порождающих равны единичному элементу. Есть два принципиально разных случая.

Во-первых, может так случиться, что произведение равно единичному элементу в любой группе, как например

$$g \cdot g^{-1}, \quad ghg^{-1}h^{-1}hgh^{-1}g^{-1}$$

(первый пример очевиден; проверьте, что и второе произведение равно единичному элементу в любой группе для любых g, h). Будем называть такие произведения групповыми тождествами.

Во-вторых, некоторые произведения порождающих равны единичному элементу в данной группе, хотя они и не являются тождествами. Например, группа перестановок S_n порождается транспозициями и для каждой транспозиции t выполняется равенство $t^2 = e$, которое, конечно же, не является групповым тождеством (даже в группе перестановок есть элементы порядка больше 2). Такие произведения будем называть *соотношениями*.

В этой главе мы рассмотрим способ задания групп указанием порождающих и соотношений между ними. Оказывается, в этом случае группа представляется как факторгруппа некоторой особой группы, называемой *свободной*.

Для простоты изложения мы ограничиваемся лишь лишь конечно-порождёнными группами.

Группа называется *конечно-порождённой*, если у неё есть конечное порождающее множество. Всякая конечная группа конечно порождена. Группы $\mathbb{Z}^n$ также конечно порождены. Порождающим множеством для $\mathbb{Z}^n$ является, например,

$$\{(1,0,\ldots,0);\ (0,1,\ldots,0);\ \ldots;\ (0,0,\ldots,1)\}. \tag{5.1}$$

Мы уже сталкивались с группами, которые не являются конечно-порождёнными. Это числовые группы рациональных, действительных и комплексных чисел (как аддитивные, так и мультипликативные). Разберём лишь один пример такого сорта.

Пример 5.1. Докажем, что группа $(\mathbb{Q}, +)$ рациональных чисел по сложению не является конечно-порождённой.

Выберем какое-нибудь конечное множество рациональных чисел

$$S = \left\{ \frac{a_1}{b_1},\ \frac{a_2}{b_2},\ \ldots,\ \frac{a_n}{b_n} \right\}.$$

Приведём все дроби к общему знаменателю, получим представление тех же чисел в виде

$$S = \left\{ \frac{A_1}{M},\ \frac{A_2}{M},\ \ldots,\ \frac{A_n}{M} \right\}.$$

Группа $(\mathbb{Q}, +)$ абелева, операцию мы записываем аддитивно. Поэтому любой элемент из подгруппы $\langle S \rangle$ выражается как целочисленная линейная комбинация порождающих:

$$\sum_{i=1}^{n} c_i \frac{A_i}{M}\,, \quad c_i \in \mathbb{Z}.$$

Это означает, что все рациональные числа из подгруппы $\langle S \rangle$ представляются обыкновенными дробями, знаменатели которых являются делителями M (что-то в знаменателе M может сократиться, но в любом случае останется делитель M).

Простых чисел, как мы знаем, бесконечно много (см. теорему 1.56). А простых делителей M конечное количество. Поэтому найдётся простое число p, которое не является делителем M. В силу сделанных выше наблюдений $1/p \notin \langle S \rangle$.

Итак, для каждого конечного множества $S \subset \mathbb{Q}$ мы нашли элемент, не принадлежащий подгруппе, порождённой S. Это и значит, что группа не является конечно-порождённой. □

5.1 Свободные абелевы группы

Начнём со случая абелевых групп, который проще общего.

В случае абелевых групп часто используется аддитивная запись для групповой операции. При такой записи степень элемента g записывается как целочисленное кратное ng, здесь $n \in \mathbb{Z}$ — показатель степени. Кроме того, при использовании аддитивной записи мы будем записывать прямое произведение групп как *прямую сумму* $G \oplus H$ (аналогично для нескольких слагаемых). Например,

$$\mathbb{Z}^n = \underbrace{\mathbb{Z} \oplus \mathbb{Z} \oplus \cdots \oplus \mathbb{Z}}_{n \text{ прямых слагаемых}},$$

где $\mathbb{Z}^n$ — группа целочисленных n-мерных векторов с операцией покомпонентного сложения.

Замечание 5.2. Удобно рассматривать также группу $\mathbb{Z}^0$, которая по определению состоит из одного элемента 0.

Теорема 5.3. *Любая конечно-порождённая абелева группа изоморфна факторгруппе группы* $\mathbb{Z}^n$.

Доказательство. Повторим в общем виде рассуждение, которое уже использовалось ранее в анализе конкретных примеров. Пусть $g_1, \ldots, g_n$ — множество порождающих некоторой абелевой группы G. В силу коммутативности групповой операции любое произведение g_i и их обратных равно выражению вида

$$g_1^{a_1} g_2^{a_2} \ldots g_n^{a_n}, \qquad a_i \in \mathbb{Z}.$$

При аддитивной записи групповой операции получаем

$$a_1 g_1 + a_2 g_2 + \cdots + a_n g_n.$$

Отображение

$$(a_1, a_2, \ldots, a_n) \mapsto a_1 g_1 + a_2 g_2 + \cdots + a_n g_n \tag{5.2}$$

является гомоморфизмом $\mathbb{Z}^n \to G$ и этот гомоморфизм сюръективен.

Из теоремы о гомоморфизмах групп заключаем, что G изоморфна факторгруппе $\mathbb{Z}^n$ по ядру этого гомоморфизма. □

Пример 5.4. Рассмотрим прямую сумму циклических групп $Z_{n_1} \oplus \cdots \oplus Z_{n_t}$. Она порождается множеством вида (5.1):

$$e_1 = (1, 0, \ldots, 0);\ e_2 = (0, 1, \ldots, 0);\ \ldots;\ e_t = (0, 0, \ldots, 1),$$

поскольку 1 является порождающей для аддитивной группы вычетов по модулю любого числа.

Что будет ядром гомоморфизма (5.2) для такой системы порождающих? Если

$$a_1 g_1 + a_2 g_2 + \cdots + a_n g_n = (0, 0, \ldots, 0),$$

то a_1 кратно n_1, a_2 кратно n_2 и т.д. Это означает, что для любого $(a_1, \ldots, a_t)$ из ядра гомоморфизма (5.2) выполняется равенство

$$(a_1, \ldots, a_t) = x_1(n_1, 0, \ldots, 0) + x_2(0, n_2, \ldots, 0) + \cdots + x_t(0, 0, \ldots, n_t), \quad x_i \in \mathbb{Z},$$

и для любого набора коэффициентов x_i выписанная в правой части линейная комбинация принадлежит ядру. Получаем изоморфизм ядра и $\mathbb{Z}^n$. □

Группы $\mathbb{Z}^n$ являются *свободными*: у них есть такая система порождающих (множество единичных векторов (5.1)), для которой описанный выше гомоморфизм имеет нулевое ядро и является тем самым изоморфизмом.

Как мы знаем, факторгруппа однозначно определяется ядром канонического гомоморфизма, то есть некоторой нормальной подгруппой. В случае абелевых групп все подгруппы нормальные. Поэтому для описания конечно-порождённых абелевых групп нужно знать, какие бывают подгруппы у $\mathbb{Z}^n$. На этот вопрос легко ответить.

Лемма 5.5. *Пусть $R < \mathbb{Z}^n$. Тогда $R \cong \mathbb{Z}^m$, $0 \leqslant m \leqslant n$.*

Доказательство. Индукция по n. База индукции $n = 1$ говорит, что всякая собственная подгруппа $\mathbb{Z}$ (бесконечной циклической группы) изоморфна $\mathbb{Z}$. Это часть утверждения леммы 3.29, которое мы доказывали выше.

Индуктивный переход. Предположим, что утверждение леммы верно при всех $n' < n$. Рассмотрим подгруппу $R < \mathbb{Z}^n$. Выделим в ней подгруппу

$$R_0 = \{x \in R : x_n = 0\}$$

(проверьте, что это и в самом деле подгруппа!).

Поскольку R_0 изоморфна подгруппе $\mathbb{Z}^{n-1}$ (равные нулю последние компоненты можно опустить), то по предположению индукции $R \cong \mathbb{Z}^k$, $0 \leqslant k \leqslant n-1$.

Если $R_0 = R$, утверждение леммы доказано.

В противном случае рассмотрим подмножество $\tilde{R}_n$ целых чисел, состоящее из последних компонент элементов $r \in R$. Это множество является нетривиальной

подгруппой $\mathbb{Z}$ (содержит не только 0). По лемме 3.29 это бесконечная циклическая группа с порождающим $d \neq 0$.

Выберем какой-нибудь $r = (r_1, r_2, \ldots, r_{n-1}, d) \in R$, у которого последняя компонента равна d. Докажем, что любой $x \in R$ однозначно представляется суммой $x_0 + yr$, $x_0 \in R_0$, $y \in \mathbb{Z}$. Действительно, $x_n = yd$ для некоторого $y \in \mathbb{Z}$ по построению, так что коэффициент y определяется однозначно. Но тогда $x_0 = x - yr \in R_0$ (в последней компоненте стоит $x_n - yd = 0$).

Теперь докажем, что $R \cong R_0 \oplus \langle r \rangle$. Построим гомоморфизм $\alpha\colon R_0 \oplus \langle r \rangle \to R$ по правилу

$$\alpha\colon (g, tr) \mapsto g + tr.$$

(Этот гомоморфизм аналогичен общей конструкции (5.2) гомоморфизма $\mathbb{Z}^n$ на конечно-порождённую группу.)

Из доказанной выше однозначности разложения элементов R в сумму $x_0 + yr$, $x_0 \in R_0$, $y \in \mathbb{Z}$, следуют и инъективность, и сюръективность этого гомоморфизма. Сюръективность прямо следует из приведённого выше рассуждения. Если бы гомоморфизм α не был инъективным, то нарушилась бы однозначность для нейтрального элемента подгруппы.

Итак,

$$R \cong R_0 \oplus \langle r \rangle \cong \mathbb{Z}^k \oplus \mathbb{Z} \cong \mathbb{Z}^m,$$

где $0 \leqslant m \leqslant n$.

Лемма теперь следует по принципу математической индукции. □

5.2 Матрицы и подгруппы $\mathbb{Z}^n$. Элементарные преобразования матриц

Итак, лемма 5.5 говорит, что любая подгруппа $G < \mathbb{Z}^n$ изоморфна некоторой группе $\mathbb{Z}^k$. Выделим в этой подгруппе те элементы b_1, b_2, …, b_k, которым изоморфизм с $\mathbb{Z}^k$ сопоставляет векторы из базиса (5.1). Это n-мерные векторы. Записывая их координаты в строки, получаем целочисленную матрицу размера $k \times n$

$$M(G) = \begin{pmatrix} b_{11} & b_{12} & \ldots & b_{1n} \\ b_{21} & b_{22} & \ldots & b_{2n} \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots\ldots \\ b_{k1} & b_{k2} & \ldots & b_{kn} \end{pmatrix}.$$

И обратно, любой матрице M размера $k \times n$ отвечает подгруппа $R(M) = \mathbb{Z}^n$, состоящая из целочисленных линейных комбинаций строк матрицы. Конечно, такая подгруппа не обязательно изоморфна $\mathbb{Z}^k$. Например, если все строки одинаковы, то получится циклическая подгруппа $\mathbb{Z}^n$, которая изоморфна $\mathbb{Z}^1$.

Пример 5.6 (продолжение примера 5.4)**.** Какая матрица задаёт подгруппу G, порождённую векторами

$$(n_1, 0, \ldots, 0) + x_2(0, n_2, \ldots, 0) + \cdots + x_t(0, 0, \ldots, n_t)?$$

(В примере 5.4 мы установили, что $\mathbb{Z}^t/G \cong C_{n_1} \oplus \cdots \oplus C_{n_t}$.)

Это диагональная матрица

$$\begin{pmatrix} n_1 & 0 & \dots, & 0 \\ 0 & n_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & n_t \end{pmatrix}$$

□

Одну и ту же подгруппу можно задавать разными матрицами. Начнём с двух примеров.

Пример 5.7. Рассмотрим две матрицы

$$M_1 = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{pmatrix}, \qquad M_2 = \begin{pmatrix} -b_{11} & -b_{12} & \dots & -b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \\ \dots & \dots & \dots & \dots \\ b_{k1} & b_{k2} & \dots & b_{kn} \end{pmatrix}$$

и докажем, что они задают одну и ту же подгруппу: $R(M_1) = R(M_2)$.

Пусть $x \in R(M_1)$. По определению это означает, что x представляется целочисленной линейной комбинацией строк матрицы M_1:

$$x = x_1 b_1 + x_2 b_2 + \dots + x_k b_k, \quad x_i \in \mathbb{Z}.$$

Но тогда

$$x = (-x_1) \cdot (-b_1) + x_2 b_2 + \dots + x_k b_k,$$

то есть x является целочисленной линейной комбинацией строк матрицы M_2. Значит, $x \in R(M_2)$.

Мы доказали включение $R(M_1) \subseteq R(M_2)$. Обратное включение доказывается аналогично (можно ещё заметить, что перестановка индексов матриц ничего не меняет в этом рассуждении, а первое включение превращается во второе). □

Пример 5.8. Рассмотрим две матрицы

$$M_1 = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} & b_{22} & \dots & b_{2n} \end{pmatrix}, \quad M_2 = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1n} \\ b_{21} + t b_{11} & b_{22} + t b_{12} & \dots & b_{2n} + t b_{1n} \end{pmatrix},$$

где $t \in \mathbb{Z}$, и докажем, что они задают одну и ту же подгруппу: $R(M_1) = R(M_2)$.

Для строк b_1, b_2 матрицы M_1 и b'_1, b'_2 матрицы M_2 выполняются равенства

$$b'_1 = b_1, \quad b'_2 = b_2 + t b_1; \qquad b_1 = b'_1, \quad b_2 = b'_2 - t b'_1.$$

Пусть $x \in R(M_1)$, то есть $x = x_1 b_1 + x_2 b_2$, $x_i \in \mathbb{Z}$. Тогда

$$x = x_1 b'_1 + x_2 (b'_2 - t b'_1) = (x_1 - t x_2) b'_1 + x_2 b'_2 \in R(M_2),$$

то есть $R(M_1) \subseteq R(M_2)$.

Аналогично для $x \in R(M_2)$: пусть $x = x_1 b'_1 + x_2 b'_2$, $x_i \in \mathbb{Z}$. Тогда

$$x = x_1 b_1 + x_2 (b_2 + t b_1) = (x_1 + t x_2) b'_1 + x_2 b_2 \in R(M_1),$$

то есть $R(M_2) \subseteq R(M_1)$.

Два доказанных включения равносильны равенству подгрупп. □

Ясно также, что перестановка строк матрицы не изменяет задаваемую ей подгруппу.

Будем называть перестановки строк, умножение строки на -1 и прибавление к строке целого кратного другой строки *элементарными строчными преобразованиями* целочисленных матриц.

По сути мы доказали лемму.

Лемма 5.9. *Если одна матрица получается из другой последовательностью элементарных строчных преобразований, то эти две матрицы задают одну и ту же подгруппу.*

Контрольный вопрос 5.10. Запишите аккуратное доказательство леммы.

Аналогично элементарным строчным преобразованиям матриц определим *элементарные столбцовые преобразования*: перестановки столбцов, умножение столбца на -1 и прибавление к столбцу целого кратного другого столбца.

Столбцовые преобразования уже не сохраняют подгруппу, задаваемую матрицей.

Пример 5.11. Рассмотрим две матрицы

$$M_1 = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}, \quad M_2 = \begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}.$$

Матрица M_2 получается из M_1 прибавлением к первому столбцу второго.

Эти матрицы задают разные подгруппы. В целочисленной линейной комбинации строк первой матрицы первая координата всегда чётна. Для второй матрицы это не так (вторая строка имеет нечётную первую координату). □

Однако группы, задаваемые матрицами, полученными одна из другой столбцовыми преобразованиями, «похожи». Точный смысл этого утверждения состоит в том, что одна переводится в другую автоморфизмом $\mathbb{Z}^n$.

Нам не потребуется вся группа автоморфизмов $\mathbb{Z}^n$. Достаточно будет лишь некоторых конкретных автоморфизмов.

Любое преобразование $\mathbb{Z}^n$, задаваемое линейными формулами, является гомоморфизмом: образ суммы равен сумме образов, что следует из дистрибутивности сложения и умножения целых чисел.

Утверждение 5.12. *Следующие отображения являются автоморфизмами $\mathbb{Z}^n$:*

- *перестановка координат;*
- *умножение одной из координат на -1:*
$$(x_1, \ldots, x_i, \ldots, x_n) \mapsto (x_1, \ldots, -x_i, \ldots, x_n);$$
- *прибавление к одной координате целого кратного другой*
$$(x_1, \ldots, x_i, \ldots, x_j, \ldots x_n) \mapsto (x_1, \ldots, x_i + tx_j, \ldots, x_j, \ldots, x_n).$$

Доказательство. Эти преобразования линейные, то есть являются гомоморфизмами $\mathbb{Z}^n$. Их биективность легко проверяется. Обратные преобразования задаются соответственно

- обратной перестановкой координат;
- умножением той же координаты x_i на -1;
- формулой

$$(x_1, \dots, x_i, \dots, x_j, \dots x_n) \mapsto (x_1, \dots, x_i - tx_j, \dots, x_j, \dots, x_n).$$

Таким образом, все эти преобразования являются автоморфизмами $\mathbb{Z}_n$. □

Лемма 5.13. *Если матрица M_1 получается из матрицы M_2 последовательностью элементарных столбцовых преобразований, то существует автоморфизм $\mathbb{Z}^n$, который переводит подгруппу $R(M_1)$ в подгруппу $R(M_2)$.*

Доказательство. Композиция автоморфизмов является автоморфизмом, поэтому достаточно проверить утверждение леммы для одного элементарного столбцового преобразования.

Необходимые автоморфизмы уже описаны в утверждении 5.12. Соответствие между элементарными столбцовыми преобразованиями и автоморфизмами из утверждения 5.12 буквальное: перестановке столбцов отвечает перестановка координат (ведь каждый столбец матрицы отвечает некоторой координате); умножению столбца на -1 отвечает умножение соответствующей координаты на -1; прибавлению к i-му столбцу целого кратного j-го столбца отвечает прибавление к координате i целого кратного координаты j.

Каждая строка M_2 получается из соответствующей строки M_1 применением одного и того же автоморфизма α. Поэтому множество векторов, которые являются целочисленными линейными комбинациями строк $\alpha(b_i)$ матрицы M_2 в точности совпадает с образом при автоморфизме α множества векторов, которые являются целочисленными линейными комбинациями строк b_i матрицы M_1:

$$x = \sum_i x_i b_i \quad \Leftrightarrow \quad \alpha(x) = \alpha\left(\sum_i x_i b_i\right) = \sum_i x_i \alpha(b_i).$$

□

Применяя лемму 4.39, получаем такие следствия.

Следствие 5.14. *Если матрица M_1 получается из матрицы M_2 последовательностью элементарных столбцовых преобразований, то $\mathbb{Z}^n/R(M_1) \cong \mathbb{Z}^n/R(M_2)$.*

Определение 5.15. Назовём целочисленные матрицы *элементарно эквивалентными*, если одна получается из другой последовательностью элементарных строчных и элементарных столбцовых преобразований.

Упражнение 5.16. Проверьте, что отношение элементарной эквивалентности действительно является отношением эквивалентности (то есть оно рефлексивно, симметрично и транзитивно).

Из леммы 5.9 и следствия 5.14 выводим принципиальный для анализа факторгрупп $\mathbb{Z}^n$ (то есть всех конечно-порождённых групп) факт.

Теорема 5.17. *Пусть M_1 и M_2 элементарно эквивалентны. Тогда $\mathbb{Z}^n/R(M_1) \cong \mathbb{Z}^n/R(M_2)$.*

Эта теорема даёт удобный способ анализа конечно-порождённых абелевых групп. Дело в том, что для каждой целочисленной матрицы есть элементарно эквивалентная матрица очень простого вида.

Теорема 5.18 (нормальная форма Смита)**.** *Для любой целочисленной матрицы существует такая элементарно эквивалентная матрица, у которой все матричные элементы вне главной диагонали равны* 0*, а для диагональных элементов m_{ii} выполняется условие*: *m_{ii} делит $m_{(i+1)(i+1)}$.*

Доказательство. Если матрица состоит из одних нулей, то утверждение теоремы справедливо.

Пусть в матрице M есть ненулевые элементы. Будем применять к M элементарные преобразования до тех пор, пока можно уменьшить минимальную абсолютную величину ненулевого элемента матрицы. Затем перестановками строк и столбцов (и умножением строк на -1) добьёмся, чтобы m_{11} стал положительным элементом с минимальной абсолютной величиной среди элементов матрицы.

Докажем, что все ненулевые элементы матрицы делятся на m_{11}.

Пусть элемент первого столбца m_{i1} не делится на m_{11}. Разделим число m_{i1} на m_{11} с остатком: $m_{i1} = qm_{11}+r$ и $0 < r < m_{11}$. Если вычесть из i-й строки первую строку, умноженную на q, то $m'_{i1} = r$, то есть мы уменьшили минимальную абсолютную величину ненулевых элементов матрицы. Аналогично рассуждаем для элементов первой строки.

Пусть теперь m_{ij} не делится на m_{11}, то есть, как и выше, $m_{ij} = qm_{11}+r$, $0 < r < m_{11}$. Как уже доказано, $m_{i1} = am_{11}$, $m_{1j} = bm_{11}$. И в этом случае элементарными преобразованиями матрицы можно уменьшить минимальную абсолютную величину ненулевых элементов матрицы. Вычтем из i-й строки первую строку, умноженную на $(a-1)$, а затем из j-го столбца вычтем первый столбец, умноженный на $q-(a-1)b$. После этих преобразований (i,j)-й элемент матрицы будет равен r, как показывают вычисления (приводим только минор матрицы, образованный первой и i-й строкой и первым и j-м столбцом):

$$\begin{pmatrix} m_{11} & bm_{11} \\ am_{11} & qm_{11}+r \end{pmatrix} \to \begin{pmatrix} m_{11} & bm_{11} \\ m_{11} & (q-(a-1)b)m_{11}+r \end{pmatrix} \to$$
$$\to \begin{pmatrix} m_{11} & (b-q+(a-1)b)m_{11} \\ m_{11} & r \end{pmatrix}.$$

Поскольку все ненулевые элементы матрицы делятся на m_{11}, можно элементарными преобразованиями сделать равными нулю все элементы первого столбца и первой строки, за исключением m_{11} (вычитая подходящие кратные первой строки

или первого столбца). После этого матрица приобретает вид

$$M = \begin{pmatrix} m_{11} & 0 \\ 0 & M' \end{pmatrix},$$

где все ненулевые элементы матрицы M' делятся на m_{11}, а размер M' меньше, чем у исходной матрицы.

К матрице M' можно применить такие же преобразования, какие были применены к M и т.д. В конце концов получим требуемую форму матрицы. □

Группу, отвечающую матрице M в нормальной форме Смита, легко описать. Нулевые строки не влияют на подгруппу $R(M)$ (любое кратное нулевого вектора — нулевой вектор). Поэтому считаем в последующем рассуждении, что нулевых строк нет (но могут быть нулевые столбцы).

Лемма 5.19. *Пусть M — матрица в нормальной форме Смита, в которой есть k нулевых столбцов. Тогда*

$$\mathbb{Z}^n / R(M) \cong G = \bigoplus_{i=1}^{n-k} C_{m_{ii}} \oplus \mathbb{Z}^k.$$

Доказательство. Рассмотрим сюръективный гомоморфизм $\varphi\colon \mathbb{Z}^n \to G$, задаваемый правилом:

$$\varphi\colon (x_1, x_2, \ldots, x_n) \mapsto (x_1 \bmod m_{11}, \ldots, x_{n-k} \bmod m_{(n-k)(n-k)}, x_{n-k+1}, \ldots, x_n).$$

Ядро этого гомоморфизма совпадает с $R(M)$: если $x \in \operatorname{Ker}\varphi$, то для координаты $i \leqslant n-k$ должно выполняться условие $x_i \equiv 0 \pmod{m_{ii}}$, а для координаты $i > n-k$ должно выполняться условие $x_i = 0$. Это в точности целочисленные линейный комбинации строк M.

По теореме о гомоморфизмах получаем искомый изоморфизм $\mathbb{Z}^n / R(M) \cong G$. □

Из леммы 5.19 следует неизоморфизм групп $\mathbb{Z}^n$ и $\mathbb{Z}^k$ при $n \neq k$.

Следствие 5.20. *$\mathbb{Z}^n \ncong \mathbb{Z}^k$ при $n \neq k$.*

Доказательство. Пусть $n > k$ и группа $G = \mathbb{Z}^n$ изоморфна $\mathbb{Z}^k$. Обозначим через M_S матрицу в нормальной форме Смита, элементарно эквивалентную матрице $M(G)$. Поскольку в этой матрице строк меньше, чем столбцов, то $\mathbb{Z}^n / R(M_S)$ бесконечная (есть прямое слагаемое $\mathbb{Z}^k$, $k > 0$).

Но факторгруппа $\mathbb{Z}^n / R(M(G))$ единичная по предположению об изоморфизме $\mathbb{Z}^n$ и $\mathbb{Z}^k$. Полученное противоречие доказывает, что $\mathbb{Z}^n \ncong \mathbb{Z}^k$. □

Из теорем 5.17, 5.18 и леммы 5.19 получаем как следствие основной структурный результат о конечно-порождённых абелевых группах.

Следствие 5.21. *Любая конечно-порождённая абелева группа изоморфна прямой сумме циклических групп (конечных и бесконечных).*

5.3 Конечные абелевы группы

Как уже отмечено, любая конечная группа конечно порождена. Поэтому к конечным абелевым группам применимы теоремы из предыдущего раздела и получаем такое важное следствие: любая конечная абелева группа является прямой суммой циклических. Эта структурная теорема о конечных абелевых группах позволяет легко доказывать много утверждений о конечных абелевых группах, сводя их к изучению циклических прямых слагаемых. Приведём пример.

Как мы уже проверяли, обращение теоремы Лагранжа для порядков элементов неверно: если d — делитель порядка группы, то в группе необязательно есть элемент порядка d. Но оказывается, что если d ещё и простое число, то элемент порядка d обязательно найдётся.

Для общего случая это несложное следствие теоремы Силова (о которой в этом вводном курсе речи идти не будет). Для абелевых групп это утверждение легко следует из анализа структуры абелевой группы.

Лемма 5.22. *Пусть p — простое число, которое делит порядок n абелевой группы G. Тогда в G есть элемент порядка p.*

Доказательство. Группа G изоморфна прямой сумме циклических групп. Поэтому её порядок равен произведению порядков этих циклических групп. По основной теореме арифметики p делит порядок одной из циклических групп. Но для циклических групп обращение теоремы Лагранжа справедливо: если порядок $\langle g\rangle$ равен n, а $k \mid n$, то порядок $\langle g^{n/k}\rangle$ равен в точности k. □

Используя эту лемму, легко закончить анализ гомоморфизма $x \mapsto x^a$, начатый в разделе 3.1, как обещано в замечании 3.19.

Лемма 5.23. *Гомоморфизм возведения в степень a является автоморфизмом абелевой группы G порядка n тогда и только тогда, когда* $\text{НОД}(a, n) = 1$.

Доказательство. В одну сторону это утверждение 3.18 (достаточность). Докажем необходимость.

В утверждении 3.18 мы использовали аддитивную форму записи групповой операции. Сейчас для разнообразия используем мультипликативную.

Итак, пусть $\text{НОД}(a, n) = d > 1$. Возьмём какой-нибудь простой делитель p числа d. В группе G есть элемент g порядка p (это лемма 5.22). Но тогда $g^a = g^{pb} = e$, то есть g лежит в ядре гомоморфизма возведения в степень a. Поэтому такой гомоморфизм не является автоморфизмом. □

Продолжим анализ конечных абелевых групп и связанных с ними матриц.

Если матрица в нормальной форме Смита отвечает конечной группе G, её без ограничения общности можно считать квадратной (нулевых столбцов нет, а выбрасывание нулевых строк не изменяет факторгруппу $G \cong \mathbb{Z}^n/R(M)$).

Для диагональной матрицы M порядка n очевидно соотношение: $|\det M| = |\mathbb{Z}^n/R(M)|$. Как известно из линейной алгебры, элементарные преобразования

строк и столбцов не изменяют модуля детерминанта матрицы. Поэтому порядок группы, заданной матрицей M, легко найти, вычислив детерминант M, даже если эта матрица и не в нормальной форме.

Пример 5.24. Каков порядок абелевой группы G с двумя порождающими a, b порядка 15, для которых выполняются соотношения $a^3b^5 = e$ и $a^5b^3 = e$?

Запишем матрицу, строки которой порождают ядро гомоморфизма из $\mathbb{Z}^2$ в G:

$$\begin{pmatrix} 3 & 5 \\ 5 & 3 \end{pmatrix}.$$

Детерминант этой матрицы равен $3^2 - 5^2 = -16$. Поэтому порядок G равен 16.

Является ли G циклической группой? Для ответа на этот вопрос знания детерминанта недостаточно.

Преобразуем матрицу к нормальной форме Смита:

$$\begin{pmatrix} 3 & 5 \\ 5 & 3 \end{pmatrix} \xrightarrow{(1)} \begin{pmatrix} 3 & 2 \\ 5 & -2 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 1 & 2 \\ 7 & -2 \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} 1 & 0 \\ 7 & -16 \end{pmatrix} \xrightarrow{(4)} \begin{pmatrix} 1 & 0 \\ 0 & -16 \end{pmatrix}$$

Выполнены элементарные преобразования: (1) вычитание из второго столбца первого; (2) вычитание из первого столбца второго; (3) вычитание из второго столбца удвоенного первого; (4) вычитание из второй строки семикратной первой строки.

Из вида этой матрицы заключаем, что $G \cong C_1 \times C_{16} \cong C_{16}$, то есть группа и впрямь циклическая. □

Контрольный вопрос 5.25. Почему преобразования (1) и (2) не обратны друг другу?

Та же матричная техника может использоваться и для анализа факторгрупп циклических групп.

Пример 5.26 (продолжение примера 4.38)**.** В том примере мы сравнивали две факторгруппы

$$G_1 = (Z_8 \oplus Z_{16})/\langle(2,0)\rangle \quad \text{и} \quad G_2 = (Z_8 \oplus Z_{16})/\langle(4,4)\rangle.$$

Найдём их разложения в прямую сумму циклических, используя приведение матрицы к нормальной форме Смита.

Обе группы имеют два порождающих (смежные классы, содержащие $(1,0)$ и $(0,1)$ соответственно), обозначим их a, b. Для порождающих G_1 выполняются соотношения $a^8 = e$, $b^{16} = e$, $a^2 = 2$ (так как это факторгруппа по $\langle(2,0)\rangle$). Первое соотношение является следствием третьего. В матричном формализме это означает, что в матрице

$$\begin{pmatrix} 8 & 0 \\ 0 & 16 \\ 2 & 0 \end{pmatrix}$$

первая строка становится нулевой после вычитания из неё учетверённой третьей. Нулевую строку можно отбросить, не меняя факторгруппы, задаваемой матрицей. После перестановки столбцов и строк получаем матрицу

$$\begin{pmatrix} 2 & 0 \\ 0 & 16 \end{pmatrix},$$

которая соответствует факторгруппа $Z_2 \times Z_{16}$.

Пока мы фактически повторили рассуждение из примера 4.38. Но теперь тем же методом получим структуру группы G_2. Преобразуем соответствующую ей матрицу к нормальной форме:

$$\begin{pmatrix} 8 & 0 \\ 0 & 16 \\ 4 & 4 \end{pmatrix} \xrightarrow{(1)} \begin{pmatrix} 0 & -8 \\ 0 & 16 \\ 4 & 0 \end{pmatrix} \xrightarrow{(2)} \begin{pmatrix} 0 & -8 \\ 0 & 0 \\ 4 & 0 \end{pmatrix} \xrightarrow{(3)} \begin{pmatrix} 4 & 0 \\ 0 & 8 \\ 0 & 0 \end{pmatrix},$$

здесь выполнены элементарные преобразования: (1) вычитание из второго столбца первого; (2) сложение второй строки с удвоенной первой; (3) перестановки строк и столбцов и умножения на -1.

Отсюда видим, что $G_2 \cong Z_4 \oplus Z_8 \ncong Z_2 \oplus Z_{16} \cong G_1$. □

Любая конечная группа изоморфна прямой сумме конечных циклических групп. Некоторые прямые суммы оказываются изоморфными, как мы уже видели раньше на примерах. В частности, китайская теорема 3.49 утверждает, что $C_{pq} \cong C_p \oplus C_q$ для взаимно простых чисел p, q. Этот изоморфизм напоминает разложение числа на множители. В случае прямых сумм циклических групп также есть теорема однозначности, но она формулируется сложнее.

Продолжая разложение по китайской теореме, получим изоморфизм данной циклической группы с прямой суммой циклических групп, порядки которых — степени простых. Такие группы будем называть *примарными*.

Если применить китайскую теорему к прямой сумме циклических групп, получим разложение в прямую сумму примарных компонент. То есть, для любой конечной абелевой группы G существует изоморфизм

$$G \cong \bigoplus_{p,k} C_{p^s}^{k_{p,s}}, \tag{5.3}$$

в котором p — простые, а $k_{p,s}$ — *кратность*, с которой примарная группа C_{p^s} входит в прямую сумму. (Если для каких-то p, s в сумме нет слагаемого C_{p^s}, полагаем кратность равной нулю.)

Пример 5.27. Для групп G_1 и G_2 из примера 5.26 получаем такие разложения в примарные компоненты:

$$G_1 \cong C_2 \oplus C_{16}, \quad G_2 \cong C_4 \oplus C_8$$

и эти группы неизоморфны. В этих разложениях все кратности равны 1 (или 0 для отсутствующих слагаемых). Легко написать разложения с кратностью больше 1:

$$G_3 \cong C_2^3 \oplus C_4, \quad G_4 \cong C_2 \oplus C_4^2.$$

Все эти группы попарно неизоморфны. Но если G_1 и G_2 отличаются возможными порядками элементов (в G_1 есть элемент порядка 16, в G_2 нет элементов порядка 16, но есть элемент порядка 8), возможные порядки элементов в группах G_3 и G_4 одни и те же: 1, 2, 4. Поэтому разницу между ними нужно формулировать иначе.

Рассмотрим подгруппы квадратов в этих двух группах. В G_3 порядок группы квадратов равен 2: в первых трёх компонентах все элементы имеют порядок 2, так

что их квадраты совпадают с нейтральным элементом; а в последней компоненте порядок группы квадратов равен 2. Аналогичный анализ для группы G_4 показывает, что порядок группы квадратов равен $2^2 = 4$. Поэтому эти группы неизоморфны. □

Аналогичные этому примеру рассуждения показывают, что набор кратностей примарных компонент определён однозначно для данной группы, то есть разложение (5.3) единственно с точностью до изоморфизма.

Теорема 5.28. *Любая конечная абелева группа изоморфна прямой сумме циклических групп порядков p^k, p — простое, причём количество слагаемых порядка p^k одинаково для любого разложения в прямую сумму.*

В доказательстве теоремы мы восстановим часть рассуждений, пропущенных в разборе предыдущего примера.

Доказательство. Первая часть теоремы уже доказана выше. Осталось проверить единственность. Рассмотрим разложение (5.3) в прямую сумму примарных компонент. Нужно показать единственность кратностей $k_{p,s}$.

Выберем простое число p и докажем, что числа $k_s = k_{p,s}$ не зависят от выбора разложения (5.3).

Возведение в степень $\varphi_n\colon x \mapsto nx$ является гомоморфизмом абелевой группы, так как $n(x+y) = nx+ny$. Обозначим через $G^{(n)}$ образ G при гомоморфизме возведения в n-ю степень. Порядок группы $|G| = p^{a_0}q$, где $p \nmid q$. Из основной теоремы арифметики следует, что число a_0 однозначно определяется порядком группы $|G|$.

Обозначим $a_t = |G^{(p^t q)}|$, $1 \leqslant t < a_0$. В определении чисел a_t не используется разложения вида (5.3). Значит, если мы выразим k_s через a_t, то докажем, что числа k_s также не зависят от выбора разложения.

Докажем, что

$$G^{(p^t q)} \cong \bigoplus_{s>t} C_{p^s}^{k_s}. \tag{5.4}$$

Рассмотрим примарное слагаемое, порядок которого не делится на p. Тогда он делит q и поэтому возведение в степень $p^t q$ отображает такую циклическую группу в единичную.

Аналогичное рассуждение справедливо и для слагаемых C_{p^s} при $s \leqslant t$ (в этом случае порядок такого слагаемого делит p^t).

Если же $s > t$, то $C_{p^s}^{(p^t q)} \cong C_{p^s}^{(p^t)}$ (так как возведение в степень q, взаимно простую с порядком группы, является автоморфизмом циклической группы). Ясно также, что $C_{p^s}^{(p^t)} \cong C_{p^{s-t}}$ (кратные p^t имеют вид $p^t u a$, $0 \leqslant u < p^{k-t} - 1$, где a — порождающий C_{p^s}).

Порядок группы равен произведению порядков прямых слагаемых. Поэтому по-

лучаем систему уравнений (приравниваем показатели в степенях p)

$$\begin{aligned} k_1 + 2k_2 + 3k_3 + \cdots + ik_i + (i+1)k_{i+1} + \cdots + jk_j &= a_0, \\ k_2 + 2k_3 + \cdots + (i-1)k_i + ik_{i+1} + \cdots + (j-1)k_j &= a_1, \\ &\cdots \\ k_i + 2k_{i+1} + \cdots + (j-i+1)k_j &= a_i, \\ &\cdots \\ k_j &= a_{j-1}, \end{aligned}$$

в которой для простоты обозначений полагаем $j = a_0$. Из этой системы k_j однозначно выражаются через a_t: из последнего уравнения находим k_j, после чего из предпоследнего находим k_{j-1} и так далее. □

Пример 5.29. Приведём простой пример использования этой теоремы. Пусть порядки абелевых групп G_1 и G_2 не взаимно просты, то есть имеют общий простой делитель p. Докажем, что тогда группа $G_1 \oplus G_2$ не является циклической.

Обозначим через N наибольшую степень p, на которую делится произведение порядков G_1 и G_2.

Если группа $G_1 \oplus G_2$ циклическая, то в её разложении в прямую сумму примарных прямых слагаемых будет C_{p^N} (так как в циклической группе есть элементы всех порядков, которые делят порядок группы, в том числе и порядка p^N).

Однако если разложить по отдельности G_1 и G_2, такой примарной компоненты не будет.

Действительно, N равно сумме наибольшей степени p, на которую делится порядок G_1 и аналогичного числа для G_2. Каждое из чисел положительно по условию. Значит, каждое меньше N.

Из теоремы 5.28 заключаем, что группа $G_1 \oplus G_2$ не циклическая. □

5.4 Порождающие и соотношения в общем случае

Повторим ещё раз общую структуру анализа конечно-порождённой абелевой группы. Мы доказали, что любая такая группа является факторгруппой группы $\mathbb{Z}^n$ (свободная абелева группа ранга n). Любая подгруппа $\mathbb{Z}^n$ изоморфна $\mathbb{Z}^m$, где $m \leqslant n$. Эту подгруппу можно задать матрицей, причём элементарные преобразования матриц оставляют факторгруппу неизменной (с точностью до изоморфизма). В каждом классе эквивалентности отношения элементарной эквивалентности матриц есть нормальная форма Смита, для которой факторгруппа легко находится — это прямая сумма циклических групп (конечных или бесконечной).

В общем случае не обязательно абелевой группы ситуация отчасти похожая, но более сложная. Любая конечно-порождённая группа с n порождающими является факторгруппой некоторой группы F^n — свободной группы ранга n. Любая подгруппа свободной группы свободна (теорема Нильсена – Шрайера). Однако теперь её ранг необязательно меньше n и вообще может оказаться бесконечным. Кроме того,

для неабелевых групп нет столь же простого способа описания подгрупп, который гарантирует приведение подгруппы к нормальной форме. Это делает задачу классификации конечно-порождённых групп безнадёжно трудной (не существует алгоритма, который проверял бы изоморфизм конечно-порождённых групп).

Мы не будем доказывать теорему Нильсена – Шрайера и теорему об алгоритмической неразрешимости проверки изоморфизма конечно-порождённых групп. Ограничимся определением свободных групп и примерами задания групп порождающими и соотношениями.

5.4.1 Свободные группы

Напомним, что подгруппа $\langle S\rangle$, $S \subseteq G$, состоит из всех возможных произведений элементов S и обратных к ним. Неформально, свободная группа F порождается некоторым множеством своих элементов S так, что все произведения элементов S и обратных к ним различны.

Буквально это неверно, разумеется: для любого элемента группы выполняются равенства $gg^{-1} = g^{-1}g = e$. Поэтому более точное описание свободной группы получается добавлением уточнения: все произведения различны, за исключением тех, которые одинаковы из-за групповых аксиом. Это, конечно, не определение в математическом смысле: не очень понятно, как его использовать.

Чтобы дать точное определение, нужно прежде всего уточнить, о каких математических объектах мы говорим, когда говорим о произведениях элементов S и их обратных. Поскольку групповая операция ассоциативна, такое произведение однозначно задаётся последовательностью элементов $\Gamma = S \cup S^{-1}$, где через S^{-1} обозначено множество элементов, обратных к элементам S.

Такие последовательности принято называть *словами в алфавите* Γ. Конкатенации слов (приписыванию одного слова в конец другого) отвечает умножение соответствующих элементов группы. Поэтому мы снабжаем множество слов операцией конкатенации. Эта операция ассоциативна (проверьте!) и относительно неё есть нейтральный элемент — *пустое слово* (последовательность длины 0). Обратного нет ни для одного непустого слова.

Множество с ассоциативной операцией, относительно которой есть нейтральный элемент, называется *моноидом*. В частности, слова в алфавите Γ с операцией конкатенации образуют *свободный моноид с множеством порождающих* Γ (не будем объяснять, что это такое: заинтересованному читателю предлагается самостоятельно изобрести нужные определения и доказать, что этот моноид и впрямь свободен).

Поскольку в группе есть обратный для каждого элемента (в том числе и для порождающих), некоторым словам в алфавите Γ отвечают одни и те же элементы. Более точно, выполняются *тривиальные соотношения* на словах:

$$ugg^{-1}v = uv,$$
$$ug^{-1}gv = uv$$

для любых слов u, v и любого $g \in S$.

Используем эти соотношения, чтобы задать отношение эквивалентности на словах в алфавите Г.

Определение 5.30. Два слова u, v в алфавите $\Gamma = S \cup S^{-1}$ называются эквивалентными, если одно получается из другого цепочкой преобразований вида

$$w'gg^{-1}w'' \to w'w'', \quad w'w'' \to w'gg^{-1}w'', \quad w'g^{-1}gw'' \to w'w'', \quad w'w'' \to w'g^{-1}gw''. \tag{5.5}$$

Будем обозначать это отношение эквивалентности $\sim$.

Разумеется, хотя мы и назвали слова эквивалентными, это не означает, что для заданного таким образом отношения выполняются свойства отношения эквивалентности. Это нужно доказывать.

Утверждение 5.31. *Определение* 5.30 *задаёт отношение эквивалентности.*

Доказательство. Рефлексивность очевидна: слово u получается из слова u после пустого множества преобразований вида (5.5).

Симметричность легко следует из вида преобразований: они симметричны. Если слово u получается из слова v преобразованием вида (5.5), то и слово v получается из слова u преобразованием вида (5.5).

Осталось проверить транзитивность. Это также легко: если v получается из u, а w из v цепочками преобразований вида (5.5), то w получится из u последовательным применением этих двух цепочек. □

Определение 5.32. *Свободная группа* F_S с множеством порождающих S состоит из классов эквивалентности слов в алфавите $\Gamma = S \cup S^{-1}$. Групповая операция — конкатенация слов.

Если множество S состоит из n элементов, то говорим о свободной группе ранга n и обозначаем её F_n.

Мы уже много раз встречались с такими определениями (факторобъектов). Для них обязательно нужно доказывать корректность: ведь буквально определение операции зависит от выбора представителей в классе эквивалентности.

Утверждение 5.33. *Определение* 5.32 *задаёт группу.*

Доказательство. Начнём с проверки корректности операции. Пусть $u \sim u'$, $v \sim v'$. Нужно доказать, что тогда $uv \sim vv'$. Это нетрудно: сделаем вначале преобразования, который переводят слово u в u', получим из uv слово $u'v$. Далее выполним те преобразования, которые переводят слово v в слово v' и получим слово $u'v'$.

Теперь уже ясно, что ассоциативность групповой операции следует из ассоциативности конкатенации:

$$[u] \cdot ([v] \cdot [w]) = [u] \cdot [vw] = [uvw] = [uv] \cdot [w] = ([u] \cdot [v]) \cdot [w].$$

Здесь через $[u]$ обозначается класс эквивалентности, содержащий слово u.

Нейтральным элементом будет класс эквивалентности пустого слова.

Осталось доказать, что для любого слова есть обратное. Это доказательство повторяет рассуждение о тождестве $(xy)^{-1} = y^{-1}x^{-1}$ в общем случае.

Рассмотрим некоторое слово $u = g_1 g_2 \dots g_r$ в алфавите $\Gamma = S \cup S^{-1}$. Построим по нему слово $\bar{u} = \bar{g}_r \bar{g}_{r-1} \dots \bar{g}_1$. Здесь на символах алфавита используется соответствие: если $g \in S$, то $\bar{g} = g^{-1}$ (это также символ алфавита, он принадлежит множеству S^{-1}); если $g = h^{-1} \in S^{-1}$, то $\bar{g} = h$.

Проверим, что

$$[u\bar{u}] = [\bar{u}u] = [],$$

то есть, что $[\bar{u}]$ обратный к $[u]$. Слово

$$u\bar{u} = g_1 g_2 \dots g_r \bar{g}_r \bar{g}_{r-1} \dots \bar{g}_1$$

преобразуется в пустое последовательным удалением самой внутренней пары символов $g_r \bar{g}_r$ (по построению, такое преобразование имеет вид (5.5)):

$$g_1 g_2 \dots g_r \bar{g}_r \bar{g}_{r-1} \dots \bar{g}_1 \to g_1 g_2 \dots g_{r-1} \bar{g}_{r-1} \dots \bar{g}_1 \to \dots \to g_1 \bar{g}_1 \to \varepsilon$$

(через ε обозначили пустое слово).

Точно так же проверяется второе равенство $[\bar{u}u] = []$. □

Как уже сказано, имеет место теорема.

Теорема 5.34 (теорема Нильсена–Шрайера)**.** *Подгруппа свободной группы свободна.*

Мы не будем доказывать эту теорему. Повторим лишь, что подгруппа свободной группы ранга n может иметь бóльший ранг (и даже бесконечный ранг).

Теорема 5.35. *Всякая конечно порождённая группа G с n порождающими является фактором свободной группы F_n.*

Доказательство. Сюръективный гомоморфизм из F_n на G задаётся очевидным образом: слову из порождающих и обратных сопоставим произведение этих порождающих и обратных в группе G.

Свойство гомоморфизма следует из ассоциативности умножения в G. Сюръективность — из того, что любой элемент группы является произведением порождающих и обратных. □

Далее для простоты обозначений мы всюду подразумеваем этот гомоморфизм и говорим о словах в алфавите $\Gamma = S \cup S^{-1}$ как об элементах G.

5.4.2 Задание группы порождающими и соотношениями

Пользуясь свойствами групповой операции, любое равенство $u = v$ можно преобразовать к виду $uv^{-1} = e$. Пусть имеется группа G с множеством порождающих S. Если для некоторого слова w в алфавите $\Gamma = S \cup S^{-1}$ выполняется равенство $w = e$, мы называем слово w *соотношением* между порождающими (напомним, что слова в этом алфавите канонически отображаются в элементы группы G, как указано в доказательстве теоремы 5.35).

Определение 5.36. Через $\langle\langle S\rangle\rangle$, где S — подмножество группы G обозначаем наименьшую нормальную подгруппу группы G, содержащую множество S.

Пример 5.37. В неабелевом случае $\langle\langle S\rangle\rangle$ не всегда равно $\langle S\rangle$. Как мы помним, группа A_5 не содержит нетривиальных нормальных подгрупп. Поэтому $\langle\langle(1\ 2\ 3)\rangle\rangle = A_5$, а $\langle(1\ 2\ 3)\rangle$ — циклическая подгруппа порядка 3.

Причина такой разницы понятна: из свойства нормальности следует, что $\langle\langle S\rangle\rangle$ вместе с каждым элементом $g \in S$ содержит и все его сопряжённые hgh^{-1}, $h \in G$. □

Определение 5.38. Пусть R — множество соотношений между порождающими из множества S. Тогда группа (S, R) — это факторгруппа $F_S/\langle\langle R\rangle\rangle$ (мы рассматриваем соотношения как элементы свободной группы).

Группу (R, S) можно задать и более конструктивно: как множество классов эквивалентности на множестве слов в алфавите $S \cup S^{-1}$. Эта эквивалентность задаётся преобразованиями, похожими на (5.5):

$$w'uw'' \to w'w'', \quad w'w'' \to w'uw'', \quad u \in R \cup \{gg^{-1}, g^{-1}g : g \in S\}. \tag{5.6}$$

Упражнение 5.39. Проверьте, что цепочки преобразований (5.6) задают отношение эквивалентности на словах в алфавите $S \cup S^{-1}$ и что классы эквивалентности этого отношения с операцией конкатенации слов образуют группу.

Будем обозначать это отношение эквивалентности через $=_R$. Ясно, что классы эквивалентности $=_R$ являются объединениями классов эквивалентности отношения, задающего свободную группу (преобразования (5.5) входят в преобразования (5.6)). Поэтому дальше мы говорим о классах эквивалентности отношения $=_R$ на элементах свободной группы F_S.

Утверждение 5.40. *Группа классов эквивалентности отношения, задаваемого цепочками преобразований* (5.6)*, изоморфна* (R, S)*.*

Доказательство. Для этого достаточно проверить, что класс эквивалентности пустого слова задаёт наименьшую нормальную подгруппу в F_S, содержащую соотношения из R.

То, что класс эквивалентности пустого слова является нормальной подгруппой, содержащей R, легко проверить: он замкнут относительно сопряжения. Если $[w] =_R []$, то и $[uwu^{-1}] =_R []$ (сначала преобразуем w к пустому слову, а потом сократим слово uu^{-1} как в доказательстве утверждения 5.33).

С другой стороны, если $R \subseteq H \lhd F_S$, то H обязана содержать все слова из класса $[]$. Единичный элемент оно содержит. Далее, $w'w'' \in H$ равносильно $w'gg^{-1}w'' \in H$, аналогично $w'w'' \in H$ равносильно $w'gg^{-1}w'' \in H$.

Рассмотрим слово $w'uw''$, $u \in R$. Пусть оно принадлежит H. Докажем, что тогда и слово $w'w'' \in H$: слово $(w')^{-1}w'uw''(w') = uw''w'$ принадлежит H (так как подгруппа H нормальная, то есть замкнута относительно сопряжения), значит, и слово $u^{-1}(uw''w') = w''w' \in H$, так как $u^{-1} \in H$. Но тогда и $w'w'' = (w'')^{-1}(w''w')w'' \in H$, так как H нормальная.

Аналогично и в другую сторону. Пусть $w'w'' \in H$. Тогда $(w')^{-1}w'w''w' = w''w' \in H$, откуда $uw''w' \in H$. Но тогда $w'(uw''w')(w')^{-1} = w'uw'' \in H$.

Мы проверили, что применение преобразований 5.6 к элементам подгруппы H не выводит за пределы H. Поэтому класс эквивалентности пустого слова (единичный элемент, который всегда входит в подгруппу) целиком содержится в H. □

5.4.3 Примеры

Приведём несколько примеров задания групп порождающими и соотношениями. Анализ в этих примерах следует одной общей схеме.

Пусть мы хотим задать группу G порождающими и соотношениями. Для этого нужно установить изоморфизм $G \cong (S, R)$ для некоторого множества порождающих и соотношений между ними.

Проще задавать гомоморфизм из свободной группы F_S в группу G.

Утверждение 5.41. *Для любой группы G и свободной группы F_S любое отображение $\varphi\colon S \to G$ однозначно продолжается до гомоморфизма $\varphi\colon F_S \to G$.*

Доказательство. Обратный элемент g^{-1}, $g \in S$, обязан переходить в $(\varphi(g))^{-1}$ при гомоморфизме. Так как образ произведения равен произведению образов, то любое слово $w = w_1 \dots w_n$ в алфавите $S \cup S^{-1}$ обязано переходить в элемент группы G вида

$$\varphi(w_1)\varphi(w_2)\dots\varphi(w_n).$$

Ясно также, что этот единственно возможный вариант и впрямь является гомоморфизмом, так как

$$\varphi(w_1 \dots w_n u_1 \dots u_k) = \varphi(w_1)\varphi(w_2)\dots\varphi(w_n)\varphi(u_1)\varphi(u_2)\dots\varphi(u_k) = \\ = \varphi(w_1 \dots w_n)\varphi(u_1 \dots u_k)$$

в силу ассоциативности групповой операции в G. □

Если ядро гомоморфизма $\varphi\colon F_S \to G$ содержит $\langle\langle R \rangle\rangle$, то такой гомоморфизм поднимается до гомоморфизма $\tilde{\varphi}\colon (S, R) = F_S/\langle\langle R \rangle\rangle \to G$, как утверждается в следующей лемме.

Лемма 5.42. *Пусть $\varphi\colon G \to H$ — гомоморфизм групп и $K \lhd G$ — нормальная подгруппа G, которая лежит в ядре гомоморфизма, $K \subseteq \operatorname{Ker}\varphi$.*

Тогда существует гомоморфизм $\tilde{\varphi}\colon G/K \to G$, согласованный с φ в том смысле, что $\tilde{\varphi}([x]_K) = \varphi(x)$.

Доказательство. Из условия согласованности уже следует, чему должны равняться образы для всех смежных классов по K. Нужно проверить корректность этого определения. Сохранение операции будет сразу же следовать из свойств гомоморфизмов и определения факторгруппы:

$$\tilde{\varphi}([x]_K[y]_K) = \tilde{\varphi}([xy]_K) = \varphi(xy) = \varphi(x)\varphi(y).$$

Проверим корректность. Пусть $x = yk$, $k \in K \subseteq \operatorname{Ker}\varphi$, — два элемента из одного класса смежности по подгруппе K. Тогда $\varphi(x) = \varphi(y)\varphi(k) = \varphi(y)e = \varphi(y)$. □

Если гомоморфизм $\varphi\colon F_S \to G$ сюръективный, то и индуцированный гомоморфизм $\tilde{\varphi}(S, R) \to G$ будет сюръективным. Чтобы он был биективным, нужно проверить инъективность. Это уже комбинаторное свойство: нужно проверить, что два слова u, v в алфавите $S \cup S^{-1}$, которые φ переводит в один элемент $g \in G$, получаются друг из друга последовательностью преобразований (5.6) (то есть, задают один и тот же элемент группы (S, R)).

Пример 5.43 (циклическая группа)**.** Один порождающий элемент a и одно соотношение a^n. Гомоморфизм $\varphi\colon F_1 \to C_n$ задаётся условием $a \mapsto g$, где $a \in F_1$ — порождающая F_1, а g — порождающая C_n. Ядро гомоморфизма состоит из степеней a^{nk}, $k \in \mathbb{Z}$. Ясно, что $\langle\langle a^n \rangle\rangle$ принадлежит ядру. По лемме 5.42 получаем сюръективный гомоморфизм

$$\tilde{\varphi}\colon (\{a\}, a^n) \to C_n.$$

Проверим его инъективность. Пусть $\varphi(a^s) = \varphi(a^r)$, то есть $\varphi(a^{s-r}) = e$. Это значит, что $s - r$ кратно n. Без ограничения общности $s > r$. Слово a^{s-r} сокращением слов a^n приводится к пустому. Это означает инъективность $\tilde{\varphi}$.

Получаем изоморфизм $(\{a\}, a^n) \cong Z_n$. □

Пример 5.44 (свободная абелева группа ранга 2)**.** Два порождающих a, b, которые коммутируют. То есть одно соотношение $aba^{-1}b^{-1}$ (равенство $ab = ba$ равносильно равенству $aba^{-1}b^{-1} = e$). Временно обозначим эту группу G и докажем, что она изоморфна $\mathbb{Z}^2$.

Построим гомоморфизм $F_2 \to \mathbb{Z}^2$

$$\varphi\colon a \mapsto (1, 0), \quad \varphi\colon b \mapsto (0, 1).$$

Тогда слово $w \in \{a, b, a^{-1}, b^{-1}\}$ обязано переходить в пару (n_a, n_b), где n_a — алгебраическая сумма букв a в слове w, а n_b — алгебраическая сумма букв b (буквы a, b дают вклад $+1$, а буквы a^{-1}, b^{-1} дают вклад -1). Поэтому $\varphi(aba^{-1}b^{-1}) = (0, 0)$. Это означает, что ядро φ содержит подгруппу $\langle\langle aba^{-1}b^{-1} \rangle\rangle$. По лемме 5.42 гомоморфизм φ задаёт гомоморфизм $\tilde{\varphi}\colon G \to \mathbb{Z}^2$. Этот гомоморфизм очевидно сюръективный. Осталось доказать его инъективность.

Инъективность означает, что прообраз пары $(0, 0)$ единственный в группе G. Если посмотреть на элементы G как на классы эквивалентности свободной группы F_2, то видим, что требуется доказать такое утверждение: если у слова u алгебраические суммы букв равны 0, то это слово получается из пустого преобразованиями из (5.6), где $R = \{aba^{-1}b^{-1}\}$ (то есть задаёт единичный элемент в группе $(\{a, b\}, R)$).

Сокращение двух рядом стоящих взаимно обратных букв входит в допустимые преобразования. Докажем, что $ab =_R ba$. Так как

$$\varepsilon \to_R b^{-1}a^{-1}ab \to_R b^{-1}a^{-1}(aba^{-1}b^{-1})ab = a^{-1}b^{-1}ab,$$

то

$$ab \to_R (baa^{-1}b^{-1})ab = ba(a^{-1}b^{-1}ab) \to_R ba.$$

Контрольный вопрос 5.45. Докажите, что $a^{-1}b^{-1} =_R b^{-1}a^{-1}$.

Теперь ясно, что перестановками букв и сокращениями взаимно обратных букв можно добиться, чтобы буквы a (или обратные) шли в слове до букв b (или обратных). Поскольку в исходном слове алгебраическая сумм по каждой букве равна 0, в таком слове не будет букв вообще. То есть это пустое слово, что и требовалось доказать. □

Пример 5.46 (диэдральная группа)**.** Напомним, что группа D_n — это группа симметрий правильного n-угольника. Обозначим поворот против часовой стрелки на угол $2\pi/n$ вокруг центра n-угольника через r, а отражение относительно некоторой прямой, проходящей через вершину n-угольника, через p. Очевидны соотношения

$$r^n = e, \quad p^2 = e. \tag{5.7}$$

Кроме того, нетрудно проверить, что сопряжение поворота n-угольника на угол $2\pi/n$ отражением относительно прямой — это поворот на тот же угол, но в другом направлении (в других терминах мы это проверили в примере 4.51).

Поэтому есть ещё одно соотношение

$$prp = r^{-1}, \text{ что равносильно } (pr)^2 = e. \tag{5.8}$$

Три соотношения (5.7), (5.8) порождают группу диэдра. Действительно, из (5.8) следует, что

$$pr^k p = pr^{k-1}pprp = pr^{k-1}pr^{-1} = \ldots = r^{-k}\,,$$

поэтому любое произведение элементов p и r равно такому произведению, в котором p встречается не более одного раза. С учётом (5.7) таких выражений $1 + 3(n-1)$ штук: $1 = p^0 = r^0$, r^k, pr^k, $r^k p$, где $1 \leqslant k \leqslant n-1$. Но из (5.8) следует, что $rp = pr^{-1}$, поэтому

$$r^k p = r^{k-1}pr^{-1} = \ldots = pr^{-k} = pr^{n-k}.$$

Итак в группе, порождённой соотношениями (5.7), (5.8), не более $2n$ элементов.

Поскольку есть сюръективный гомоморфизм на D_n (аналогично предыдущим разобранным случаям), в ней ровно $2n$ элементов и она изоморфна D_n. □

6 Действия групп

Специальный случай гомоморфизмов — гомоморфизмы в группу биекций множества — оказывается очень полезным в самых разных приложениях, не только в самой теории групп, но и в комбинаторике, и других областях математики.

Определение 6.1. *Действием группы G на множестве X* называется гомоморфизм $\varphi\colon G \to S(X)$ группы G в группу $S(X)$ биекций множества X (взаимно однозначных отображений множества X на себя). Говорят также, что группа G действует на множестве X.

Если ясно, о каком действии идёт речь, то $\varphi(g)(x)$ записывают как $g(x)$. Элементы множества X будем называть точками, чтобы отличать их от элементов группы G. Элементы группы g будем называть биекциями на X, имея в виду образ элемента при гомоморфизме φ.

В этой главе мы рассмотрим некоторые простейшие свойства действий и их приложения к комбинаторике.

Другой важный пример гомоморфизмов — гомоморфизмы в группы невырожденных линейных преобразований (*представления групп*) — здесь не рассматривается. Представления групп важны не только в математике, но и в квантовой физике.

6.1 Комбинаторные и геометрические примеры

Группа перестановок S_n естественным образом действует на множестве $\{1, 2, \ldots, n\}$. Это действие задаётся тождественным гомоморфизмом из S_n в $S(\{1, 2, \ldots, n\})$ (разные обозначения для одного и того же).

Однако с группой перестановок связаны и более интересные действия.

Пример 6.2. Действие на подмножествах. Обозначим P_n множество всех подмножеств n-элементного множества $\{1, 2, \ldots, n\}$. Группа перестановок S_n действует на P_n по правилу:

$$\pi(X) = \{\pi(i) : i \in X\}.$$

Это биекция на P_n: единственным прообразом множества Y является множество $\{\pi^{-1}(y) : y \in Y\}$ (здесь существенно, что π является биекцией на $\{1, 2, \ldots, n\}$).

Проверка свойства гомоморфизма сводится к проверке равенства

$$(\sigma \circ \pi)(X) = \sigma(\pi(X)),$$

которое легко следует из определения. Нужно доказать равенство множеств

$$\{(\sigma \circ \pi)(i) : i \in X\} \quad \text{и} \quad \{(\sigma(j) : j \in \pi(X)\}.$$

Но второе множество совпадает с множеством $\{(\sigma(\pi(i)) : i \in X\}$, которое равно первому, поскольку по определению композиции $(\sigma \circ \pi)(i) = \sigma(\pi(i))$ для любого i. □

Композиция гомоморфизмов является гомоморфизмом. Поэтому любое действие S_n на m-элементном множестве продолжается до действия S_n на подмножествах этого множества. Приведём важный для теории графов пример.

Пример 6.3 (действие на помеченных графах)**.** Из действия S_n на подмножествах n-элементного множества можно выделить действие на парах элементов: ясно, что количество элементов в $\pi(X)$ и X одинаково (π — биекция).

Общее количество пар: $\binom{n}{2} = n(n-1)/2$. Действие S_n на парах переносится на действие на подмножествах пар.

При чём тут графы? Простой неориентированный граф — это как раз по определению подмножество пар вершин. Перестановка вершин переводит один простой неориентированный граф в другой. Так что мы фактически определили действие группы перестановок вершин на графах.

Разумеется, от переименования вершин разумные свойства графа (определяемые через смежность вершин) не изменяются. Графы, которые получаются один из другого переименованием вершин, называются *изоморфными*.

Заметим, что мы различаем вершины графа между собой. Это и называется в комбинаторике «помеченным случаем». «Непомеченные» графы определяются с помощью указанного действия перестановок вершин, мы сделаем это ниже. □

Диэдральные группы и группы правильных многогранников действуют на множестве вершин: каждая симметрия многоугольника или многогранника переводит вершины в вершины и композиции таких симметрий отвечает композиция соответствующих им перестановок вершин (свойство гомоморфизма).

Мы уже использовали эти действия при подсчёте порядков этих групп. В этом разделе мы используем действия для описания групп многогранников в алгебраических терминах. Для этого нам потребуются действия не только на множествах вершин.

Пример 6.4. Впрочем, для группы тетраэдра, обозначим её G, достаточно рассмотреть действие на вершинах. Оно задаёт гомоморфизм $G \to S_4$. Этот гомоморфизм инъективный: если все вершины остались на месте, то и сама симметрия тождественная.

Определение 6.5. Если гомоморфизм $G \to S(X)$ инъективный, то действие называется *точным*.

Теперь найдём образ этого гомоморфизма.

Из каких поворотов состоит группа тетраэдра? Помимо тождественного, есть повороты двух типов: на 120° вокруг прямой, проходящей через вершину тетраэдра и центр противоположной грани, а также на 180° вокруг прямой, проходящей через середины скрещивающихся рёбер, см. рис. 22.

Перестановки вершин, отвечающие этим поворотам это циклы длины 3 и пары циклов длины 2. Эти перестановки чётные. Значит, образ группы тетраэдра лежит в группе чётных перестановок A_4 и совпадает с ней (так как ядро единичное, а в группе тетраэдра столько же элементов, сколько в A_4: $12 = 4!/4$).

Поскольку гомоморфизм инъективный, по теореме о гомоморфизме группа тетраэдра изоморфна образу, то есть A_4. □

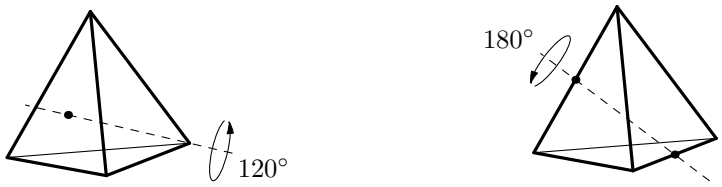


Рис. 22: Поворотные симметрии тетраэдра

Пример 6.6. Теперь рассмотрим группу куба. Помимо единичного элемента, она состоит из поворотов, указанных на рис. 23: повороты на 120° вокруг больших диагоналей (проходящих через противоположные вершины куба); повороты на 90° и на 180° вокруг прямых, проходящих через центры противоположных граней; и повороты на 180° вокруг прямых, проходящих через середины противоположных рёбер.

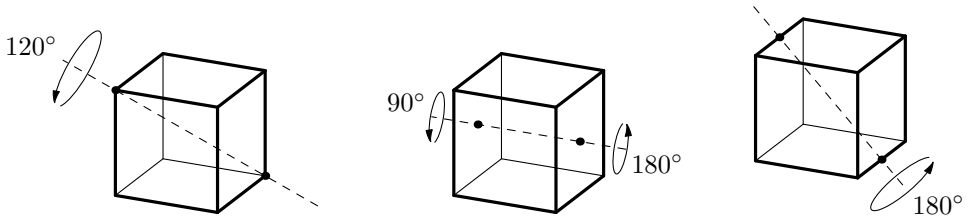


Рис. 23: Поворотные симметрии куба

Действие на вершинах куба по-прежнему инъективно. Получаем вложение группы куба в группу перестановок S_8. Однако понять структуру группы куба такое вложение не очень помогает.

Более удачным примером является действие группы куба на больших диагоналях. Их 4 и они обязаны переходить в себя, так как противоположные вершины куба находятся на максимальном расстоянии в кубе.

Получаем гомоморфизм из группы куба в группу S_4. В его образе содержатся все транспозиции. Действительно, поворот на 180°, показанный на рис. 23 справа, пару больших диагоналей переставляет, а вторую пару (те диагонали, которые перпендикулярны оси симметрии) оставляет на месте. Значит, образ группы куба совпадает со всей группой S_4.

Но количество элементов в группе куба $24 = 4!$. Сюръективный гомоморфизм из группы порядка 24 в группу порядка 24 неминуемо инъективен.

Приходим к выводу, что группа куба изоморфна S_4. □

Пример 6.7. Заметим, что группа куба действует не только на вершинах, но и на парах противоположных граней. (Разумеется, при симметрии куба пара противоположных граней перейдёт в какую-то пару противоположных граней — других пар граней без общих точек в кубе нет).

Получаем гомоморфизм $S_4 \to S_3$ (ведь группа куба изоморфна S_4). Этот гомоморфизм сюръективен, поскольку образ при таком гомоморфизме содержит все транспозиции в S_3. Чтобы это понять, нужно убедиться, что поворот на 90°, показанный на рис. 23 в центре, переставляет две пары противоположных граней (а ту пару граней, через центры которых проходит ось симметрии, оставляет на месте).

Это показывает, что в группе S_4 есть нормальная подгруппа порядка $|S_4|/|S_3| = 4$. Факторгруппа по этой подгруппе изоморфна S_3.

Нетрудно убедиться, что эта нормальная подгруппа состоит из тождественной перестановки и трёх перестановок, в цикловом разложении которых есть пара циклов длины 2. Такие перестановки образует класс сопряжённости в S_4, так что эта подгруппа нормальная.

Более простой способ увидеть этот гомоморфизм состоит в том, чтобы рассмотреть действие S_4 на 4 точках . После этого можно заметить, что S_4 действует также и на разбиениях 4 точек на непересекающиеся пары (таких разбиений ровно 3). Это даёт тот же самый сюръективный гомоморфизм $S_4 \to S_3$. □

Пример 6.8. Имея нормальную подгруппу $N \lhd S_4$ индекса 6 (ядро описанного выше гомоморфизма $\varphi\colon S_4 \to S_3$), легко построить подгруппу порядка 8 в S_4. Для этого возьмём прообраз $\varphi^{-1}((a\ b))$ (для наглядности договоримся, что S_3 переставляет не числа $1, 2, 3$, а буквы a, b, c, которые мы обозначили разбиения

$$\{1,3\} \sqcup \{2,4\}, \quad \{1,4\} \sqcup \{2,3\}, \quad \{1,2\} \sqcup \{3,4\}$$

соответственно). Это смежный класс по N, как и любой прообраз элемента при гомоморфизме групп: $\varphi^{-1}((a\ b)) = \pi N$. Тогда $N \cup \pi N$ и будет искомой подгруппой порядка 8. Почему?

Дело в том, что $\pi^2 \in N$ (так как $\varphi(\pi)^2 = ()$). Поэтому объединение этих смежных классов замкнуто относительно композиции перестановок (и конечно). Этого достаточно, чтобы оно было подгруппой.

Нетрудно указать явно порождающие подгруппы порядка 8. Для этого заметим, что $\varphi((1\ 2)) = (a\ b)$ в указанных выше обозначениях. Поэтому подгруппа

$$\langle (1\ 2), (1\ 2)(3\ 4), (1\ 3)(2\ 4), (1\ 4)(2\ 3) \rangle$$

имеет порядок 8 (читателю рекомендуется явно выписать недостающие 4 перестановки из этой подгруппы). □

Пример 6.9. Рассмотрим, наконец, группу додекаэдра. В ней 60 элементов: столько же, сколько в знакопеременной группе A_5. Оказывается, как мы уже упоминали выше, A_5 изоморфна группе додекаэдра.

Доказать это можно аналогичным геометрическим построением. Прежде всего нужно найти множество из 5 элементов, на которых действует группа додекаэдра. Это 5 кубов, вписанных в додекаэдр, см. рис. 19 в центре.

Действие группы додекаэдра на вписанных кубах инъективно. Для проверки инъективности нужно убедиться, что любая нетождественная поворотная симметрия додекаэдра не сохраняет хотя бы один куб.

Помимо тождественного преобразования, поворотные симметрии додекаэдра такие: 24 вращения вокруг каждой из шести осей, проходящих через противоположные вершины, на углы $2\pi/5$, $4\pi/5$, $6\pi/5$ и $8\pi/5$; 20 вращений вокруг каждой из десяти осей, проходящих через центры противоположных граней, на углы $2\pi/3$ и $4\pi/3$; 15 вращений вокруг каждой из пятнадцати осей, проходящих через середины противоположных рёбер, на угол π. Аккуратной проверкой убеждаемся, что для каждого нетождественного вращения найдётся куб, переводящийся данным вращением в другой куб. Поэтому только тождественному вращению соответствует тождественная подстановка систем рёбер.

Получаем инъективный гомоморфизм группы додекаэдра в группу S_5. Подгруппа порядка 60 в S_5 единственная, это утверждение задачи 4.49 (если вы её ещё не решили, самое время вернуться к ней и решить; обратите внимание на указание). □

6.2 Орбиты и стабилизаторы

Введём несколько общих понятий, относящихся к действиям групп. Зафиксируем некоторое действие и будем писать $g(x)$ для обозначения образа точки x при действии элемента g.

Определение 6.10. Пусть группа G действует на множестве X. Для любого $x \in X$ множество $\mathrm{Stab}_x(G)$ элементов группы G, оставляющих точку x неподвижной, называется *стабилизатором* точки x:

$$\mathrm{Stab}_x(G) = \{g \in G : g(x) = x\}.$$

Если группа G фиксирована, то в обозначении стабилизатора она опускается.

Пример 6.11. Для действия группы многогранника на вершинах стабилизатор вершины состоит из поворотов вокруг оси симметрии, проходящей через эту вершину (вершины, соседние с данной вершиной должны переходить в соседние).

Это подгруппа группы многогранника, порядок которой совпадает с количеством рёбер, выходящих из вершины. Мы использовали эти подгруппы при подсчёте порядков групп многогранников. □

Пример 6.12. Рассмотрим действие циклической группы C_n на множестве X из k точек. По определению действия это означает, что мы рассматриваем гомоморфизм $\varphi\colon C_n \to S_k$. Если образ $\varphi(a)$ порождающей a циклической группы имеет порядок n, то этот гомоморфизм инъективный.

Контрольный вопрос 6.13. Докажите предыдущее утверждение.

Возьмём для определённости $k = 15$, $n = 60$, а

$$\varphi(a) = (1)(2\ 3)(4\ 5\ 6)(7\ 8\ 9\ 10)(11\ 12\ 13\ 14\ 15).$$

Порядок этой перестановки $\mathrm{НОК}(1, 2, 3, 4, 5) = 60$.

Найдём Stab_1. Очевидно, что это все степени a, так как 1 — неподвижная точка перестановки $\varphi(a)$. Таким образом, $\mathrm{Stab}_1 = C_{60}$.

Найдём Stab_2. Если $\varphi(a)^k(2) = 2$, то k чётно. Поэтому

$$\mathrm{Stab}_2 = \langle a^2 \rangle \cong C_{30}.$$ □

Контрольный вопрос 6.14. Найдите Stab_{11} в предыдущем примере.

Во всех рассмотренных примерах стабилизатор оказывался подгруппой группы, действующей на множестве. Это не случайно: так будет всегда.

Утверждение 6.15. *Стабилизатор* $\mathrm{Stab}_x(G)$ *— подгруппа* G*.*

Доказательство. Проверим свойства подгруппы. Во-первых, $e \in \mathrm{Stab}_x(G)$, так как $e(x) = x$ (при гомоморфизме нейтральный элемент переходит в нейтральный, а нейтральный элемент в группе $S(X)$ — тождественное отображение). Во-вторых, так как $g^{-1}(g(x)) = (g^{-1}g)(x) = e(x) = x$, то из $g \in \mathrm{Stab}_x(G)$ следует $g^{-1} \in \mathrm{Stab}_x(G)$. В-третьих, если $g, h \in \mathrm{Stab}_x(G)$, то и $gh \in \mathrm{Stab}_x(G)$, поскольку $g(h(x)) = g(x) = x$. □

Второе важное понятие, связанное с действием группы: орбита.

Определение 6.16. Пусть группа G действует на множестве X. *Орбитой действия* называется множество образов некоторой точки x:

$$\mathrm{Orb}_x(G) = \{y \in X : y = g(x),\ g \in G\}.$$

Как и в случае стабилизатора, если группа G фиксирована, то в обозначении орбиты точки она опускается.

Пример 6.17 (продолжение примера 6.12)**.** В том примере мы рассматривали действие группы C_{60} на 15 точках, задаваемое гомоморфизмом

$$\varphi\colon a^k = \big((1)(2\ 3)(4\ 5\ 6)(7\ 8\ 9\ 10)(11\ 12\ 13\ 14\ 15)\big)^k,$$

где a — порождающий циклической группы.

Орбита точки 11 состоит из тех $i \in \{1, \ldots, 15\}$, которые могут быть образом 11 при действии какой-нибудь степенью перестановки $\varphi(a)$. Легко видеть, что это в точности числа из цикла (11 12 13 14 15). Любое такое число можно получить из 11, применив перестановку $\varphi(a)$ несколько раз. Никакое другое число не получится: начиная с $\varphi(a)^5$, образы числа 11 циклически повторяются.

Ясно, что это рассуждение годится для любой перестановки. Получаем описание циклового разложения перестановки n точек в терминах действия этой перестановки на этих точках: циклы (в том числе и циклы длины 1, то есть неподвижные точки) это и есть орбиты действия. □

В примере мы получили разбиение множества X на орбиты. Оказывается, так будет для любого действия группы на любом множестве.

Утверждение 6.18. *Орбиты действия разбивают точки множества* X *на непересекающиеся множества (классы эквивалентности отношения «точка* x *принадлежит орбите точки* y*»).*

Доказательство. Проверим свойства отношения эквивалентности для отношения «точка x принадлежит орбите точки y».

Рефлексивность: $e(x) = x$, так как при гомоморфизме единичный элемент переходит в единичный.

Симметричность: если $y \in \mathrm{Orb}_x$, то $x \in \mathrm{Orb}_y$, так как из $y = g(x)$ следует $g^{-1}(y) = g^{-1}(g(x) = (g^{-1}g)(x) = e(x) = x$ (во втором равенстве использовано свойство гомоморфизма для действия).

Транзитивность: если $y \in \mathrm{Orb}_x$, $z \in \mathrm{Orb}_y$, то $z \in \mathrm{Orb}_x$, так как из $y = g_1(x)$, $z = g_2(y)$ следует $z = g_2(y) = (g_2 g_1)(x)$ (тут опять использовано свойство гомоморфизма для действия). □

Определение 6.19. Действие называется *транзитивным*, если у него ровно одна орбита. (То есть любая точка переводится в любую другую действием какого-нибудь элемента группы.)

Пример 6.20. Группы многогранников действуют транзитивно и на множестве вершин, и на множестве граней, и на множестве рёбер. □

Рассмотрим несколько примеров нетранзитивных действий из комбинаторики.

Пример 6.21 (продолжение примера 6.3)**.** Для действия перестановок вершин на множестве графов орбитами будут классы изоморфизма графов. Эти орбиты часто называются графами с «непомеченными» вершинами. Смысл названия в том, что графы можно задавать картинками. Если при этом указать на картинке названия вершин (скажем, номера), то получаем граф с помеченными вершинами. А если не указывать, то получаем граф с непомеченными вершинами. Любая разметка такого графа даст какой-то помеченный граф из орбиты, задаваемой непомеченным графом. Эта ситуация проиллюстрирована на рис. 24.

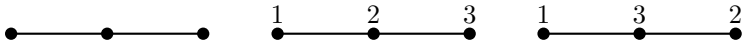


Рис. 24: Непомеченный граф (слева) задаёт орбиту действия перестановок вершин на помеченных графах (две точки этой орбиты изображены справа и в центре)

Неизоморфные графы задают разные (непересекающиеся) орбиты, как следует из утверждения 6.18. На рис. 25 изображены два неизоморфных графа. Нетрудно

Рис. 25: Неизоморфные графы

сообразить, что при любой нумерации вершин левого графа получится помеченный

граф, отличный от всех помеченных графов, получающихся нумерациями вершин правого графа (в первом случае обязательно будет вершина степени 3, а во втором такой вершины никогда не будет, поэтому совпадение графов невозможно). □

Пример 6.22 (продолжение примера 6.2)**.** В том примере мы перенесли действие группы $G = S_n$ с множества $\{1, 2, \ldots, n\}$ на множество P_n, состоящее из всех подмножеств множества $\{1, 2, \ldots, n\}$.

Как хорошо известно, между подмножествами $\{1, 2, \ldots, n\}$ и всюду определёнными функциями из $\{1, 2, \ldots, n\}$ в $\{0, 1\}$ существует биекция. Подмножеству S ставится в соответствие *характеристическая функция* χ_S (другое название — индикаторная функция), которая определяется как

$$\chi_S(x) = \begin{cases} 1, & \text{если } x \in S, \\ 0, & \text{если } x \notin S. \end{cases}$$

Действие S_n на подмножествах переносится на действие на функциях со значениями в множестве $\{0, 1\}$.

В перечислительной комбинаторике используется обобщение этого примера, которое получается так. Действие группы G на множестве X определяет действие той же группы на множестве всюду определённых функций $X \to Y$ из множества X в множество Y: для $g \in G$ и $\alpha\colon X \to Y$ образ α при действии элемента g задаётся равенствами

$$g(\alpha)(x) = \alpha(g^{-1}(x)). \tag{6.1}$$

для всех $x \in X$.

Почему это действие? (И к чему лишнее усложнение формулы в правой части?) Давайте разберёмся. Рекомендуем вспомнить об изоморфизме группы и транспонированной к ней (пример 3.12).

Во-первых, проверим, что заданное формулой (6.1) отображение для любого g является биекцией на множестве функций из X в Y. Заметим, что прообраз β функции $\alpha\colon X \to Y$ обязан для всех $x \in X$ удовлетворять равенству

$$g(\beta)(x) = \beta(g^{-1}(x)) = \alpha(x)$$

и выполнения этого равенства достаточно для того, чтобы $g(\beta) = \alpha$. Поскольку g — биекция на X, написанное равенство равносильно равенству

$$\beta(y) = \alpha(g(y)) \tag{6.2}$$

(для наглядности мы заменили переменную, полагая $y = g^{-1}(x)$).

Равенство (6.2) однозначно задаёт функцию β по функции α. Таким образом, прообраз существует и единственен, отображение (6.1) биекция.

Во-вторых, нужно проверить свойство гомоморфизма. Оно вполне очевидно, но выполним формальную проверку. Нужно доказать, что

$$(gh)(\alpha) = g(h(\alpha)).$$

Раскрывая определение (6.1), получаем для любого $x \in X$ цепочку равенств

$$(gh)(\alpha)(x) = \alpha\big((gh)^{-1}(x)\big) = \alpha((h^{-1}g^{-1})(x)) = \alpha(h^{-1}(g^{-1}(x))) = g(h(\alpha))(x).$$

Из этой выкладки видно, что использование обратного к g элемента в формуле (6.1) важно (по-другому можно было бы сделать так: определить действие транспонированной группы и потом применить изоморфизм между группой и её транспонированной).

Приведём пример, в котором это действие возникает в комбинаторной задаче.

Рассмотрим кубик, грани которого можно красить в несколько цветов, для определённости будем говорить о раскрасках в 3 цвета, в остальных случаях всё аналогично. Кубик можно вращать в пространстве и естественно считать раскраски одинаковыми, если одна получается из другой вращениями (два кубика, одинаково раскрашенных в этом смысле, невозможно различить — один совмещается с другим поворотом).

Занумеруем грани кубика числами от 1 до 6. Тогда раскраскам граней в 3 цвета взаимно однозначно соответствуют функции $\{1,2,3,4,5,6\} \to \{1,2,3\}$.

Группа куба действует на множестве его граней. Раскраски (функции) считаем одинаковыми, если одна получается из другой действием поворотной симметрии куба. Другими словами, раскраски одинаковы, если они лежат в одной орбите этого действия группы куба на функциях.

Этот пример аналогичен определению «непомеченных» графов (классов изоморфизма графов). В каждом случае возникает сложный комбинаторный объект, который является орбитой некоторого действия. □

Теперь рассмотрим связь между стабилизаторами и орбитами. Фактически мы уже использовали эту связь при подсчёте порядков групп многогранников и диэдральной группы.

Лемма 6.23. *Пусть группа G действует на множестве X. Отображение $\varphi\colon y \mapsto \{g \in G : g(x) = y\}$ сопоставляет каждой точке орбиты Orb_x левый смежный класс по стабилизатору Stab_x. Это соответствие взаимно однозначно.*

В частности, для порядков группы, стабилизатора и размера орбиты выполняется соотношение

$$|G| = |\operatorname{Stab}_x| \cdot |\operatorname{Orb}_x|. \tag{6.3}$$

Доказательство. Вначале докажем, что образом любой точки при отображении φ является класс смежности. Условие $g(x) = h(x)$ равносильно условию $h^{-1}(g(x)) = x$, которое означает, что $h^{-1}g \in \operatorname{Stab}_x$ и потому $g \in h \operatorname{Stab}_x$.

Из определения φ ясно, что прообраз класса $g \operatorname{Stab}_x$ при отображении φ определён однозначно: поскольку $g \in g \operatorname{Stab}_x$, то этот прообраз равен $g(x)$.

Осталось доказать, что образ орбиты при отображении φ — всё множество смежных классов по Stab_x. Это очевидно: левый смежный класс $g \operatorname{Stab}_x$ является образом точки $g(x)$ при отображении φ.

Итак, размер орбиты равен индексу стабилизатора, $|\operatorname{Orb}_x| = (G : \operatorname{Stab}_x)$.

Поэтому соотношение (6.3) является переформулировкой теоремы Лагранжа. □

Орбита действия «однородна» относительно действия группы. Это неформальное высказывание можно уточнять разными способами. Например, можно заметить, что стабилизаторы точек вдоль орбиты сопряжены.

Лемма 6.24. *Пусть группа G действует на множестве X, $y \in \mathrm{Orb}_x$. Тогда* $\mathrm{Stab}_y = g\,\mathrm{Stab}_x\, g^{-1}$, *где* $g(x) = y$.

Доказательство. Прямая проверка. Пусть $h \in \mathrm{Stab}_y$, то есть $h(y) = y$. Но это означает, что $h(g(x)) = g(x)$, что равносильно $x = (g^{-1}hg)(x)$. Таким образом, $g^{-1}hg \in \mathrm{Stab}_x$ и $\mathrm{Stab}_y \subseteq g\,\mathrm{Stab}_x\, g^{-1}$.

В обратную сторону: пусть $h \in \mathrm{Stab}_x$. Тогда, многократно применяя свойство гомоморфизма для действия, получаем цепочку равенств

$$(ghg^{-1})(y) = (gh)(g^{-1}(g(x))) = (gh)(x) = g(h(x)) = g(x) = y,$$

то есть $ghg^{-1} \in \mathrm{Stab}_y$ и $g\,\mathrm{Stab}_x\, g^{-1} \subseteq \mathrm{Stab}_y$. □

6.3 Действие группы сдвигами. Теорема Кэли

Перейдём к алгебраическим примерам действий групп. В этом разделе разберём действие группы на себе самой сдвигами.

В этом случае $X = G$, а группа действует по правилу

$$g(h) = gh \tag{6.4}$$

(в данном случае это левый сдвиг на элемент $g \in G$).

Проверим, что это действие. Для любого $g \in G$ отображение $h \mapsto gh$ является биекцией: обратное отображение задаётся правилом $h \mapsto g^{-1}h$.

Контрольный вопрос 6.25. Проверьте, что композиция таких отображений является тождественным отображением на группе.

Проверим свойство гомоморфизма:

$$(g_1g_2)(h) = g_1g_2h = g_1(g_2h) = g_1(g_2(h)).$$

Утверждение 6.26. *Действие группы на себе самой сдвигами точное (гомоморфизм инъективный).*

Доказательство. Если $g(h) = gh = h = eh$, то по закону сокращения $g = e$. □

Отсюда получаем интересный факт.

Теорема 6.27 (теорема Кэли)**.** *Любая конечная группа порядка n изоморфна некоторой подгруппе симметрической группы S_n.*

Доказательство. Группа G порядка n действует сдвигами на себе самой, то есть на множестве из n элементов. Мы уже проверили, что это действие точное. По теореме о гомоморфизмах G изоморфна образу при инъективном гомоморфизме, то есть подгруппе группы перестановок n элементов. □

Теорема Кэли показывает, что симметрические группы в некотором смысле «самые сложные» конечные группы: любая конечная группа является подгруппой какой-нибудь симметрической группы. Вопрос о наименьшем n, для которого данная группа G изоморфна подгруппе S_n в общем случае очень трудный. Рассмотрим два примера.

Пример 6.28 (циклическая группа)**.** Теорема Кэли гарантирует, что $C_n \cong G < S_n$. Из доказательства ясно даже, что это за подгруппа G: она порождена циклом длины n.

Однако это не наименьшее k для которого в S_k существует подгруппа, изоморфная C_n. Действительно, любой элемент группы, порядок которого равен n, порождает подгруппу, изоморфную C_n.

Вспоминая формулу для порядка перестановки, получаем такую неявную характеризацию тех k, для которых в S_k есть элемент порядка n: существует такое разбиение k на слагаемые, $k = k_1 + k_2 + \cdots + k_t$, что $\mathrm{НОК}(k_1, k_2, \ldots, k_t) = n$.

Для небольших значений n минимальное k можно найти перебором. Приведём таблицу

n	1	2	3	4	5	6	7	8	9	10	11	12
k	1	2	3	4	5	5	7	8	9	7	11	7

Проверить корректность этой таблицы нетрудно. Для $n = 1$ всё очевидно. Далее, если $n = p^s$ — степень простого числа, то, как мы уже выясняли ранее, $k = p^s$ (в перестановке порядка $n = p^s$ обязательно есть цикл длины n). В оставшихся случаях $n = 6, 10, 12$ легко подобрать разбиения чисел 5 и 7 с требуемыми свойствами:

$$5 = 2 + 3,\ \mathrm{НОК}(2,3) = 6, \quad 7 = 2 + 5,\ \mathrm{НОК}(2,5) = 10, \quad 7 = 3 + 4,\ \mathrm{НОК}(3,4) = 12.$$

Минимальность этих значений k проверяется перебором возможных разбиений чисел 4 и 6 и вычислением НОК частей этих разбиений. □

Пример 6.29 (диэдральная группа)**.** Случай группы D_n аналогичен. Так как $|D_n| = 2n$, теорема Кэли гарантирует, что $D_n \cong G < S_{2n}$. Однако из определения D_n ясно, что она точно действует на множестве вершин правильного n-угольника, поэтому $D_n \cong G < S_n$.

Пусть $D_n \cong G < S_k$. Как мы выяснили в примере 5.46, это означает, что $G = \langle \rho, \pi \rangle$, где ρ — перестановка порядка n, а π — перестановка порядка 2, причём $\pi \circ \rho \circ \pi = \rho^{-1}$.

Всегда ли для перестановки ρ найдётся перестановка порядка 2 (инволюция, как говорят), сопряжение которой (заметьте, что $\pi^{-1} = \pi$) совпадает с ρ^{-1}? Достаточно найти такую перестановку для одного цикла и мы это уже фактически сделали: поскольку D_n действует на множестве из n вершин, сопряжение цикла длины n посредством отражения относительно прямой переводит цикл в обратный. Для общего случая нужно просто объединить инволюции для циклов, поскольку они не пересекаются.

В частности, получаем интересный факт: D_{12} изоморфна подгруппе S_7. □

Контрольный вопрос 6.30. Проверьте, что $D_{12} \cong \langle (1\ 2\ 3\ 4\ 5\ 6\ 7), (1\ 2)(3\ 4)(5\ 6) \rangle$.

Группа действует сдвигами не только на себе самой, но и на множестве смежных классов по любой подгруппе $H < G$. Правило аналогично (6.4):

$$g(xH) = (gx)H, \tag{6.5}$$

оно показано на рис. 26.

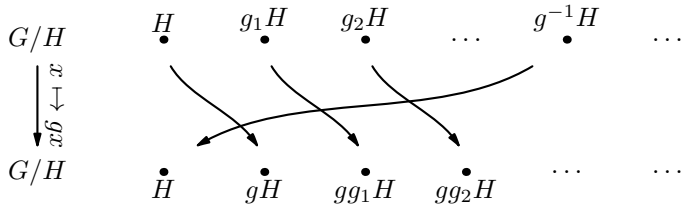


Рис. 26: Действие группы на смежных классах сдвигами

Проверка свойств действия рутинная. Для $g \in G$ обратное отображение к (6.5) задаётся правилом $xH \mapsto (g^{-1}x)H$. Свойство гомоморфизма проверяется точно так же, как для действия (6.4) группы на самой себе сдвигами.

Действие группы сдвигами на смежных классах по подгруппе определено для любой подгруппы, необязательно нормальной. Для нормальной подгруппы это действие совпадает с действием факторгруппы на себе самой сдвигами. В общем случае это действие может быть устроено очень сложно.

Напомним, что действие называется *транзитивным*, если у него ровно одна орбита (любую точку можно перевести в любую действием элемента группы).

Пусть G действует транзитивно на множестве X и Stab_x — стабилизатор некоторой точки x. Как говорит лемма 6.23, каждой точке X, которое в данном случае совпадает с орбитой x, отвечает класс смежности по стабилизатору. Пусть $y = g(z)$, $g \in G$, $y, z \in X$. Тогда класс смежности, отвечающий точке y равен $gh\,\mathrm{Stab}_x$, где класс смежности $h\,\mathrm{Stab}_x$ соответствует точке z. То есть данное действие фактически совпадает с действием G на смежных классах по стабилизатору (мы пропускаем точное определение выражения «фактически совпадает», предлагаем читателю придумать его самостоятельно).

Контрольный вопрос 6.31. Пусть группа G действует сдвигами на смежных классах по подгруппе H. Чему равен стабилизатор H относительно этого действия?

6.4 Действие группы сопряжениями

Есть ещё одно естественное действие группы на себе самой: действие сопряжениями. Оно задаётся правилом

$$g(x) = gxg^{-1}. \tag{6.6}$$

Напомним, что (6.6) задаёт гомоморфизм G в группу автоморфизмов G. Автоморфизмы, разумеется, являются подгруппой $S(G)$, так как любой автоморфизм является биекцией.

Это действие не является точным в общем случае. Как мы помним, ядром гомоморфизма $G \to \mathrm{Aut}\,G$, задаваемого сопряжениями, является центр группы $Z(G)$. Для абелевых групп $Z(G) = G$ и действие сопряжениями тривиально: любой элемент группы оставляет любой элемент группы на месте.

Действие сопряжениями не является транзитивным. Скажем, нейтральный элемент группы всегда образует орбиту, так как для любого $g \in G$ выполняется $g(e) =$

$= geg^{-1} = e$. Орбиты действия сопряжениями — это классы сопряжённости, как нетрудно видеть из определений.

Из леммы 6.23 получаем интересное арифметическое следствие.

Следствие 6.32. *Количество элементов в классе сопряжённости делит порядок группы.*

А что является стабилизатором действия сопряжениями? По определению, Stab_h состоит из тех $g \in G$, для которых $h = ghg^{-1}$, что равносильно $hg = gh$ (то есть это те элементы группы, которые коммутируют с h). Это подгруппа G, как мы знаем, и её стандартное название — *нормализатор* h (обозначение $N(h)$).

Действие сопряжениями полезно в доказательствах разных свойств групп. Для примера рассмотрим p-группы.

Определение 6.33. Группа называется *p-группой*, если её порядок равен степени простого числа p.

Как мы знаем, группы порядка p циклические. Группы порядка p^2 уже не обязательно циклические, например, группа $C_p \times C_p$. Однако все они абелевы, как мы сейчас докажем.

Лемма 6.34. *Пусть порядок группы G равен p^n, p — простое число. Тогда центр группы нетривиален, $Z(G) \neq \{e\}$. Другими словами, в группе обязательно найдётся неединичный элемент, который коммутирует со всеми элементами группы.*

Доказательство. Рассмотрим классы сопряжённости в группе G. Среди них есть единичные, это в точности элементы центра группы. По следствию 6.32 количество элементов в любом классе сопряжённости делит порядок группы. В случае p-группы количество элементов в каждом неединичном классе сопряжённости является положительной степенью p, то есть делится на p.

Запишем разложение порядка группы на слагаемые, равные количествам элементов в классах сопряжённости:

$$|G| = p^n = \underbrace{1 + 1 + \cdots + 1}_{|Z(G)| \text{ штук}} + k_1 + \cdots + k_s, \qquad k_i > 1,\ p \mid k_i.$$

Но тогда и $|Z(G)|$ должен делиться на p. Поэтому $|Z(G)| > 1$. □

Следствие 6.35. *Любая группа G порядка p^2 абелева.*

Доказательство. По предыдущей лемме 6.34 центр G неединичный. Его порядок равен p или p^2.

Первый случай невозможен, так как тогда $|G/Z(G)| = p$ и факторгруппа $G/Z(G)$ обязана быть циклической. Но это невозможно по теореме 4.59.

Остаётся лишь случай $|Z(G)| = p^2 = |G|$, в котором G абелева, так как совпадает со своим центром. □

Пример 6.36. Построим неабелеву группу G порядка p^3. Аналогично предыдущему анализу проверяем, что $|Z(G)| = p$, то есть центр группы порождается некоторым элементом z порядка p.

Факторгруппа по центру $G/Z(G)$ имеет порядок p^2. Она абелева и потому должна быть изоморфна $C_p \oplus C_p$ (проверьте, пользуясь техникой из предыдущей главы). Выберем смежные классы $aZ(G)$, $bZ(G)$, которые соответствуют порождающим $(1,0)$ и $(0,1)$ группы $C_p \oplus C_p$.

Мы хотим, чтобы умножение в группе было некоммутативно. Потребуем, чтобы $ba = abz$, это гарантирует некоммутативность. Тогда

$$(a^i b^j)(a^k b^s) = a^i a^k z^{kj} b^j b^s = a^{i+j} b^{k+s} z^{kj},$$

и общее правило умножения приобретает вид (учитываем, что $z \in Z(G)$):

$$(a^i b^j z^t)(a^k b^s z^u) = a^{i+k} b^{j+s} z^{t+u+kj}, \quad a^p = b^p = z^p = e.$$

Это похоже на правило умножения в группе $\langle a\rangle \times \langle b\rangle \times \langle z\rangle$, за исключением «подкрутки» в третьей координате.

Нужно проверить, что определённая таким образом операция и впрямь задаёт группу.

Легко проверить, что единичный элемент отвечает показателям $i = 0$, $j = 0$, $t = 0$, а обратный к элементу $a^i b^j z^t$ имеет вид $a^{-i} b^{-j} z^{-t+ij}$. Осталось проверить ассоциативность операции. Это требует более развёрнутого вычисления

$$\begin{aligned}(a^i b^j z^t)\big((a^{i'} b^{j'} z^{t'})(a^{i''} b^{j''} z^{t''})\big) &= (a^i b^j z^t)(a^{i'+i''} b^{j'+j''} z^{t'+t''+j'i''}) = \\ = a^{i+i'+i''} b^{j+j'+j''} z^{t+t'+t''+(i'+i'')j+j'i''} &= a^{i+i'+i''} b^{j+j'+j''} z^{t+t'+t''+i'j+(j+j')i''} = \\ = (a^{i+i'} b^{j+j'} z^{t+t'+i'j})(a^{i''} b^{j''} z^{t''}) &= \big((a^i b^j z^t)(a^{i'} b^{j'} z^{t'})\big)(a^{i''} b^{j''} z^{t''}).\end{aligned}$$

□

Пример 6.37. Действие группы сопряжениями переносится на действие на подмножествах группы.

Нормализатором $N(S)$ подмножества $S \subset G$ элементов группы G называется множество таких её элементов g, что выполнено равенство $Sg = gS$. Нормализатор $N(S)$ всегда не пуст ($e \in N(S)$). Если группа коммутативна, то $N(S) = G$.

Аналогично действию на самой группе, нормализатор множества S это стабилизатор действия G на подмножествах G сопряжениями. Поэтому он является подгруппой. □

Контрольный вопрос 6.38. Чему равен $N(H)$, где $H \lhd G$ — нормальная подгруппа?

Пример 6.39. Подгруппы H_1 и H_2 группы G называются сопряжёнными, если $H_2 = gH_1g^{-1}$ для некоторого $g \in G$.

Докажем, что количество подгрупп, сопряжённых данной, является делителем порядка группы. Действительно, подгруппы, сопряжённые данной, образуют орбиту относительно действия группы G сопряжениями на подмножествах G. Из леммы 6.23 получаем, что их количество обязано делить порядок группы. □

Пример 6.40. Пусть G — подгруппа S_8, составленная из перестановок, которые чётные числа переводят в чётные. Сколько в S_8 подгрупп, сопряжённых G?

Сопряжённая G группа $\pi G\pi^{-1}$ сохраняет множество $X = \pi(E)$, где E — множество чётных чисел, так как из $x = \pi(2k)$, $g \in G$ следует $(\pi g\pi^{-1})(x) = (\pi g)(2k) = = \pi(2s) \in X$.

По той же причине верно и обратное: если G_X — подгруппа перестановок, сохраняющая множество X, то $\pi^{-1}G_X\pi$ сохраняет множество чётных чисел E.

Осталось заметить, что $|E| = 4$ и что если g сохраняет X, то она же сохраняет и $\bar{X}$. Поэтому $G_X = G_{\bar{X}}$, а общее количество сопряжённых G групп в два раза меньше, чем количество 4-элементных подмножеств 8-элементного множества.

Ответ: $\frac{1}{2}\binom{8}{4} = 35$. □

6.5 Лемма Бернсайда

Как мы уже видели на примерах, многие задачи перечислительной комбинаторики о подсчёте количества «непомеченных» объектов сводятся к нахождению количества орбит некоторого действия группы. Существует развитая и технически довольно громоздкая теория перечисления для таких задач. Однако в её основе лежит очень простой алгебраический факт, который мы докажем в этом разделе и покажем примеры его использования.

Лемма 6.41 (лемма Бернсайда)**.** *Пусть конечная группа G действует на конечном множестве X. Количество орбит действия даётся формулой:*

$$\#\text{орбит} = \frac{1}{|G|}\sum_{g\in G}|X_g|, \tag{6.7}$$

где $X_g = \{x \in X \mid gx = x\}$ — множество неподвижных точек действия элемента g.

Доказательство. Будем доказывать равносильное равенство

$$|G|\#\text{орбит} = \sum_{g\in G}|X_g|$$

двойным подсчётом. Это означает, что мы выразим количество элементов в некотором множестве и как левую, и как правую части равенства (тогда слева и справа в равенстве стоят одинаковые числа, то есть равенство верное).

С этой целью рассмотрим двудольный граф, доли которого образуют G и X. Ребро (g, x) соединяет элемент группы $g \in G$ и точку множества $x \in X$ тогда и только тогда, когда $g(x) = x$ (эта точка является неподвижной для действия элемента g).

Подсчитаем количество рёбер в таком графе двумя способами: суммируя концы рёбер в левой доле (множество G) и в правой доле (множество X).

В первом случае получаем как раз правую часть равенства: $|X_g|$ в точности равно количеству рёбер, выходящих из вершины g; чтобы найти общее количество рёбер в графе, нужно просуммировать по всем $g \in G$.

Во втором случае количество рёбер, выходящих из вершины x, равно мощности стабилизатора Stab_x: ведь именно элементы группы, входящие в Stab_x оставляют точку x на месте. Значит, количество рёбер в графе равно

$$\sum_{x\in X} \mathrm{Stab}_x = \sum_{x\in X} \frac{|G|}{|\,\mathrm{Orb}_x\,|} = |G| \sum_{x\in X} \frac{1}{|\,\mathrm{Orb}_x\,|}.$$

Здесь мы в очередной раз воспользовались леммой 6.23.

Поскольку орбиты задают разбиение множества X, а сумма $1/|\,\mathrm{Orb}_x\,|$ по каждой орбите очевидно равна 1, то в правой части получаем в точности $|G| \cdot$ #орбит. □

В применениях формулы Бернсайда полезным оказывается такое наблюдение.

Утверждение 6.42. *Пусть группа G действует на множестве X, а элементы группы g_1 и g_2 сопряжены. Тогда $|X_{g_1}| = |X_{g_2}|$.*

Доказательство. Пусть $g_2 = hg_1h^{-1}$. Тогда условие $g_2(x) = x$ равносильно условию $(hg_1h^{-1})(x) = h(g_1(h^{-1}(x))) = x$, что в свою очередь равносильно $g_1(h^{-1}(x)) = h^{-1}(x)$.

Другими словами, $x \in X_{g_2}$ тогда и только тогда, когда $h^{-1}(x) \in X_{g_1}$, то есть $X_{g_2} = h^{-1}(X_{g_1})$. Поскольку преобразование $x \mapsto h^{-1}(x)$ является биекцией, размеры этих множеств одинаковы. □

Пример 6.43 (продолжение примера 6.22)**.** В том примере мы дали точную формулировку такой перечислительной задачи: сколько есть раскрасок граней куба в 3 цвета, если считать одинаковыми раскраски, получающиеся действием группы поворотных симметрий куба?

Как мы выяснили в примере 6.22, нужно найти количество орбит действия группы куба на функциях $\{1,2,3,4,5,6\} \to \{1,2,3\}$.

Чтобы воспользоваться леммой Бернсайда (с учётом утверждения 6.42), нам нужно найти для каждого класса сопряжённости группы куба количество функций $\{1,2,3,4,5,6\} \to \{1,2,3\}$, которые сохраняются при действии симметрий из этого класса сопряжённости.

Группа куба изоморфна S_4, поэтому в ней 5 классов сопряжённости. Мы их описали в примере 6.6. Это тождественное преобразование, повороты на 120° вокруг больших диагоналей (проходящих через противоположные вершины куба), в группе S_4 им соответствуют циклы $(i\ j\ k)$ длины 3, всего таких циклов $2 \cdot 4 = 8$; повороты на 90° вокруг прямых, проходящих через центры противоположных граней, в S_4 им соответствуют циклы длины 4, всего таких циклов $4!/4 = 6$; повороты на 180° вокруг тех же прямых, в S_4 им отвечают пары циклов длины 2, всего таких перестановок 3; и наконец, повороты на 180° вокруг прямых, проходящих через середины противоположных рёбер, в S_4 им отвечают транспозиции, всего транспозиций $\binom{4}{2} = 6$.

Теперь нужно занумеровать грани куба и для каждого класса сопряжённости записать действие элемента из этого класса на множестве граней, то есть чисел от 1 до 6.

Таблица 1: Вычисление $|X_g|$ для классов сопряжённости группы куба

S_4	действие на гранях	элементов в классе	$\lvert X_g\rvert$
()	()	1	3^6
$(i\ j\ k)$	(1 4 2)(3 5 6)	8	3^2
$(i\ j\ k\ l)$	(1 2 6 5)	6	3^3
$(i\ j)(k\ l)$	(1 6)(2 5)	3	3^4
$(i\ j)$	(1 4)(2 5)(3 6)	6	3^3

Результаты вычислений представлены в таблице 1. Первые три столбца заполнены в соответствии с нумерацией, указанной на рис. 27, а представителями классов сопряжённости выбраны повороты, указанные на рис. 23. В последнем столбце указаны величины $|X_g|$ для каждого класса сопряжённости.

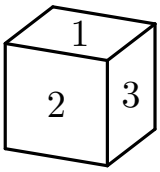


Рис. 27: Сумма номеров на противоположных гранях равна 6

Для первой строки всё очевидно: любая функция $\{1,2,3,4,5,6\} \to \{1,2,3\}$ сохраняется тождественным преобразованием, а всего таких функций 3^6.

Во второй строке указано количество функций, которые сохраняются перестановкой (1 4 2)(3 5 6). Такие функции должны удовлетворять условиям

$$f(1) = f(4) = f(2), \quad f(3) = f(5) = f(6).$$

Поэтому они однозначно задаются значениями $f(1)$ и $f(3)$, причём любая пара значений возможна. Поэтому общее количество функций равно 3^2.

Аналогично для следующих строк. В третьей строке нужно посчитать количество функций, которые сохраняются перестановкой (1 2 6 5), такие функции однозначно задаются значениями $f(1)$, $f(3)$, $f(4)$ и любой набор значений возможен. Всего функций 3^3.

Легко видеть, что в общем случае количество функций равно 3^k, где k — количество циклов в перестановке граней, отвечающей данному классу сопряжённости группы куба (включая циклы длины 1).

С помощью этой таблицы применение леммы Бернсайда сводится к простому арифметическому вычислению:

$$\#\text{раскрасок} = \frac{1}{24}\left(3^6 + 8\cdot 3^2 + 6\cdot 3^3 + 3\cdot 3^4 + 6\cdot 3^3\right) = 57.$$

Итак, есть 57 различных раскрасок граней куба в три цвета. Найти это число прямым перебором вариантов не так уж просто, очень легко пропустить какой-нибудь вариант или посчитать какой-то вариант дважды. □

Пример 6.44 (подсчёт ожерелий)**.** Ожерелье состоит из 12 бусин чёрного или белого цвета, которые жёстко закреплены на круглом кольце через равные расстояния. Ожерелья разные, если их нельзя совместить движением. Сколько есть разных ожерелий?

Занумеруем местá бусин в порядке их следования вдоль кольца и для каждого места укажем цвет бусины. Получили множество N из 2^{12} «помеченных» ожерелий (функций из 12-элементного множества в множество из двух цветов). На этом множестве «помеченных» ожерелий действует диэдральная группа D_{12} и те «помеченные» ожерелья, которые попадают в одну орбиту, задают одинаковое в смысле нашей задачи ожерелье. Задача, как и предыдущая, свелась к подсчёту числа орбит.

Подсчитаем число орбит, пользуясь леммой Бернсайда и утверждением 6.42.

Теперь рассмотрим классы сопряжённости группы D_{12}. Напомним, что мы их находили в примере 4.51 для случая произвольного n.

Первый класс: нейтральный элемент. Он действует тождественно, поэтому сохраняются все $g = e$. Тождественное отображение сохраняет все $2^{12} = 4096$ «помеченных» ожерелий.

Второй класс: повороты на $\pm 2\pi/12$, $\pm 5 \cdot 2\pi/12$. Действие на 12 бусинах таких поворотов будет одним циклов длины 12, так как 1 и 5 взаимно просты с 12. Поэтому неподвижными точками будут только 2^1 «помеченных» ожерелий, в которых все бусины одинакового цвета. Этот случай даёт вклад в формулу (6.7) в размере $4 \cdot 2^1 = 8$.

Третий класс: повороты на $\pm 2 \cdot 2\pi/12$. На чётных и нечётных местах должны стоять бусины одинакового цвета, поэтому общий вклад этого случая $2 \cdot 2^2 = 8$.

Далее аналогично.

Четвёртый класс: повороты на $\pm 3 \cdot 2\pi/12$ даёт вклад $2 \cdot 2^3 = 16$, так как цвета бусин, идущих через две, должны быть одинаковы.

Пятый класс: повороты на $\pm 4 \cdot 2\pi/12$ даёт вклад $2 \cdot 2^4 = 32$,

Шестой класс: поворот на $6 \cdot 2\pi/12$ даёт вклад $1 \cdot 2^6 = 64$.

Осталось рассмотреть ещё два класса: отражения относительно диагонали 12-угольника, в этом случае вклад равен $6 \cdot 2^7 = 768$; и отражения относительно прямой, проходящей через середины противоположных сторон, в этом случае вклад равен $6 \cdot 2^6 = 384$.

Собирая эти вычисления вместе, получаем ответ:

$$\frac{1}{24}\Big(4096 + 8 + 8 + 16 + 32 + 64 + 768 + 384\Big) = \frac{5376}{24} = 224$$

различных ожерелья. □

Пример 6.45. Рассмотрим ещё один пример применения леммы Бернсайда и решим такую задачу.

На каждой грани правильного тетраэдра рисуют стрелку, соединяющую середину ребра с противоположной вершиной. Тетраэдр можно поворачивать, при этом расстановки стрелок, которые совмещаются поворотом тетраэдра, будем считать одинаковыми. Найдите количество таких расстановок.

Расстановка стрелок однозначно задаётся выбором в каждой грани одной из её вершин (конец стрелки, начало обязано лежать на середине ребра между двумя другими вершинами грани).

Если занумеровать грани и вершины тетраэдра числами от 1 до 4, то речь будет идти о действии группы тетраэдра на некотором подмножестве функций из $\{1,2,3,4\}$ в $\{1,2,3,4\}$. Некотором, потому что одну из вершин тетраэдра данной грани сопоставить невозможно. Обратите также внимание, что это не то действие, которое мы определили в примере 6.22, поскольку симметрия тетраэдра переставляет и грани, и вершины.

Грани тетраэдра занумеруем произвольно, а вершины по правилу: номер вершины равен номеру противоположной этой вершине грани. Тогда интересующее нас множество функций задаётся условиями

$$f(x) \neq x \tag{6.8}$$

для всех x. Всего таких функций 3^4 (для каждой грани есть три возможных варианта значения функции), будем называть их *корректными*.

В примере 6.4 мы нашли классы сопряжённости группы тетраэдра. Их три: тождественное преобразование; повороты на 120° вокруг прямых, проходящих через вершину и середину противоположной стороны, таких поворотов $2 \cdot 4 = 8$; повороты на 180° вокруг прямых, проходящих через середины скрещивающихся рёбер (см. рис. 22), таких поворотов $6/2 = 3$.

Для применения формулы из леммы Бернсайда нужно найти количество функций, остающихся неизменными при действиях симметрий из каждого класса сопряжённости.

Первый класс: тождественное преобразование. Оставляет неизменными все $3^4 = 81$ функций.

Второй класс: представитель — поворот R на 120° вокруг прямой, проходящей через вершину 1. Такие повороты не оставляют неизменными ни одну корректную функцию. Действительно, рассмотрим некоторую функцию и пусть $f(1) = a \neq 1$. Рассматриваемый поворот R переводит грань 1 в себя, но не оставляет на месте ни одну вершину этой грани. Поэтому $Rf(1) = R(a) \neq a$, что и означает изменение функции.

Третий класс: представитель — поворот R на 180° вокруг прямой, проходящей через середины рёбер $(1,2)$ и $(3,4)$ (указаны номера вершин). Действие R и на гранях, и на вершинах задаётся при нашей нумерации одной и той же перестановкой $(1\ 2)(3\ 4)$. Чтобы различать действие R на гранях и вершинах, будем обозначать перестановку F и V соответственно. Условие сохранения функции f при действии R имеет вид:

$$V(f(F^{-1}(i))) = f(i).$$

Упражнение 6.46. Докажите справедливость последнего рассуждения, обобщив рассуждения из примера 6.22.

Теперь уже видно, что можно произвольно (но с выполнением условия (6.8)) выбрать значения $f(1)$ и $f(3)$, после чего значения инвариантной при действии R функции однозначно восстанавливаются: $f(2) = V(f(1))$, $f(4) = V(f(3))$.

Поэтому общее количество инвариантных относительно R функций равно 3^2.

Применяя лемму Бернсайда, получаем

$$\#\text{расстановок стрелок} = \frac{1}{12}\left(3^4 + 3 \cdot 3^2\right) = \frac{81 + 27}{12} = 9.$$

□

7 Кольца и поля: определения, примеры и простейшие свойства

Во второй части курса мы рассмотрим примеры алгебраических систем с двумя (бинарными) операциями. В начальной главе 1 мы уже рассматривали такие примеры. Сейчас дадим формальные определения, вернёмся к этим примерам и рассмотрим некоторые другие.

7.1 Виды колец, делители нуля, нильпотентные элементы

Начнём с определения кольца.

Определение 7.1. Кольцо — это множество R с двумя бинарными операциями сложения (обозначается $+$) и умножения (обозначается $\cdot$, иногда опускается, как это принято в формулах элементарной алгебры), для которых выполняются следующие свойства (*аксиомы кольца*):

R1: относительно сложения R — коммутативная группа (которая называется *аддитивной группой кольца*), нейтральный элемент относительно сложения называется нулём и обозначается обычно как 0;

R2: умножение ассоциативно;

R3: $a \cdot (b + c) = a \cdot b + a \cdot c$; $(b + c) \cdot a = b \cdot a + c \cdot a$ (*дистрибутивность* умножения относительно сложения слева и справа).

В этом определении от операции умножения самой по себе требуется очень немногое. Однако условие дистрибутивности «привязывает» умножение к сложению и приводит к интересным следствиям, выполняющимся для любых колец. Мы рассмотрим эти свойства далее, а пока приведём примеры и выделим важные для дальнейшего классы колец.

Пример 7.2 (числовые кольца)**.** Обычные числовые системы — целые, рациональные, действительные и комплексные числа — являются кольцами относительно обычных операций сложения и умножения чисел.

Эти кольца обладают дополнительными свойствами: например, в них существует нейтральный элемент относительно умножения и умножение коммутативно. □

Определение 7.3. Кольцо называется *коммутативным*, если умножение в кольце коммутативно: $xy = yx$ для любых элементов кольца.

Определение 7.4. Кольцо называется *кольцом с единицей*, если в нём есть нейтральный элемент относительно умножения. Этот элемент называется единицей и обозначается 1. Таким образом, $a \cdot 1 = 1 \cdot a = a$ для любого элемента a кольца.

Пример 7.5. Чётные целые числа образуют кольцо относительно обычных операций сложения и умножения. Это кольцо, разумеется, коммутативно, как и обычное кольцо целых чисел $\mathbb{Z}$. Однако в нём нет единицы: из $xy = y$ для целых чисел следует $x = 1$ (закон сокращения), а 1 — нечётное число. □

Пример 7.6 (кольца матриц)**.** Важный пример некоммутативных колец — кольца матриц. Пусть R — некоторое кольцо. Тогда кольцо $M_n(R)$ состоит из квадратных матриц размера n с элементами из кольца R. Операции сложения и умножения определяются как обычно для матриц:

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{pmatrix} + \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{pmatrix} =$$

$$= \begin{pmatrix} x_{11}+y_{11} & x_{12}+y_{12} & \dots & x_{1n}+y_{1n} \\ x_{21}+y_{21} & x_{22}+y_{22} & \dots & x_{2n}+y_{2n} \\ \dots & \dots & \dots & \dots \\ x_{n1}+y_{n1} & x_{n2}+y_{n2} & \dots & x_{nn}+y_{nn} \end{pmatrix}$$

$$\begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{pmatrix} \cdot \begin{pmatrix} y_{11} & y_{12} & \dots & y_{1n} \\ y_{21} & y_{22} & \dots & y_{2n} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nn} \end{pmatrix} = \begin{pmatrix} z_{11} & z_{12} & \dots & z_{1n} \\ z_{21} & z_{22} & \dots & z_{2n} \\ \dots & \dots & \dots & \dots \\ z_{n1} & z_{n2} & \dots & z_{nn} \end{pmatrix},$$

$$z_{ij} = \sum_{k=1}^{n} x_{ik} y_{kj}.$$

Выполнение аксиомы R1 очевидно: $M_n(R)$ относительно сложения является прямой суммой n^2 аддитивных групп кольца R (как видно из определения, сложение матриц покомпонентное).

Проверка двух других аксиом требует прямого вычисления с суммами. Ассоциативность умножения (аксиома R2):

$$(X(YZ))_{ij} = \sum_{k=1}^{n} X_{ik}(YZ)_{kj} = \sum_{k=1}^{n} X_{ik} \sum_{p=1}^{n} Y_{kp} Z_{pj} =$$

$$= \sum_{k,p=1}^{n} X_{ik} Y_{kp} Z_{pj} =$$

$$= \sum_{p=1}^{n} \left(\sum_{k=1}^{n} X_{ik} Y_{kp} \right) Z_{pj} = \sum_{p=1}^{n} (XY)_{ip} Z_{pj} =$$

$$= ((XY)Z)_{ij}$$

Дистрибутивность (проверяем левую дистрибутивность, для правой аналогично) использует дистрибутивность в кольце матричных элементов R:

$$(X(Y+Z))_{ij} = \sum_{k=1}^{n} X_{ik}(Y+Z)_{kj} = \sum_{k=1}^{n} (X_{ik}Y_{kj} + X_{ik}Z_{kj}) =$$

$$= \left(\sum_{k=1}^{n} X_{ik} Y_{kj} \right) + \left(\sum_{k=1}^{n} X_{ik} Z_{kj} \right) = (XY)_{ij} + (XZ)_{ij} = (XY + XZ)_{ij}$$

Некоммутативность умножения матриц легко продемонстрировать на примере кольца с единицей и матриц порядка 2:

$$\begin{pmatrix}0&1\\0&0\end{pmatrix}\cdot\begin{pmatrix}0&0\\0&1\end{pmatrix}=\begin{pmatrix}0&1\\0&0\end{pmatrix}\qquad\begin{pmatrix}0&0\\0&1\end{pmatrix}\cdot\begin{pmatrix}0&1\\0&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix}$$

Если в кольце матричных элементов R есть единица, то в кольце матриц $M_n(R)$ также есть единица (будем обозначать эту матрицу I): матрица, у которой элементы на главной диагонали равны 1, а вне главной диагонали равны 0, то есть

$$I_{jk}=\begin{cases}1, & \text{если } j=k,\\ 0 & \text{в противном случае.}\end{cases}$$

Проверка свойства единицы очень простая:

$$(XI)_{ij}=\sum_k X_{ik}I_{kj}=X_{ij}=\sum_k I_{ik}X_{kj}=(IX)_{ij}.$$

Равенство

$$\begin{pmatrix}0&1\\0&0\end{pmatrix}\cdot\begin{pmatrix}0&0\\0&1\end{pmatrix}=\begin{pmatrix}1&0\\0&1\end{pmatrix}\cdot\begin{pmatrix}0&0\\0&1\end{pmatrix}$$

показывает, что в кольцах матриц не выполняется, вообще говоря, закон сокращения. □

Определение 7.7. Элемент $a\neq 0$ кольца R называется *левым делителем нуля*, если существует такой $b\neq 0$, что $ab=0$. Аналогично, a называется *правым делителем нуля*, если существует такой $b\neq 0$, что $ba=0$.

Для коммутативных колец разницы между левыми и правыми делителями нуля нет, поэтому говорят просто о делителях нуля.

Пример 7.8. Равенство

$$\begin{pmatrix}0&0\\0&1\end{pmatrix}\cdot\begin{pmatrix}0&1\\0&0\end{pmatrix}=\begin{pmatrix}0&0\\0&0\end{pmatrix}$$

показывает, что в кольцах матриц делители нуля встречаются.

Из линейной алгебры легко вывести критерии левого и правого делителя нуля для матрицы.

Матрица A является и левым, и правым делителем нуля тогда и только тогда, когда она вырождена.

Условие вырожденности необходимо, так как из существования обратной матрицы A^{-1} и равенства $AX=0$ следует $A^{-1}(AX)=IX=X=0$. Аналогично справа: из равенства $XA=0$ следует $(XA)A^{-1}=XI=X=0$.

Из теоремы о ранге и теоремы о разрешимости систем однородных линейных уравнений заключаем, что вырожденность A равносильна как существованию такого вектора-столбца x, что $Ax=0$, так и существованию такой строки y, что $yA=0$. Но тогда нетрудно проверить, что $AX=0$ и $YA=0$, где каждый столбец матрицы X равен x, а каждая строка матрицы Y равна y. □

Не всегда множества левых и правых делителей нуля совпадают. Рассмотрим пример.

Пример 7.9. Пусть множество R_n состоит из матриц порядка $n > 2$ с действительными коэффициентами, у которых все строки, начиная со второй, нулевые. Это множество является кольцом относительно обычных операций матричного сложения и умножения. Так как сложение матриц покомпонентное, то такие матрицы относительно сложения образуют группу (она изоморфна группе $\mathbb{R}^n$ по сложению). Замкнутость относительно умножения нужно проверить. Пусть $A, B \in R_n$. Тогда

$$(AB)_{ij} = \sum_{k=1}^{n} a_{ik} b_{kj} = a_{i1} b_{1j},$$

при $i > 1$ получаем $(AB)_{ij} = 0$.

Из этой же формулы видно, как умножаются первые строки матриц из R_n: первая строка AB — это первая строка B, умноженная на самый левый верхний элемент A_{11} матрицы A. Поэтому все матрицы с $A_{11} = 0$ и только они являются левыми делителями нуля.

С другой стороны, любая матрица из R_n является правым делителем нуля, как показывает та же самая формула (умножение любой матрицы $B \in R_n$ слева на матрицу, в первой строке которой стоит $(0, 1, \ldots, 1)$ даст нулевую матрицу).

Этот же пример показывает, что в кольце могут быть левые единицы, но не быть правых единиц. Левыми единицами, для которых выполняется равенство $EX = X$ для всех $X \in R_n$, будут матрицы с $E_{11} = 1$. А правых единиц в этом кольце нет: равенство $XE = X$ означает в этом кольце, что первые строки матриц X и E пропорциональны. Но среди строк длины $\geqslant 2$ есть непропорциональные. Поэтому ни одна матрица кольца R_n не является правой единицей. □

Между нарушением закона сокращения и существованием делителей нуля есть связь. Чтобы её установить, нам потребуются некоторые общие свойства операций в кольце.

Определение 7.10. *Вычитание* в кольце — это сложение с противоположным, то есть $x - y = x + (-y)$, где в правой части $-y$ обозначает противоположный к y элемент кольца: $y + (-y) = 0$.

Утверждение 7.11 (дистрибутивность вычитания)**.** *В любом кольце $a(b-c) = ab-ac$, $(b-c)a = ba-ca$ для любых элементов a, b, c.*

Доказательство. Докажем дистрибутивность слева. Из дистрибутивности сложения получаем

$$a(b-c) + ac = a(b + (-c) + c) = a(b+0) = ab,$$

дистрибутивность вычитания получается из этого равенства вычитанием ac из обеих частей. Заметим, что по сложению кольцо является группой, так что относительно сложения закон сокращения в кольце выполняется всегда. □

Контрольный вопрос 7.12. Докажите правую дистрибутивность вычитания (второе равенство в утверждении 7.11).

Утверждение 7.13. *В любом кольце* $a \cdot 0 = 0 = 0 \cdot a$.

Доказательство. Используя дистрибутивность вычитания, получаем

$$a \cdot 0 = a(b - b) = ab - ab = 0.$$

Аналогично для второго равенства. □

Это утверждение отчасти объясняет, почему в определении делителей нуля нас интересуют только ненулевые элементы: умножение на 0 в любом кольце даёт 0.

Ещё одно полезное общее свойство колец в школьной математике называется «правилом знаков».

Утверждение 7.14. *Пусть* R *— кольцо с* 1. *Тогда* $(-1) \cdot (-1) = 1$.

Доказательство. Применяя утверждение 7.13 и дистрибутивность, получаем

$$0 = (-1) \cdot 0 = (-1) \cdot (1 + (-1)) = (-1) + (-1) \cdot (-1)$$

и правило знаков получается прибавлением к обеим частям равенства 1. □

Для коммутативных колец запас простых и привычных тождеств больше. Приведём пример.

Утверждение 7.15. *В коммутативном кольце* $a^2 - b^2 = (a - b)(a + b)$.

Доказательство. Раскрывая скобки (то есть пользуясь дистрибутивностью), получаем из коммутативности кольца

$$(a - b)(a + b) = a^2 + (-b)a + ab + (-b)b = a^2 + (-b)b + a((-b) + b) = a^2 + (-b)b.$$

Осталось доказать, что $-b^2 = (-b)b$. Прибавив к обеим частям равенства b^2, получим $0 = b^2 + (-b)b = (b + (-b))b = 0 \cdot b$. □

Теперь сформулируем связь между законом сокращения и отсутствием делителей нуля.

Лемма 7.16. *Если элемент* $a \neq 0$ *кольца* R *не является левым делителем нуля, то из* $ax = ay$ *следует* $x = y$. *И наоборот: если элемент* $a \neq 0$ *кольца* R *является левым делителем нуля, то для некоторых* $x \neq y$ *выполняется* $ax = ay$.

Доказательство. Дистрибутивность вычитания означает, что из равенства $ax = ay$ следует равенство $ax - ay = a(x - y) = 0$. Если $a \neq 0$ не является делителем нуля, то из этого равенства следует $x - y = 0$, то есть $x = y$.

Второе утверждение совсем просто: пусть $a \neq 0$ является делителем нуля, то есть $ab = 0 = a \cdot 0$, $b \neq 0$. Это и есть второе утверждение леммы. □

Аналогичное утверждение справедливо, конечно, и для правых делителей нуля.

Закон сокращения очень удобен в рассуждениях, как и отсутствие делителей нуля. Вспомним, что для обычных чисел из равенства $ab = 0$ следует, что $a = 0$ или $b = 0$. Это следствие и есть утверждение об отсутствии делителей нуля. В такой форме оно используется, например, при решении уравнений.

Целые числа являются коммутативным кольцом без делителей нуля. Эти свойства целых чисел настолько важны, что оказывается удобным выделить их и ввести особый класс колец.

Определение 7.17. Коммутативное кольцо с единицей и без делителей нуля называется *областью целостности*.

Название «область целостности» говорит о близости таких колец к кольцам целых чисел. Например, любое подкольцо комплексных чисел будет областью целостности.

Пример 7.18 (гауссовы целые)**.** Это кольцо состоит из комплексных чисел с целыми действительной и мнимой частями. Легко видеть, что такое множество комплексных чисел замкнуто относительно сложения и умножения. Аксиомы кольца выполняются в нём, так как они выполняются для комплексных чисел. Делителей нуля нет по той же причине — их нет в комплексных числах. □

Однако эту близость не нужно понимать буквально. В частности, самый естественный пример колец, которые получаются из целых чисел, — кольца вычетов — не всегда оказывается областью целостности.

Пример 7.19 (кольца вычетов)**.** Мы уже фактически проверяли, что вычеты по модулю n образуют кольцо. Будем его обозначать $\mathbb{Z}/(n)$ (смысл этого обозначения станет ясен дальше).

Алгебраические свойства кольца вычетов зависят от n. Скажем, в кольце $\mathbb{Z}/(4)$ есть делители нуля, так как $2 \cdot 2 = 0$ в этом кольце.

Нетрудно видеть, что если $n = pq$ составное, то вычеты p и q являются делителями нуля: в этом кольце $p \cdot q = n = 0$.

Верно и обратное: если n — простое, то $\mathbb{Z}/(n)$ является областью целостности. Этот факт по сути доказан в лемме 2.57 и на нём основано доказательство однозначности разложения целых на простые. □

Пример кольца $\mathbb{Z}/(4)$ показывает, что среди делителей нуля встречаются совсем удивительные: такие элементы, некоторая степень которых равна 0.

Определение 7.20. Ненулевой элемент a кольца называется *нильпотентным*, если для какого-то целого положительного $k > 1$ выполняется равенство $a^k = 0$.

Пример 7.21. Вычисление

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

показывает, что в кольце матриц с действительными элементами есть нильпотентные элементы. □

Чтобы поупражняться с определениями, проверим несколько простых свойств нильпотентных элементов и делителей нуля.

Из мультипликативного свойства нуля $a \cdot 0 = 0 \cdot a = 0$ следует, что если $a^n = 0$, то и $a^{n+1} = 0$. Поэтому множество тех положительных целых чисел n, для которых $a^n = 0$, либо пусто, либо имеет вид $\{n : n \geqslant k\}$. Это наименьшее число k называется *индексом нильпотентности* элемента a.

Утверждение 7.22. *Любой нильпотентный элемент является делителем нуля.*

Доказательство. Если k — индекс нильпотентности элемента a, то $a^k = a \cdot a^{k-1} = 0$ и $a^{k-1} \neq 0$. Это и означает, что a — делитель нуля. □

Докажем похожее утверждение про делители нуля.

Утверждение 7.23. *Если a^k, $k > 0$, — делитель нуля, то и a — делитель нуля.*

Доказательство. Рассмотрим множество D тех положительных чисел d, для которых a^d является левым делителем нуля.

Пусть $d \in D$. Докажем, что $a^{d-1} \in D$. Отсюда будет следовать, что все a^k, $1 \leqslant k \leqslant d$, принадлежат D, то есть являются левыми делителями нуля.

Пусть для некоторого $b \neq 0$ выполняется $a^d b = 0$, $a^d \neq 0$. Второе условие гарантирует, что a^{d-1} также не 0. Получаем $a^{d-1}(ab) = 0$. Если $ab \neq 0$, получаем, что a^{d-1} — левый делитель нуля. Если $ab = 0$, то a — левый делитель нуля. При $d > 2$ имеем $a^{d-1}b = a^{d-2} \cdot (ab) = a^{d-2} \cdot 0 = 0$, так что a^{d-1} и в этом случае является левым делителем нуля.

Для правых делителей нуля рассуждение аналогичное. □

Легко видеть, что в коммутативном кольце произведение нильпотентных элементов либо равно 0, либо является нильпотентным элементом: если $a^k = 0$ и $b^n = 0$, то

$$(ab)^{\max(k,n)} = a^{\max(k,n)} b^{\max(k,n)} = 0.$$

Аналогичное утверждение справедливо и для суммы нильпотентных элементов, но в его доказательстве есть небольшая тонкость.

Лемма 7.24. *Нильпотентные элементы коммутативного кольца и 0 образуют подкольцо.*

Доказательство. Как мы только что проверили, это множество замкнуто относительно умножения. Оно содержит 0 по определению.

Проверим, что это множество замкнуто относительно сложения. Пусть $a^k = 0$ и $b^n = 0$. Тогда $(a + b)^{n+k}$ является суммой произведений $a^i b^j$, $i + j = k + n$. Поэтому или $i \geqslant k$, или $j \geqslant n$. В любом из этих случаев каждое такое слагаемое равно нулю, то есть $(a + b)^{n+k} = 0$.

Однако мы ещё не доказали, что по сложению нильпотентные элементы и 0 образуют группу. Требуется замкнутость относительно взятия противоположного

элемента. Заметим, что в коммутативном кольце выполняется обычное равенство $a^2 - b^2 = (a-b)(a+b)$ (утверждение 7.15). Поэтому

$$a^2 - (-a)^2 = (a-(-a))(a+(-a)) = (a+a)\cdot 0 = 0,$$

то есть $a^2 = (-a)^2$. Значит, из $a^n = 0$ следует

$$(-a)^{2n} = ((-a)^2)^n = (a^2)^n = a^{2n} = 0.$$

На этом доказательство закончено. □

Для некоммутативных колец это утверждение неверно.

Пример 7.25. Мы уже видели примеры нильпотентных элементов в кольце матриц

$$\begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

Однако сумма этих нильпотентных матриц обратима

$$\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

поэтому нильпотентные матрицы не замкнуты относительно сложения. □

Аналогично делителям нуля определяются *делители единицы*, или обратимые элементы.

Определение 7.26. Элемент a кольца с единицей имеет *левый обратный* b, если $ba = 1$. Аналогично, элемент a имеет *правый обратный* b, если $ab = 1$.

Элемент называется *обратимым*, если у него есть и левый, и правый обратные.

Из определения ясно, что любой обратимый элемент не является делителем нуля. Мы уже проводили это рассуждение для матриц, повторим его в общем виде.

Утверждение 7.27. *Пусть a имеет правый обратный b. Тогда a не является правым делителем нуля.*

Аналогично, из существования левого обратного следует, что элемент не является левым делителем нуля.

Обратимый элемент не является никаким делителем нуля.

Доказательство. Третье утверждение следует из первых двух. Докажем первое, второе доказывается аналогично.

Из свойства правого обратного получаем $xab = x \cdot 1 = x$. Поэтому из равенства $xa = 0$ следует $x = xab = 0 \cdot b = 0$. Значит, a не является правым делителем нуля. □

Утверждение 7.28. *Пусть a — нильпотентный элемент кольца R с 1. Тогда элемент $1 - a$ обратим.*

Доказательство. Пусть $a^k = 0$. Тогда, используя дистрибутивность, получаем

$$(1-a)(1+a+a^2+\cdots+a^{k-1}) = (1+a+a^2+\cdots+a^{k-1}) - (a+a^2+\cdots+a^k) = 1-a^k = 1,$$

то есть нашли правый обратный. Перестановкой множителей получаем левый обратный (умножение степеней a коммутативно). Это и означает обратимость элемента a. □

Мы уже выяснили, что делители нуля есть в кольце вычетов по модулю n тогда и только тогда, когда n составное. Теперь разберёмся, для каких n в кольце вычетов по модулю n есть нильпотентные элементы.

Лемма 7.29. *Кольцо $\mathbb{Z}/(n)$ содержит нильпотентные элементы тогда и только тогда, когда n делится на p^s, где p — простое, а $s > 1$.*

Доказательство. Запишем разложение n по степеням простых

$$n = p_1^{d_1} p_2^{d_2} \cdot \ldots \cdot p_t^{d_t}.$$

Пусть a — ненулевой остаток по модулю n. Запишем и его разложение по степеням простых:

$$a = p_1^{f_1} p_2^{f_2} \cdot \ldots \cdot p_t^{f_t}.$$

Здесь мы считаем, что $p_1, \ldots, p_t$ — все простые делители чисел n или a, а в разложении $d_i = 0$ ($f_i = 0$), если n (a) не делится на p_i. Мы не включаем в список простых лишние числа, поэтому без ограничения общности полагаем, что одновременно d_i и f_i в ноль не обращаются.

Как в этих терминах выразить условие $a^k = 0$ в кольце $\mathbb{Z}/(n)$? Оно равносильно делимости a^k на n, что, в свою очередь, равносильно неравенствам на показатели $kf_i \geqslant d_i$ для всех $1 \leqslant i \leqslant t$. Ясно, что при достаточно большом k эти неравенства будут выполнены, если $f_i \neq 0$ для всех i (напомним, что мы договорились не брать лишних простых, поэтому одновременное обращение в 0 и d_i, и f_i невозможно).

Но условие $a \neq 0$ означает выполнение неравенств $f_i \leqslant d_i$, причём хотя бы одно из неравенств строгое. Если все $d_i = 1$, соблюсти оба ограничения невозможно (или $f_i = 0$, или $f_i \geqslant 1 = d_i$). Поэтому в этом случае нильпотентных элементов в кольце $\mathbb{Z}/(n)$ нет.

Если, скажем, $d_1 > 1$, то нильпотентные элементы есть: это, например, вычет, который содержит остаток n/p_1. При $d_1 > 1$ у числа $a = n/p_1$ в точности те же простые делители, что и у числа n. Поэтому достаточно большая степень a делится на n. □

Пример 7.30. В кольце $\mathbb{Z}/(15)$ есть делители нуля: $3 \cdot 5 = 0$, но нет нильпотентных элементов, так как $15 = 3^1 \cdot 5^1$. □

Пример 7.31. А в кольце $\mathbb{Z}/(10^6)$ нильпотентные элементы есть, так как $10^6 = 2^6 \cdot 5^6$.

Нетрудно даже подсчитать их количество. Из предыдущего рассуждения видно, что нильпотентные вычеты — это в точности те вычеты, которые содержат остатки, делящиеся и на 2, и на 5, то есть кратные 10. Количество таких ненулевых остатков равно $10^6/10 - 1 = 99\,999$. □

7.2 Поля, определение и простейшие свойства

Кольцо вычетов $\mathbb{Z}/(p)$, где p — простое, обладает даже более сильным свойством, чем отсутствие делителей нуля. А именно, ненулевые элементы этого кольца образуют группу по умножению (любой ненулевой вычет взаимно прост с p и потому обратим по модулю p, это одна из первых наших теорем, теорема 1.48).

В любом кольце $a \cdot 0 = 0 = 0 \cdot a$. Поэтому в случае $0 \neq 1$ всё кольцо не может быть группой по умножению. А вот если отбросить 0, то оставшиеся элементы могут образовывать группу по умножению, как показывают примеры рациональных, действительных и комплексных чисел, а также вычетов по модулю простого числа. Приходим к важному определению.

Определение 7.32. Кольцо с $0 \neq 1$ называется *телом*, если ненулевые элементы кольца образуют группу по умножению.

Если тело вдобавок коммутативно, то оно называется *полем*.

Итак, поле — это такое множество F с двумя операциями, называемыми сложением и умножением, что выполняются следующие свойства:

F1: Относительно сложения F является абелевой группой, нейтральный элемент обозначается 0.

F2: Относительно умножения $F^* = F \setminus \{0\}$ является абелевой группой, нейтральный элемент которой обозначается 1.

F3: Выполняется аксиома дистрибутивности: $a(b+c) = ab + ac$ (из-за коммутативности умножения достаточно потребовать лишь левой дистрибутивности).

F4: $0 \neq 1$.

Пример 7.33 (простые поля)**.** Рациональные числа с операциями сложения и умножения образуют поле $\mathbb{Q}$. Вычеты по модулю простого числа p также образуют поле. В нём p элементов и помимо обозначения $\mathbb{Z}/(p)$ мы будем использовать обозначение $\mathbb{F}_p$. □

Поскольку ненулевые элементы поля образуют группу по умножению, каждый ненулевой элемент обратим. Это означает, как мы уже видели, что в поле нет делителей нуля. Этот простой признак позволяет отсеивать многие кольца при выборе полей.

Пример 7.34. Кольцо вычетов $\mathbb{Z}/(p^n)$ по модулю степени простого при $n \geqslant 2$ содержит делители нуля: $[p] \cdot [p^{n-1}] = [p^n] = 0$. Поэтому оно заведомо не является полем. □

Замечание 7.35. Тем не менее поля из p^n элементов существуют при любом n для любого простого p. Далее мы изложим технику, позволяющую строить такие поля, выполнять вычисления в них и изучать их свойства (очень интересные).

Как раз конечные поля нас и будут интересовать главным образом. Однако окажется, что многие бесконечные поля устроены похожим образом (конечные расширения простых полей).

Объясним, чем выделяются простые поля. Это «наименьшие» поля в том смысле, что любое поле содержит ровно одно простое подполе, а в простом поле подполей нет. (Как обычно, подполем называется такое подмножество элементов поля, которое само является полем относительно тех же операций.)

Лемма 7.36. *Каждое поле содержит простое подполе. В простом поле подполей нет.*

Доказательство. Рассмотрим какое-нибудь поле $\Bbbk$. Оно содержит 0 и 1 и является группой по сложению, как всякое уважающее себя кольцо. Выделим в аддитивной группе поля $\Bbbk$ подгруппу $\langle 1\rangle$, порождённую единицей. Эта группа состоит из элементов вида

$$\pm(\underbrace{1+1+\cdots+1}_{n \text{ слагаемых}}),$$

которые естественно обозначать целыми числами (всякое целое число — это сумма какого-то количества единиц или какого-то количества минус единиц). Мы так и будем делать далее (и уже делали выше), несмотря на некоторую путаницу в обозначениях целых чисел и элементов полей. Элементы этой группы называются *кратными единицы*.

Группа кратных единицы замкнута не только относительно сложения, но и относительно умножения, как показывает равенство

$$m\cdot n=(\underbrace{1+1+\cdots+1}_{m \text{ слагаемых}})\cdot(\underbrace{1+1+\cdots+1}_{n \text{ слагаемых}})=\underbrace{1+1+\cdots+1}_{mn \text{ слагаемых}}=mn \qquad (7.1)$$

(если раскрыть скобки, пользуясь дистрибутивностью, получится как раз mn слагаемых).

Далее нужно рассмотреть два случая.

I. Пусть порядок подгруппы $\langle 1\rangle$ в аддитивной группе поля конечен и равен p. Тогда p — простое число. Действительно, $p=0$ в поле (сумма p единиц равна нейтральному элементу аддитивной группы, то есть нулю). Если $p=mn$ составное, то $m\neq 0$, $n\neq 0$, а $mn=p=0$. А в поле делителей нуля нет, пришли к противоречию.

Формула (7.1) и аналогичная формула для сложения кратных единицы показывает, что кратные единицы складываются и умножаются по модулю p. Поэтому они образуют простое поле $\mathbb{Z}/(p)$ вычетов.[4)]

II. Пусть порядок подгруппы $\langle 1\rangle$ в аддитивной группе поля бесконечен. Тогда все кратные единицы в поле различны и им взаимно однозначно соответствуют целые числа.

[4)] Аккуратнее говорить, что они образуют поле, изоморфное полю вычетов. Далее мы введём и изучим изоморфизмы колец и полей. Пока эту тонкость можно игнорировать.

Множество целых чисел не является полем, поскольку незамкнуто относительно взятия обратного. В поле $\Bbbk$, тем не менее, обратные есть у любого ненулевого элемента поля. Будем обозначать обратный к n в поле $\Bbbk$ через n^{-1}, как обычно.

Рассмотрим множество элементов поля вида mn^{-1}, где $m, n \in \langle 1 \rangle$ — кратные единицы, $n \neq 0$. Проверим, что это подполе поля $\Bbbk$, по сути совпадающее с полем рациональных чисел $\mathbb{Q}$.

Пусть $mn^{-1} = uv^{-1}$ в поле $\Bbbk$. Умножая на nv, получаем равенство $mv = un$. Это то же самое равенство, которое задаёт равенство простых дробей.

Это множество содержит 0 и 1 (это кратные единицы и они представляются в виде $0 \cdot 1^{-1}$ и $1 \cdot 1^{-1}$ соответственно).

Проверка замкнутости относительно сложения и умножения состоит в проверке равенств

$$mn^{-1} + uv^{-1} = (mv + un) \cdot (nv)^{-1}, \qquad mn^{-1} \cdot uv^{-1} = (mu)(nv)^{-1},$$

которая ничем не отличается от проделанной в первой главе проверки для суммы и произведения простых дробей (домножение на nv обеих частей равенства).

Наконец, нужно проверить замкнутость относительно взятия противоположного и обратно, то есть равенства

$$mn^{-1} + (-m)n^{-1} = (mn - mn)n^{-2} = 0 \cdot n^{-2} = 0,$$
$$mn^{-1} \cdot nm^{-1} = (mn)(mn)^{-1} = 1.$$

На этом построение рациональных чисел из элементов поля $\Bbbk$ закончено.

Осталось доказать последнее утверждение леммы, что очень просто. Действительно, поле $\mathbb{F}_p$ попросту совпадает с множеством кратных единицы, которое должно входить в любое подполе. А множество целых чисел, то есть кратных единицы в поле рациональных чисел, порождает всё поле $\mathbb{Q}$ ровно тем способом, который описан выше. □

Возникшая в доказательстве леммы 7.36 величина — порядок группы кратных единицы в поле — очень важна и постоянно нужна в рассуждениях. Однако принято использовать немного другую величину.

Определение 7.37. Если порядок группы кратных единицы в поле $\Bbbk$ равен p, то *характеристика поля* $\operatorname{char}\Bbbk$ равна p.

Если порядок группы кратных единицы в поле $\Bbbk$ бесконечный, то *характеристика поля* $\operatorname{char}\Bbbk$ равна 0.

Из доказательства леммы 7.36 следует, что если характеристика поля ненулевая, то это простое число.

В доказательстве леммы 7.36 мы фактически повторили построение поля рациональных чисел за той лишь разницей, что материалом служили не целые числа, а соответствующие им элементы поля $\Bbbk$ (кратные единицы). Та же самая конструкция обобщается на более широкий случай областей целостности.

Пусть R — область целостности, то есть коммутативное кольцо с единицей и без делителей нуля. Построим по ней *поле частных* F. Для этого на парах (x, y), $x, y \in$

R, $y \neq 0$, введём отношение $(x, y) \sim (u, v)$. Оно выполняется тогда и только тогда, когда $xv = uy$ в кольце R.

Точно так же, как для целых чисел, для любой области целостности доказывается, что это отношение эквивалентности. Поле частных как раз и состоит из классов эквивалентности этого отношения, операции на которых определяются аналогично рациональным числам. Доказательства корректности определения операций и свойства поля в точности такие же, как для рациональных чисел: внимательно посмотрев на них, можно убедиться, что в тех доказательствах используются только свойства области целостности.

Далее у нас появятся многие другие примеры полей. А пока приведём пример тела, которое не является полем.

Пример 7.38 (кватернионы)**.** Определим *алгебру кватернионов* $\mathbb{H}$ аналогично алгебраическому определению комплексных чисел. Кватернионы — это множество

$$\mathbb{H} = \{a_0 + a_1 i + a_2 j + a_3 k : a_i \in \mathbb{R}\};$$

сложение кватернионов покомпонентное

$$(a_0+a_1i+a_2j+a_3k)+(b_0+b_1i+b_2j+b_3k) = (a_0+b_0)+(a_1+b_1)i+(a_2+b_2)j+(a_3+b_3)k;$$

умножение на множестве «мнимых единиц» $\{i, j, k\}$ задаётся таблицей

	i	j	k
i	-1	k	$-j$
j	$-k$	-1	i
k	j	$-i$	-1

и продолжается на все кватернионы, используя условие дистрибутивности:

$$\begin{aligned}(a_0 + a_1 i + a_2 j + a_3 k) \cdot (b_0 + b_1 i + b_2 j + b_3 k) = \\ =(a_0b_0 - a_1b_1 - a_2b_2 - a_3b_3)+ \\ +(a_0b_1 + a_1b_0 + a_2b_3 - a_3b_2)i+ \\ +(a_0b_2 + a_2b_0 - a_1b_3 + a_3b_1)j+ \\ +(a_0b_3 + a_2b_0 + a_1b_2 - a_2b_1)k.\end{aligned} \tag{7.2}$$

Из определения сложения кватернионов ясно, что по сложению они образуют группу, изоморфную группе $\mathbb{R}^4$.

Определение умножения кватернионов выглядит замысловато. На первый взгляд неочевидна даже ассоциативность такого умножения. Смысл такого умножения, как и в случае комплексных чисел, геометрический.

Будем записывать кватернион как «сумму» $a + \boldsymbol{v}$ действительного числа и вектора в трёхмерном евклидовом пространстве с ортонормированным базисом i, j, k. Тогда формула (7.2) умножения кватернионов выражается через скалярное $(\boldsymbol{u}, \boldsymbol{v})$ и векторное $[\boldsymbol{u}, \boldsymbol{v}]$ произведения трёхмерных векторов:

$$(a + \boldsymbol{u}) \cdot (b + \boldsymbol{v}) = \big(ab - (\boldsymbol{u}, \boldsymbol{v})\big) + a\boldsymbol{v} + b\boldsymbol{u} + [\boldsymbol{u}, \boldsymbol{v}]$$

(проверьте!). Такая вид формулы умножения упрощает вычисления с кватернионами. В частности, дистрибутивность умножения кватернионов теперь легко следует из линейности скалярного и векторного произведения по каждому из аргументов.

Легко также увидеть, что произведение *чисто действительного* кватерниона $a = a + \mathbf{0}$ на любой кватернион состоит в умножении всех координат кватерниона на a:

$$(a + \mathbf{0}) \cdot (b + \boldsymbol{v}) = \big(ab - (\mathbf{0}, \boldsymbol{v})\big) + a\boldsymbol{v} + b\mathbf{0} + [\mathbf{0}, \boldsymbol{v}] = ab + a\boldsymbol{v},$$

хотя это и из формулы (7.2) нетрудно увидеть.

Это означает, что $1 = 1 + \mathbf{0}$ является единицей среди кватернионов.

Ещё заметим, что умножение кватерниона $h = a+\boldsymbol{v}$ на *сопряжённый кватернион* $\bar{h} = a - \boldsymbol{v}$ является неотрицательным действительным числом:

$$h \cdot \bar{h} = (a + \boldsymbol{v}) \cdot (a - \boldsymbol{v}) = a^2 - (\boldsymbol{v}, -\boldsymbol{v}) + [\boldsymbol{v}, -\boldsymbol{v}] = a^2 + (\boldsymbol{v}, \boldsymbol{v}) \geqslant 0,$$

так как векторное произведение двух коллинеарных векторов равно 0.

Контрольный вопрос 7.39. Проверьте, что $h \cdot \bar{h} = \bar{h} \cdot h$.

Нормой кватерниона h назовём корень квадратный из $h \cdot \bar{h}$, это аналог модуля комплексного числа. Равенства

$$h \cdot \frac{\bar{h}}{h \cdot \bar{h}} = 1 = \frac{\bar{h}}{h \cdot \bar{h}} \cdot h$$

выполняются для любого ненулевого кватерниона h и показывают, что каждый ненулевой кватернион обратим.

Из всех свойств тела нам осталось проверить ассоциативность умножения. Разумеется, можно написать длинную формулу, раскрывая в двух произведениях трёх кватернионов скобки и приводя подобные.

Вместо этого мы укажем на другой способ проверки ассоциативности.

Кватернион h нормы r представляется в виде $r \cdot \big(\cos(\theta/2) + \sin(\theta/2)\boldsymbol{v}\big)$, где $\boldsymbol{v}$ — вектор единичной длины, $r \geqslant 0$ — чисто действительный кватернион. Странный вид коэффициента, выраженного через косинус половинного угла, объясняется тем, что единичному кватерниону нужно сопоставить поворот трёхмерного пространства вокруг оси, задаваемой вектором v, на угол θ.

Тогда выполняется замечательное утверждение («геометрический смысл умножения кватернионов»): при умножении кватернионов их нормы перемножаются, а с поворотами выполняется композиция. Из этого утверждения ассоциативность умножения становится очевидной.

Однако само по себе утверждение совсем неочевидно. Мы оставляем его для самостоятельного доказательства вдумчивым читателем. □

7.3 Кольца функций и прямые суммы колец

По двум кольцам R_1, R_2 можно определить их *прямую сумму* $R_1 \oplus R_2$ аналогично прямому произведению групп.

Кольцо $R_1 \oplus R_2$ состоит из всевозможных пар (r_1, r_2), где $r_1 \in R_1$, $r_2 \in R_2$. Операции определяются так:

$$(r_1, r_2) + (r_1', r_2') = (r_1 + r_1', r_2 + r_2'),$$
$$(r_1, r_2) \cdot (r_1', r_2') = (r_1 \cdot r_1', r_2 \cdot r_2')$$

(для простоты обозначений мы используем одни и те же символы для операций в обоих кольцах).

Проверка аксиом кольца для прямой суммы колец выполняется механически. По сложению это группа, которая есть прямое произведение аддитивных групп колец R_1 и R_2. Ассоциативность умножения и дистрибутивность проверяются покомпонентно.

В прямой сумме колец, в каждом из которых есть хотя бы два элемента, всегда есть делители нуля:

$$(a, 0) \cdot (0, b) = (a \cdot 0, 0 \cdot b) = (0, 0)$$

(так как в любом кольце $x \cdot 0 = 0 \cdot x = 0$, утверждение 7.13).

Пример 7.40. С другой стороны, если в кольцах R_1, R_2 нет нильпотентных элементов, то их нет и в прямой сумме $R_1 \oplus R_2$. Равенство $(a, b)^n = 0$ равносильно тому, что $a^n = 0$ и $b^n = 0$. □

Контрольный вопрос 7.41. Проверьте, что если в кольце R_1 есть нильпотентные элементы, то в прямой сумме $R_1 \oplus R_2$ этого кольца с любым другим кольцом также есть нильпотентные элементы.

Аналогично определяется прямая сумма нескольких колец.

Ещё один пример колец, близкий к прямым суммам колец, предоставляют *кольца функций*.

Пусть R — кольцо, а X — множество. Тогда кольцо функций R^X на X со значениями в R состоит из всех всюду определённых функций $f\colon X \to R$ с операциями поточечного сложения и умножения: $(f + g)(x) = f(x) + g(x)$, $(f \cdot g)(x) = f(x)g(x)$.

Проверка аксиом кольца для кольца функций быстро сводится к аксиомам кольца R для каждой точки $x \in X$ по отдельности.

Формально кольцо функций R^X можно рассматривать как прямую сумму экземпляров кольца R, индексированных элементами множества X.

Часто ограничение классом функций с достаточно хорошими свойствами даёт подкольцо общего кольца функций на множестве.

Пример 7.42. Множество непрерывных функций на отрезке $[0, 1]$ со значениями в действительных числа и обычными операциями поточечного сложения и умножения функций является кольцом, которое обычно обозначается $C[0, 1]$. Это легко следует из обычных теорем анализа, которые утверждают непрерывность суммы и произведения непрерывных функций. □

Аналогично этому примеру можно рассмотреть кольца дифференцируемых, бесконечно дифференцируемых или аналитических функций.

В общем кольце функций на множестве, содержащем хотя бы два элемента, всегда есть делители нуля.

Пример 7.43. Назовём *дельта-функцией* $\delta_a\colon X \to X$, $a \in X$, функцию, которая равна 1 в точке a и равна 0 в остальных точках. Ясно, что $\delta_a \cdot \delta_b = 0$ при $a \neq b$. □

Пример 7.44. Есть делители нуля и в кольце непрерывных функций, их нетрудно найти.

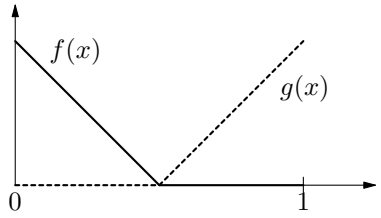


Рис. 28:

Рассмотрим две функции f, g, графики которых изображены на рис. 28. Очевидно, что произведение этих функций тождественно равно нулю, а сами функции нет. □

Замечание 7.45. Более сложно найти делители нуля в кольце бесконечно дифференцируемых функций. Однако такой пример хорошо известен и приводится в учебниках анализа.

А вот в кольце аналитических функций делителей нуля нет. Это также хорошо известный в анализе факт.

7.4 Кольца многочленов от одной переменной

Перефразируя крылатую фразу из известного сериала, «многочлены — это не то, чем они кажутся».

В школьной математике многочлен (одной переменной) определяют как сумму одночленов, которые являются произведением числа (коэффициента) и степени переменной. При таком определении многочлены являются разновидностью формул, что скрывает многие их важные свойства.

В анализе под многочленом обычно понимают функцию действительного переменного, представимую многочленом. Такое определение достаточно для анализа, но, как будет видно далее, совершенно не годится для алгебры, когда речь заходит о конечных полях.

В основу алгебраического определения многочленов берётся тот факт, что многочлен однозначно задаётся своими коэффициентами. Приведём точное формальное определение.

Определение 7.46. Пусть R — коммутативное кольцо R с единицей. *Кольцо многочленов* $R[x]$ с коэффициентами в кольце R состоит из всех *финитных* бесконечных

последовательностей $(f_0, f_1, \dots, f_n, \dots)$ элементов кольца. Слово «финитная» означает, что все элементы последовательности, начиная с некоторого места, равны 0.

Для многочлена f элемент последовательности f_n, стоящий на n-м месте, называется *коэффициентом многочлена* (при степени n, или просто n-м коэффициентом). Коэффициент f_0 по традиции называется свободным членом.

Если последовательность коэффициентов ненулевая, то наибольшее число d, для которого $f_d \neq 0$, называется *степенью* многочлена и обозначается $\deg f$.

Степень нулевого многочлена не определена.[5)]

Операции с многочленами определяются формулами, вычисляющими коэффициенты суммы и произведения многочленов в кольце R:

$$(f+g)_n = f_n + g_n, \tag{7.3}$$

$$(f\cdot g)_n = \sum_{i=0}^{n} f_i g_{n-i} = \sum_{i+j=n} f_i g_j. \tag{7.4}$$

Многочлены степени 0 называются *константами*. То есть, константа — это последовательность $(a, 0, \dots, 0, \dots)$, $a \in R$. Обычно константы обозначаются так же, как элементы кольца R. Как нетрудно видеть, константы перемножаются так же, как элементы кольца R: сложение констант $a + b$ даёт финитную последовательность $(a+b, 0, \dots, 0, \dots)$, а при умножении на константу сумма в правой части (7.4) содержит ровно одно ненулевое слагаемое.

Проверим свойства кольца для многочленов.

Утверждение 7.47. *Описанное в определении* 7.46 *множество $R[x]$ с операциями сложения и умножения является коммутативным кольцом с единицей.*

Доказательство. Из формул (7.3) и (7.4) очевидно, что константы 0 и 1 будут соответственно нулём и единицей в $R[x]$.

Из определения сложения (коэффициенты складываются по отдельности при разных степенях) видно, что коммутативность и ассоциативность сложения следуют из коммутативности и ассоциативности сложения в R. Противоположный многочлен (обратный относительно сложения) получается переменой знаков у всех коэффициентов. Поэтому по сложению многочлены образуют группу.

Коммутативность умножения очевидна из симметрии формулы (7.4), которой определяется умножение.

Дистрибутивность также легко проверяется по формулам для умножения и сложения:

$$\big(f\cdot(g+h)\big)_n = \sum_{i=0}^{n} f_i (g+h)_{n-i} = \sum_{i=0}^{n} f_i g_{n-i} + \sum_{i=0}^{n} f_i h_{n-i} = (fg)_n + (fh)_n$$

Это и есть дистрибутивность $f(g+h) = fg + fh$: ведь по определению 7.46 многочлены — это последовательности коэффициентов.

[5)] Иногда степень нулевого многочлена полагают равной $-\infty$.

Ассоциативность проверяется несложным, но длинным вычислением

$$\big(f(gh)\big)_n = \sum_{i+j=n} f_i (gh)_j = \sum_{i+j=n} \sum_{k+\ell=j} f_i g_k h_\ell = \sum_{i+k+\ell=n} f_i g_k h_\ell,$$
$$\big((fg)h\big)_n = \sum_{i+j=n} (fg)_i h_j = \sum_{i+j=n} \sum_{k+\ell=i} f_k g_\ell h_j = \sum_{k+\ell+j=n} f_k g_\ell h_j.$$

Правые части в этих равенствах различаются только именами индексов суммирования, что не влияет на результат суммирования. Поэтому $\big(f(gh)\big)_n = \big((fg)h\big)_n$, то есть умножение многочленов ассоциативно. □

Теперь посмотрим на связь многочленов с функциями.

Определение 7.48. Пусть даны многочлен $f \in R[x]$ и элемент a кольца R. *Значением* многочлена f в точке a называется элемент кольца

$$f_0 + f_1 a + f_2 a^2 + \cdots + f_d a^d = \sum_n f_n a^n,$$

где $d = \deg f$.

Каждый элемент a кольца R задаёт отображение $\mathrm{Ev}_a \colon R[x] \to R$, которое сопоставляет многочлену его значение в точке a.

Многочлен $f \in R[x]$ задаёт функцию $x \mapsto \mathrm{Ev}_x(f)$. Эта функция обычно обозначается $f(x)$. Объединяя эти функции вместе, получаем отображение кольца многочленов в кольцо функций на R. Образ этого отображения состоит из функций, *представимых многочленами* (или, как говорят, полиномиальных функций).

Операции с многочленами и значениями многочленов согласованы.

Лемма 7.49. $(f+g)(a) = f(a) + g(a)$; $(f \cdot g)(a) = f(a)g(a)$.

Доказательство. Прямые вычисления, использующие дистрибутивность и прочие свойства кольца R, показывают

$$\sum_n (f+g)_n a^n = \sum_n \big(f_n a^n + g_n a^n\big) = \sum_n f_n a^n + \sum_n g_n a^n,$$
$$\sum_n (f \cdot g)_n a^n = \sum_n \sum_{i+j=n} f_i a^i g_j a^j = \left(\sum_n f_n a^n\right) \cdot \left(\sum_n g_n a^n\right),$$

что и требовалось доказать. □

Зачастую многочлены по традиции записываются в виде

$$f(x) = f_0 + f_1 x + f_2 x^2 + \cdots + f_d x^d,$$

который подчёркивает связь многочленов и функций. Мы тоже будем использовать эту традиционную запись.

Однако важно помнить, что многочлены равны тогда и только тогда, когда их коэффициенты равны; а функции равны тогда и только тогда, когда их значения в каждой точке совпадают. Из-за этого многочлены отличаются от функций, представимых многочленами.

Пример 7.50. В поле $\mathbb{Z}/(2)$ выполняется равенство $x = x^2$ для любого $x \in \mathbb{Z}/(2)$. Поэтому два разных многочлена $(0, 1, 0, \dots)$ и $(0, 0, 1, 0 \dots)$ задают одну и ту же функцию $\mathbb{Z}/(2) \to \mathbb{Z}/(2)$, а именно, тождественное отображение. □

Пример 7.51. Более общим образом, если R — конечное кольцо, то множество функций из R в R конечно. Но если в кольце R есть хотя бы два элемента, кольцо $R[x]$ уже бесконечно. Поэтому какие-то различные многочлены будут задавать одну и ту же функцию. □

Замечание 7.52. Если F — бесконечное поле (например, рациональные, действительные или комплексные числа), то разницы между многочленами из $F[x]$ и функциями $F \to F$, которые представимы многочленами, нет. Мы подробно обсудим этот вопрос позже.

Обсудим некоторые важные свойства многочленов.

Теорема 7.53 (логарифмическое свойство степени)**.** *Если в R нет делителей нуля, то для $f, g \neq 0$ выполнено* $\deg(fg) = \deg f + \deg g$.

Доказательство. Рассмотрим ненулевые многочлены f, g степеней r и q соответственно.

По формуле (7.4) для умножения многочленов

$$(fg)_n = \sum_{i+j=n} f_i g_j.$$

При $n > r + q$ каждое слагаемое в этой сумме равно 0: если $i \leqslant r$, то $j \geqslant n - i > > (r - i) + q$, то есть $g_j = 0$; если $i > r$, то $f_i = 0$. Поэтому $\deg(fg) \leqslant \deg f + \deg g$.

При $n = r + q$ аналогичное рассуждение показывает, что есть ровно одно слагаемое, которое может быть отлично от 0; это $f_r g_q$ (в остальных один из индексов больше степени соответствующего многочлена). Если в R нет делителей нуля, то $f_r \neq 0$ и $g_q \neq 0$ влечёт $f_r g_q \neq 0$. Поэтому $\deg(fg) = \deg f + \deg g$. □

Следствие 7.54. *Если R — область целостности, то и $R[x]$ — область целостности.*

Контрольный вопрос 7.55. Докажите, что если в R есть делители нуля, то в $R[x]$ также есть делители нуля.

Следствие 7.56. *Обратимые элементы в кольце многочленов $R[x]$ над областью целостности R — это константы* (*многочлены степени* 0), *обратимые в кольце R.*

Доказательство. Обратимость константы, обратимой в R, очевидна, так как операции с константами точно такие, как в кольце R.

В силу логарифмического свойства степени умножение на многочлен положительной степени увеличивает степень многочлена (а степень 1 равна 0). Это показывает необратимость всех остальных многочленов. □

Для многочленов с коэффициентами в поле, как и для целых чисел, определено деление с остатком.

Лемма 7.57. *Пусть F — поле. Для любых многочленов $f, g \in F[x]$, $g \neq 0$, существуют таких два многочлена q и r, что*

$$f = qg + r, \qquad r = 0 \text{ или } \deg r < \deg g.$$

Как и для целых чисел, q называется неполным частным, а r — остатком. Деление многочленов с остатком выполняется обычным школьным алгоритмом деления в столбик для многочленов.

Пример 7.58. Разделим многочлен $x^3 + x - 1$ на $x^2 - 2$ в столбик:

$$\begin{array}{rrr|l} x^3 & +x & -1 & x^2 \quad -2 \\ \cline{4-4} x^3 & -2x & & x \\ \cline{1-2} & 3x & -1 & \end{array}$$

(Неполное) частное равно x, а остаток равен $3x - 1$. □

Вполне очевидно, что деление в столбик всегда даёт остаток и неполное частное. Однако приведём формальное доказательство леммы. Читателю рекомендуется убедиться, что по сути мы в этом доказательстве как раз и выполняем деление в столбик (не называя его явно, а описывая только один шаг).

Доказательство леммы 7.57. Индукция по степени многочлена f (делимого).

Пусть мы рассматриваем деление с остатком на многочлен

$$g(x) = g_0 + g_1 x + g_2 x^2 + \cdots + g_d x^d, \ d = \deg g.$$

Для многочленов f степени меньше d имеем представление $f = 0 \cdot g + f$, $\deg f < \deg g$. Это база индукции.

Теперь предположим, что мы умеем делить с остатком на g многочлены степени меньше n. Рассмотрим многочлен степени n:

$$f(x) = f_0 + f_1 x + \cdots + f_n x^n, \ f_n \neq 0.$$

Многочлен $\tilde{f} = f - f_n g_d^{-1} x^{n-d} g$ имеет степень, меньшую n, поэтому его можно разделить с остатком на g:

$$\tilde{f} = hg + r, \ \deg r < \deg g.$$

Но тогда

$$f = \tilde{f} + f_n g_d^{-1} x^{n-d} g = (h + f_n g_d^{-1} x^{n-d}) g + r.$$

Значит, f также можно разделить на g с остатком. □

Замечание 7.59. Из доказательства ясно, почему нам нужно поле: чтобы подобрать очередной коэффициент неполного частного, нужно делить коэффициент делимого на старший коэффициент делителя.

Если старший коэффициент делителя равен 1, то деление с остатком на такой многочлен возможно и в том случае, когда кольцо коэффициентов не является полем.

7.5 Лемма о корнях многочлена и её следствия

В этом разделе мы рассматриваем кольцо многочленов с коэффициентами в поле F. В этом случае многочлены можно делить с остатком, что приводит к замечательным следствиям в теории чисел, алгебре и теоретической информатике.

Начнём с хорошо известного в школьной математике факта.

Утверждение 7.60. *При делении многочлена f на многочлен $x-a$ получается остаток $f(a)$.*

Доказательство. Поскольку степень многочлена $x - a$ равна 1, степень остатка 0 (или он нулевой). Таким образом, остаток r — это константа, для которой выполняется равенство

$$f(x) = q(x)(x - a) + r$$

Вычисляя значения многочленов в левой и правой частях равенства в точке a, получаем $f(a) = q(a)(a - a) + r = r$, что и требовалось. □

Пример 7.61. Найдём остаток при делении $x^{100} + x - 2$ на $x - 1$ в кольце $\mathbb{R}[x]$.

По предыдущему утверждению он равен $1^{100} + 1 - 2 = 0$. Таким образом, многочлен $x^{100} + x - 2$ делится на $x - 1$ в кольце $\mathbb{R}[x]$. □

Определение 7.62. Элемент $a \in R$ называется корнем многочлена $f \in R[x]$, если $f(a) = 0$.

Из утверждения 7.60 и логарифмического свойства степени получаем верхнюю оценку на количество корней многочлена.

Лемма 7.63 (о числе корней многочлена)**.** *Количество корней ненулевого многочлена $f \in F[x]$ не превосходит его степени, если кольцо коэффициентов F является полем.*

Доказательство. Индукция по степени многочлена. База индукции: многочлены степени 1. Тут всё просто. Уравнение $ax + b = 0$ имеет в поле единственное решение при $a \neq 0$ и не имеет решений при $a = 0$, $b \neq 0$ (случай $a = b = 0$ означает, что многочлен нулевой).

Индуктивный переход. Предположим, что многочлены степени $d < n$ имеют не более d корней. Рассмотрим многочлен f степени n. Пусть a — его корень. Тогда $f(x) = q(x)(x - a) + f(a) = q(x)(x - a)$. Степень многочлена $q(x)$ равна $n - 1$, по индуктивному предположению у него не больше $n - 1$ корней.

С другой стороны, если b — какой-то корень многочлена f, то $q(b)(b-a) = f(b) = 0$. Так как в поле нет делителей нуля, это означает, что $q(b) = 0$ или $b = a$. Таким образом, у многочлена f не более $(n - 1) + 1 = n$ корней.

Индуктивный переход доказан, лемма справедлива по принципу математической индукции. □

Замечание 7.64. Для произвольного кольца коэффициентов лемма неверна. Мы уже видели в примере 2.63, что у сравнения $x^2 \equiv 1 \pmod 8$ есть четыре решения. Другими словами это означает, что многочлен $x^2 - 1$ имеет 4 корня в кольце $\mathbb{Z}/(8)$, что больше его степени.

Однако условие, что кольцо коэффициентов является полем, слишком сильное. Достаточно потребовать, чтобы кольцо коэффициентов R было целостным. Обобщить лемму о числе корней многочлена на случай целостного кольца можно двумя способами.

Во-первых, можно заметить, что на многочлен $x - a$ можно делить с остатком в любом кольце многочленов, поскольку старший коэффициент равен 1 и заведомо обратим. Утверждение 7.60 также справедливо для любого кольца коэффициентов. Осталось заметить, что в доказательстве леммы 7.63 мы использовали только утверждение 7.60 и отсутствие делителей нуля.

Второй способ более важен для дальнейшего. Как объяснялось выше, по любой области целостности R можно построить поле частных F. Так как $R \subseteq F$, то и $R[x] \subseteq F[x]$. Если $a \in R$ является корнем многочлена $f \in R[x]$, то он же является и корнем многочлена f как элемента кольца многочленов с коэффициентами в поле частных. Осталось применить лемму о числе корней многочлена к этому случаю.

Мы часто будем рассматривать значения многочленов с коэффициентами в некотором кольце (на самом деле, обычно это будет поле) в точках, которые лежат в более широком кольце. Приведённое выше рассуждение — один из простых примеров, когда такие значения оказываются полезными.

7.5.1 Критерий квадратичного вычета

У леммы о числе корней многочлена есть много интересных следствий. Мы уже рассматривали квадратичные вычеты и невычеты в примере 4.1.1. Сейчас можно переформулировать определение квадратичного вычета так: ненулевой элемент a поля $\mathbb{Z}/(p)$ называется квадратичным вычетом, если многочлен $x^2 - a$ имеет корень в этом поле (в противном случае a называется квадратичным невычетом). В примере 4.1.1 мы выяснили, что квадратичных вычетов и невычетов одинаковое количество и нашли необходимое условие квадратичного невычета:

$$a^{(p-1)/2} \equiv 1 \pmod p.$$

Это условие является и достаточным. Для доказательства потребуется как раз лемма о числе корней многочлена.

Теорема 7.65 (критерий квадратичного вычета)**.** *Элемент $a \in \mathbb{Z}/(p)$ является квадратичным вычетов тогда и только тогда, когда $a^{(p-1)/2} \equiv 1 \pmod p$.*

Доказательство. Необходимость доказана в примере 4.1.1. Достаточность следует из того, что у многочлена $x^{(p-1)/2} - 1$ в поле $\mathbb{Z}/(p)$ не больше $(p-1)/2$ корней (лемма о числе корней). Но все квадратичные вычеты являются корнями этого уравнения, а их как раз $(p-1)/2$ штук. Значит, других корней нет. □

Поскольку мы знаем, что различных корней многочлена $x^{(p-1)/2}-1$ в поле $\mathbb{Z}/(p)$ ровно $(p-1)/2$ штук, то выполняется разложение

$$x^{(p-1)/2}-1=\prod_{a\in Q_p}(x-a),\quad Q_p\text{ — множество квадратичных вычетов,}\tag{7.5}$$

которое можно доказать по индукции аналогично рассуждению в доказательстве леммы о числе корней.

Так как $a^{(p-1)/2}=\pm 1$ в $\mathbb{Z}/(p)$ (мы это уже выясняли, советуем читателю вспомнить доказательство), то все квадратичные невычеты являются корнями многочлена $x^{(p-1)/2}+1$ и для этого многочлена выполняется разложение

$$x^{(p-1)/2}+1=\prod_{a\in N_p}(x-a),\quad N_p\text{ — множество квадратичных невычетов.}\tag{7.6}$$

Такие разложения позволяют применить теорему Виета, которая выражает коэффициенты многочлена $f(x)=\prod_{a\in S}(x-a)$ через его корни. Обозначим $d=|S|$ степень многочлена. Раскрывая скобки и приводя подобные, получаем равенства

$$\begin{aligned}
f_0&=(-1)^d\prod_{a\in S}a=s_d(\{a:a\in S\}),\\
f_1&=(-1)^{d-1}\sum_{\substack{P\subseteq S\\|P|=d-1}}\prod_{a\in S}a=s_{d-1}(\{a:a\in S\},\\
&\dots,\\
f_{d-1}&=(-1)^1\sum_{a\in S}a=s_1(\{a:a\in S\}),\\
f_d&=1=s_0(\{a:a\in S\}).
\end{aligned}$$

Здесь $s_i(\{a:a\in S\})$ обозначает i-ю элементарную симметрическую функцию от элементов множества $\{a:a\in S\}$: сумму всех возможных произведений i элементов из множества S.

Используя эту теорему, можно вычислять разные суммы и произведения от квадратичных вычетов и невычетов.

Пример 7.66. Чему равна сумма квадратичных вычетов по модулю 59? Из разложения (7.5) и теоремы Виета видим, что это коэффициент многочлена $x^{29}-1$ при степени 28, то есть 0.

Чему равно произведение квадратичных невычетов по модулю 97? Из разложения (7.6) и теоремы Виета видим, что это свободный член многочлена $x^{48}+1$ при степени 28, то есть это произведение равно 1. □

Замечание 7.67. Мы не будем останавливаться на этом подробно, но заметим, что любая симметрическая функция выражается через элементарные симметрические. Так что можно находить значения элементарных симметрических функций от квадратичных вычетов и невычетов, используя разложения (7.5) и (7.6).

Например, из тождеств Ньютона (см., скажем, [22]) можно заключить, что сумма i-х степеней квадратичных вычетов (как и невычетов) равна 0 при $i<(p-1)/2$.

Контрольный вопрос 7.68. Найдите сумму $(p-1)/2$-х степеней квадратичных вычетов по модулю p.

7.5.2 Мультипликативная группа конечного поля

Использование леммы о числе корней позволяет полностью прояснить структуру мультипликативной группы конечного поля (ненулевые элементы поля с операцией умножения).

Теорема 7.69. *Мультипликативная группа F^* любого конечного поля F циклическая.*

Доказательство. Лемма о числе корней говорит, что количество корней у многочлена $x^d - 1$ в поле F не больше d. Отсюда следует, что если в мультипликативной группе F^* есть хотя бы один элемент порядка d, то элементов порядка d ровно $\varphi(d)$, где $\varphi(\cdot)$ — функция Эйлера.

Действительно, пусть $\operatorname{ord} a = d$, тогда циклическая подгруппа $\langle a \rangle$ имеет порождающую a и порядок d. В этой группе $\varphi(d)$ элементов порядка d. Но все элементы этой подгруппы являются корнями уравнения $x^d - 1$ (теорема Лагранжа). Значит, других элементов порядка d, кроме подсчитанных $\varphi(d)$ штук из этой подгруппы, в мультипликативной группе поля нет.

Теперь вспомним формулу суммирования Эйлера

$$n = \sum_{d|n} \varphi(n)$$

(следствие 3.27).

Обозначим $n = |F| - 1 = |F^*|$, а через D обозначим множество возможных порядков элементов мультипликативной группы F^*. Подсчитаем количество элементов в мультипликативной группе, группируя их по возможным значениям порядкам элемента и пользуясь найденным выражением для количества элементов порядка d. Получаем равенство

$$\sum_{d \in D} \varphi(d) = n = \sum_{d|n} \varphi(d).$$

Заметим, что левая сумма целиком содержится в правой: по теореме Лагранжа, если $d \in D$ порядок элемента мультипликативной группы, то он делит порядок n самой группы. Все слагаемые неотрицательные. Поэтому равенство возможно лишь в том случае, когда в левую часть входят все возможные делители n, включая само число n. Но это означает, что в группе F^* есть элемент порядка n, то есть она циклическая. □

Пока мы знаем лишь один пример конечных полей — вычеты по модулю простого числа. Уже в этом случае теорема 7.69 нетривиальна. Она говорит, что для любого простого p найдётся такой вычет g (по традиции он называется *первообразным корнем* по модулю p), степени которого дают все ненулевые вычеты по модулю p.

Для небольших значений p первообразный корень можно найти перебором. Тут полезно следующее наблюдение.

Утверждение 7.70. *Порядок элемента, который не порождает циклическую группу порядка n, обязательно делит одно из чисел n/p_1, n/p_2, ..., n/p_t, где p_i — все простые делители n.*

Доказательство. Пусть порядок $a \in C_n$ равен d. Поскольку d делит n по теореме Лагранжа, любое простое число входит в разложение d по степеням простых с показателем не бóльшим, чем в разложении n. А если $d < n$, то найдётся такой простой делитель q порядка группы n, который в разложение d входит с меньшим показателем, чем в разложении n. Тогда $d \mid n/q$. □

Пример 7.71. Найдём первообразный корень по модулю 13. Это такой вычет g, что $g^{12} \equiv 1 \pmod{13}$, но $g^k \not\equiv 1 \pmod{13}$, $1 \leqslant k < 12$.

Попробуем вычет 2. Проверять все степени 2 не обязательно. Простые делители 12 это 2 и 3. Поэтому достаточно вычислить $2^{12/2} = 2^6 = 64 \equiv -1 \pmod{13}$ и $2^{12/3} = 2^4 = 16 \equiv 3 \pmod{13}$. Если бы порядок 2 был меньше 12, то по утверждению 7.70 одна из этих степеней равнялась бы 1 по модулю 13.

Значит, 2 — первообразный корень по модулю 13. □

Пример 7.72. Найдём первообразный корень по модулю 17. Здесь ещё проще: порядок мультипликативной группы по модулю 17 равен 16, это степень 2. Поэтому достаточно искать такой вычет g, для которого $g^8 \not\equiv 1 \pmod{17}$.

Поскольку $2^8 = 16^2 \equiv (-1)^2 = 1 \pmod{17}$, вычет 2 не является первообразным корнем по модулю 17.

А вот вычет 3 является первообразным корнем по модулю 17, как показывает вычисление $3^8 = 81^2 \equiv (-4)^2 = 16 \equiv -1 \pmod{17}$. □

7.5.3 Мультипликативные группы вычетов целых чисел

Для каких ещё модулей n мультипликативная группа вычетов Z_n^* является циклической? Китайская теорема 3.58 говорит, что при взаимно простых p, q имеет место изоморфизм $Z_{pq}^* \cong Z_p^* \times Z_q^*$. Это позволяет исключить многие значения n.

Для нечётного простого p все числа $\varphi(p^t) = p^t - p^{t-1}$ чётны. Если число n имеет два разных нечётных простых делителя p и q, то его можно разложить на два взаимно простых множителя $n = mq^s$, причём $p \mid m$. Из формулы для функции Эйлера (лемма 3.54) заключаем, что $\varphi(m) = |Z_m^*|$ и $\varphi(q^s) = |Z_{q^s}^*|$ имеют общий делитель 2. Поэтому группа $Z_m^* \oplus Z_{q^s}^*$ не циклическая (мы доказали это в примере 5.29).

Для числа $n = 2^t p^s$, где $t \geqslant 2$, а p — нечётное простое, справедливо аналогичное рассуждение, так как $\varphi(2^t) = |Z_{2^t}^*| = 2^t - 2^{t-1}$ чётное и $\varphi(p^s)$ чётное.

Остаются неразобранными случаи $n = 2^t$, $t > 1$, и $n = 2p^t$ и $n = p^s$, где p — нечётное простое. Случай $n = 4$ очевиден: $\varphi(4) = 2$, группа порядка 2 единственная с точностью до изоморфизма и она циклическая. Если $n > 2$, то мы уже знаем из примера 2.71, что решений сравнения $x^2 \equiv 1 \pmod{2^n}$ четыре. Поэтому в

мультипликативной группе $Z^*_{2^n}$ элементов порядка 2 три (ещё одно решение сравнения — единица, которая имеет порядок 1). Но в циклической группе это невозможно: элементов порядка d ровно $\varphi(d)$ штук. Поэтому $Z^*_{2^n}$, $n \geqslant 3$, не является циклической.

В последних двух случаях мультипликативная группа Z^*_n циклическая. Заметим, что по китайской теореме $Z^*_{2p^n} \cong Z^*_2 \times Z^*_{p^n} \cong Z^*_{p^n}$, так как $|Z^*_2| = 1$. Поэтому осталось доказать ровно одну теорему.

Теорема 7.73. *Если p — нечётное простое число, то мультипликативные группы по модулю p^n циклические для любого n.*

Доказательство. Случай $n = 1$ покрывается теоремой 7.69. Так что далее рассматриваем случай модуля $q = p^n$, $n > 1$.

Докажем, что любой вычет из группы $Z^*_{p^n}$ представляется в виде

$$a(1 + pb), \quad 1 \leqslant a < p,\ 0 \leqslant b < p^{n-1},$$

Разлагая в p-ичной системе счисления обратимый остаток r по модулю p^n, получаем $r = a + pB$. Пусть $\bar{a}$ — обратный к a остаток по модулю p^n (он существует, так как a взаимно просто с p^n). Тогда

$$r = a + pB \equiv a\bar{a}(a + pB) = a(a\bar{a} + p\bar{a}B) \equiv a(1 + pb) \pmod{p^n},$$

что и требовалось.

Выделим в группе Z^*_q подгруппу H, состоящую из вычетов вида $1 + px$. Это подгруппа, так как это подмножество конечной группы, замкнутое относительно умножения. Порядок группы H равен p^{n-1}, как нетрудно видеть: если $p^{n-1} \mid (x' - x'')$, то $1 + px' \equiv 1 + px'' \pmod{p^n}$.

Докажем, что группа H циклическая и порождается вычетом $1 + p$. Поскольку порядок группы — степень простого, достаточно проверить, что

$$(1 + p)^{p^{n-2}} \not\equiv 1 \pmod{p^n}.$$

Проверим, что в разложении бинома

$$(1 + p)^{p^{n-2}} = 1 + p^{n-2} \cdot p + \binom{p^{n-2}}{2} p^2 + \dots$$

все слагаемые, начиная с третьего, делятся на p^n. Для третьего и четвёртого проверка прямолинейная:

$$\binom{p^{n-2}}{2} p^2 = p^n \cdot \frac{p^{n-2} - 1}{2},$$

причём число $(p^{n-2} - 1)/2$ целое, так как p нечётное;

$$\binom{p^{n-2}}{3} p^3 = p^n \cdot \frac{p(p^{n-2} - 1)(p^{n-2} - 2)}{6},$$

и множитель в правой части при p^n целый, что проверяется разбором случаев с учётом того, что $(p^{n-2} - 1)(p^{n-2} - 2)$ чётное: если $p = 3$, то $(p^{n-2} - 1)(p^{n-2} - 2)/2$ целое и $p(p^{n-2} - 1)(p^{n-2} - 2)/6$ целое; если $p \equiv 1 \pmod 3$, то $(p^{n-2} - 1)$ не только чётное, но и делится на 3, поэтому $(p^{n-2} - 1)/6$ целое; если $p \equiv -1 \pmod 3$, то на 3

делится либо $(-1)^{n-2}-1$ (при чётном n), либо $(-1)^{n-2}-2\equiv(-1)^{n-2}+1 \pmod 3$ (при нечётном n) и число $p(p^{n-2}-1)(p^{n-2}-2)/6$ целое.

Чтобы доказать утверждение для остальных слагаемых — при $k\geqslant 4$ — проверим, что показатель максимальной степени p, на которую делится $k!$ (знаменатель в формуле для биномиального коэффициента $\binom{p^{n-2}}{k}$), не больше $k-2$. Тогда слагаемое $\binom{p^{n-2}}{k}p^k$ делится на $p^{n-2}\cdot p^2=p^n$.

Показатель максимальной степени p, на которую делится $k!$, вычислен в лемме 2.74. Используя формулу из той леммы, получаем оценку

$$\left\lfloor\frac{k}{p}\right\rfloor+\left\lfloor\frac{k}{p^2}\right\rfloor+\dots+\left\lfloor\frac{k}{p^t}\right\rfloor+\dots\leqslant\frac{k}{p}\cdot\left(1+\frac{1}{p}+\dots\right)=\frac{k}{p(1-1/p)}=\frac{k}{p-1}\leqslant k-2$$

при $p>2$, $k>3$.

Итак,

$$(1+p)^{p^{n-2}}\equiv 1+p\cdot p^{n-2}=1+p^{n-1}\not\equiv 1 \pmod{p^n},$$

откуда следует цикличность группы H.

Порядок H равен p^{n-1}, а индекс $p-1$. Эти числа взаимно просты. Поэтому по лемме 4.34 имеем изоморфизм

$$Z^*_{p^n}\cong H\oplus(Z^*_{p^n}/H).$$

Выясним, как устроена факторгруппа $Z^*_{p^n}/H$. Рассуждаем аналогично доказательству леммы 4.34.

Рассмотрим гомоморфизм возведения в степень p^{n-1}: $\varphi\colon x\mapsto x^{p^{n-1}}$ на группе $Z^*_{p^n}$. Вычет вида $a(1+pb)$ он переводит в $a^{p^{n-1}}$, так как порядок группы H равен p^{n-1}, так что второй множитель при возведении в степень p^{n-1} даёт 1 по модулю p^n.

Образом G группы $Z^*_{p^n}$ при этом гомоморфизме являются вычеты вида $a^{p^{n-1}}$. Легко проверить, что они перемножаются так же, как вычеты по модулю p: если $ab=c+pt$, то

$$a^{p^{n-1}}\cdot b^{p^{n-1}}=(ab)^{p^{n-1}}=(c+pt)^{p^{n-1}}\equiv c^{p^{n-1}} \pmod{p^n},$$

так как $\binom{p^{n-1}}{k}p^k$ делится на p^n (доказательство аналогично предыдущему разбору делимости слагаемых в биноме, только проще, потому что общее рассуждение теперь проходит при $k\geqslant 2$, а для $k=1$ утверждение очевидно). В частности, ни одна такая степень не равна 1 по модулю p^n, за исключением $1^{p^{n-1}}$.

Итак, $G\cong Z^*_p\cong C_{p-1}$ (по теореме 7.69). Но $G\cong Z^*_{p^n}/H$ (мы уже проверяли это в доказательстве леммы 4.34, читателю рекомендуется сейчас вспомнить это доказательство). Таким образом

$$Z^*_{p^n}\cong H\oplus(Z^*_{p^n}/H)\cong C_{p^{n-1}}\oplus C_{p-1}\cong C_{p^n-p^{n-1}}$$

(тут мы ещё раз используем взаимную простоту p^{n-1} и $p-1$). □

7.6 Кольца многочленов от нескольких переменных и их корни

В элементарной математике и в анализе используются не только функции, представимые многочленами от одной переменной, но и аналогичные функции нескольких переменных, представимые *целыми алгебраическими выражениями*: формулами, в которые входят числа и переменные, а разрешённые операции — сложение, вычитание и умножение.

Аналогично кольцам многочленов от одной переменной можно определить и кольца многочленов от нескольких переменных. Проще всего это сделать индуктивно, используя определение кольца многочленов от одной переменной с коэффициентами в произвольном кольце.

Определение 7.74. Кольца многочленов от двух переменных с коэффициентами в кольце R определяется как $R[x_1, x_2] = (R[x_1])[x_2]$.

Аналогично определяются кольца многочленов от бóльшего числа переменных. В общем случае $R[x_1, \ldots, x_n] = (R[x_1, \ldots, x_{n-1}])[x_n]$.

Как и в случае одной переменной, многочлены от нескольких переменных задаются своими коэффициентами. В случае n переменных коэффициенты многочленов индексируются n-мерными векторами с неотрицательными целыми координатами, множество таких векторов обозначаем $\mathbb{N}^n$. Значения коэффициентов лежат в кольце R. Другими словами, коэффициенты — это функции $\mathbb{N}^n \to R$. Будем обозначать коэффициент многочлена f, индексированный вектором $(i_1, \ldots, i_n)$, как $\operatorname{coef}_f(i_1, \ldots, i_n)$.

Определим коэффициенты многочленов от нескольких переменных по индукции. Для многочленов от одной переменной коэффициенты определены (мы ведь и определяли многочлены от одной переменной как последовательности коэффициентов). Определим теперь коэффициенты многочлена $f \in R[x_1, \ldots, x_n]$. По определению кольца многочленов от нескольких переменных,

$$f = \sum_i f_i x_n^i, \qquad f_i \in R[x_1, \ldots, x_{n-1}].$$

Тогда

$$\operatorname{coef}_f(i_1, \ldots, i_{n-1}, i_n) = \operatorname{coef}_{f_{i_n}}(i_1, \ldots, i_{n-1})$$

Контрольный вопрос 7.75. Докажите (индукцией по числу переменных), что у многочлена от нескольких переменных лишь конечное количество коэффициентов отлично от 0.

Степенью ненулевого многочлена называется максимум $i_1 + i_2 + \cdots + i_n$ по всем ненулевым коэффициентам многочлена. Обозначается степень так же, как в и в случае одной переменной: $\deg f$. Из определения коэффициентов многочлена сразу получаем равенство

$$\deg(f(x_1, \ldots, x_{n-1})x_n^k) = k + \deg f(x_1, \ldots, x_{n-1}). \tag{7.7}$$

Многочлены от n переменных задают функции $R^n \to R$ аналогично многочленам от одной переменной. Для этого определим значение многочлена $f \in R[x_1, \ldots, x_n]$ в точке $a = (a_1, \ldots, a_n) \in R^n$:

$$f(a) = \sum_{(i_1,\ldots,i_n)} \operatorname{coef}_f(i_1, \ldots, i_n) a_1^{i_1} a_2^{i_2} \cdot \ldots \cdot a_n^{i_n}.$$

Из определений коэффициентов многочлена и значения многочлена в точке для многочлена

$$f = \sum_i f_i x_n^i$$

следует равенство

$$f(a_1, \ldots, a_n) = \sum_i f_i(a_1, \ldots, a_{n-1}) a_n^i. \tag{7.8}$$

Контрольный вопрос 7.76. Запишите аккуратное доказательство равенства (7.8) индукцией по числу переменных.

Точка a называется *корнем многочлена* f, если $f(a) = 0$. Как и для многочленов от одной переменной, может случиться так, что задаваемая многочленом функция тождественно равна нулю. Однако если кольцо коэффициентов является полем и в этом поле достаточно много элементов, такое невозможно.

Для многочленов от нескольких переменных с коэффициентами в поле есть обобщение леммы 7.63, которое называется *леммой Шварца–Зиппеля*.

Лемма 7.77. *Пусть F — поле, $f \in F[x_1, \ldots, x_n]$ — ненулевой многочлен степени d, а $S \subseteq F$ — подмножество элементов поля F.*

Тогда количество корней $(a_1, \ldots, a_n)$ многочлена f, для которых все $a_i \in S$, не превосходит $d \cdot |S|^{n-1}$.

Доказательство. Индукция по числу переменных. База индукции: лемма 7.63.

Индуктивный переход. Пусть $n \geqslant 2$ и лемма доказана для многочленов с числом переменных $< n$ и

$$f = \sum_i f_i x_n^i$$

ненулевой многочлен. Какие-то из коэффициентов f_i не равны 0. Выберем максимальное i, при котором $f_i \neq 0$, обозначим его k.

Многочлен f_k ненулевой и лежит в кольце $F[x_1, \ldots, x_{n-1}]$. Его степень не больше $d - k$, см. (7.7). По индуктивному предположению у него не больше $(d-k) \cdot |S|^{n-2}$ корней в множестве S^{n-1}. Поэтому корней f вида (a', a''), $f_k(a') = 0$, $a' \in S^{n-1}$, $a'' \in S$ не больше, чем $(d-k) \cdot |S|^{n-2} \cdot |S| = (d-k) \cdot |S|^{n-1}$.

Оценим количество оставшихся корней вида (a', a''), $f_k(a') \neq 0$, $a' \in S^{n-1}$, $a'' \in S$. При фиксированном a' это корни многочлена от одной переменной

$$\sum_{i=0}^{k} f_i(a'_1, \ldots, a'_{n-1}) x^i.$$

Этот многочлен ненулевой, так как $f_k(a'_1, \ldots, a'_{n-1}) \neq 0$, и его степень равна k. Поэтому у него не более k корней. Поэтому всего корней второго типа не более $k|S|^{n-1}$.

Суммируя эти две оценки, приходим к выводу, что всего корней многочлена f не больше, чем $(d-k) \cdot |S|^{n-1} + k|S|^{n-1} = d|S|^{n-1}$. Индуктивное предположение доказано, лемма теперь следует из принципа математической индукции. □

Замечание 7.78. Лемма Шварца–Зиппеля имеет многочисленные приложения в теоретической информатике. Их изложение выходит далеко за рамки нашего рассказа. Опишем неформально одно из самых важных.

Если нужно проверить, равен ли многочлен нулю (в кольце многочленов), то прямолинейный способ требует вычисления всех его коэффициентов, что может оказаться очень долго. Лемма Шварца–Зиппеля показывает, что если допускать вероятностные процедуры и поле достаточно большое (например, бесконечное), то достаточно выбрать одну-единственную случайную точку из достаточно большого множества, в ней вычислить значение многочлена и сравнить это значение с нулём. В случае нулевого многочлена получится, конечно же, 0. А в случае ненулевого с очень большой вероятностью получится не 0.

8 Линейная алгебра над полем

В начальных курсах линейной алгебры обычно рассматриваются векторные пространства над действительными или комплексными числами. Однако многие результаты линейной алгебры справедливы и в том случае, когда вместо действительных чисел берётся произвольное поле. Доказательства этих результатов также переносятся без особых изменений на случай произвольного поля. Для замкнутости изложения мы воспроизводим их в этой главе.

8.1 Базисы и размерность

Определение векторного пространства переносится на случай произвольного поля буквально.

Определение 8.1. *Векторное пространство* над полем F — это множество V (элементы которого называются *векторами*), на котором заданы следующие операции: сложение $+\colon V \times V \to V$ и умножение на «скаляры» (элементы поля) $F \times V \to V$. Свойства этих операций таковы (здесь мы обозначаем элементы поля греческими буквами, а элементы векторного пространства — латинскими, умножение на число обозначаем точкой $\cdot$, хотя почти всегда в дальнейшем точка будет опускаться):

L1: V — абелева группа по сложению, её нейтральный элемент называется нулевым вектором, мы его обозначаем 0, как и нулевой элемент поля (различить эти два нуля всегда можно по контексту);

L2: законы дистрибутивности $\alpha \cdot (v_1 + v_2) = \alpha \cdot v_1 + \alpha \cdot v_2$, $(\alpha_1 + \alpha_2) \cdot v = \alpha_1 \cdot v + \alpha_2 \cdot v$;

L3: $\alpha \cdot (\beta \cdot v) = (\alpha\beta) \cdot v$ (композиция умножений на два числа совпадает с умножением на произведение этих чисел);

L4: $1 \cdot v = v$.

Приведём сразу один из основных примеров векторных пространств.

Пример 8.2 (координатное пространство)**.** Пусть $V = F^n$ — множество последовательностей длины n, составленных из элементов поля F. Сложение и умножение на число определим покомпонентно: для $v = (\nu_1, \nu_2, \ldots, \nu_n)$, $u = (\mu_1, \mu_2, \ldots, \mu_n)$, где $\nu_i \in F$, $\mu_i \in F$, результаты операций определяются как

$$v + u = (\nu_1 + \mu_1, \nu_2 + \mu_2, \ldots, \nu_n + \mu_n),$$
$$\nu \cdot v = (\nu\nu_1, \nu\nu_2, \ldots, \nu\nu_n).$$

Легко проверить выполнение всех свойств из определения векторного пространства: свойства поля обеспечивают равенства в каждой координате.

Такое векторное пространство называется *n-мерным координатным пространством над полем F.*[6)] □

[6)] Термину «n-мерный» будет дальше придан более общий смысл, согласованный с этим названием.

Для дальнейшего нам будет особенно важен следующий пример.

Пример 8.3. Пусть поле K содержит поле F. Тогда K является векторным пространством над полем F, операции в котором задаются сложением и умножением в поле K. Свойства векторного пространства в этом случае прямо следуют из свойств операций в поле K.

В частном случае $K = \mathbb{C}$, $F = \mathbb{R}$ такая конструкция даёт координатную плоскость для комплексных чисел. □

Аналогично свойствам колец проверяются свойства вычитания и умножения на 0.

Утверждение 8.4. *В любом векторном пространстве V над полем F для любых $\alpha, \beta \in F$ и векторов $u, v \in V$ выполняются равенства*

$$\alpha \cdot v - \beta \cdot v = (\alpha - \beta) \cdot v,$$
$$\alpha \cdot u - \alpha \cdot v = \alpha \cdot (u - v),$$
$$0 \cdot v = 0,$$
$$-v = (-1) \cdot v \quad \text{(противоположный вектор получается умножением на } -1).$$

Доказательство. Вычитая $\beta \cdot v$ из обеих частей равенства

$$(\alpha - \beta) \cdot v + \beta \cdot v = (\alpha - \beta + \beta) \cdot v = \alpha \cdot v \quad \text{(дистрибутивность)},$$

получаем первое равенство утверждения.

Второе равенство получается из аналогичного равенства

$$\alpha \cdot u = \alpha \cdot \big((u - v) + v\big) = \alpha \cdot (u - v) + \alpha \cdot v$$

(второе свойство дистрибутивности) вычитанием $\alpha \cdot v$ из обеих частей равенства.

Третье равенство следует из дистрибутивности вычитания так же, как в случае колец: $0 \cdot v = (1 - 1) \cdot v = 1 \cdot v - 1 \cdot v = 0$. Из него легко следует последнее равенство, так как $v + (-1) \cdot v = 1 \cdot v + (-1) \cdot v = (1 - 1) \cdot v = 0 \cdot v = 0$. □

Повторим обычные для линейной алгебры определения линейной комбинации, независимого и порождающего множества, базиса.

Определение 8.5. Вектор

$$v = \sum_{i=1}^{m} \alpha_i v_i$$

называется *линейной комбинацией* векторов $v_1, \ldots, v_n$ с коэффициентами $\alpha_1, \ldots, \alpha_n$. Линейная комбинация называется *нетривиальной*, если не все её коэффициенты равны 0.

Множество векторов V' векторного пространства V называется *порождающим*, если любой вектор из V является линейной комбинацией каких-то векторов из V'.

Векторное пространство V называется *конечномерным*, если в нем существует конечное порождающее множество. В противном случае V называется *бесконечномерным*.

Множество векторов V' векторного пространства V называется *линейно независимым*, если любая нетривиальная линейная комбинация векторов из V' не равна нулю. То есть, из равенства $\sum_{i=1}^{k} \alpha_i v_i = 0$, где $\alpha_i \in F$, $v_i \in V$, следует $\alpha_i = 0$ для всех i.

Упорядоченное множество векторов V' векторного пространства V называется *базисом*, если оно одновременно является и порождающим, и линейно независимым множеством.

Замечание 8.6. От порядка векторов зависят коэффициенты линейных комбинаций. Поэтому удобно рассматривать не просто множества векторов, а упорядоченные множества (то есть последовательности). Далее мы называем упорядоченные множества векторов *системами векторов*, чтобы не использовать лишний раз перегруженное слово «последовательность».

Пример 8.7. Координатное пространство F^n является конечномерным. Базис в нём образуют *координатные* векторы

$$e_i = (\underbrace{0, \ldots, 0}_{i-1 \text{ нулей}}, 1, \underbrace{0, \ldots, 0}_{n-i \text{ нулей}}), \quad 1 \leqslant i \leqslant n.$$

Разложение в линейную комбинацию координатных векторов задаётся координатами вектора:

$$(\alpha_1, \ldots, \alpha_n) = \sum_{i=1}^{n} \alpha_i e_i.$$

Из этого же равенства нетрудно видеть, что у равной нулю линейной комбинации координатных векторов все коэффициенты равны 0 (они всегда совпадают с координатами вектора, равного линейной комбинации). □

Пример 8.8. Многочлены $F[x]$ с коэффициентами в поле F образуют векторное пространство над F с операциями сложения многочленов и умножения на константу (элемент поля F). Поскольку многочлены — это кольцо с 1, свойства векторного пространства сразу следуют из свойств кольца.

Это пространство бесконечномерное: степень многочленов в любом конечном множестве многочленов ограничена, скажем, числом d, степень любой линейной комбинации этих многочленов ограничена тем же числом d (или линейная комбинация нулевая). Поэтому многочлен x^{d+1}, степень которого $d + 1$, не равен линейной комбинации многочленов из этого множества.

Множество многочленов $\{x^k : k \in \mathbb{N}\}$ является порождающим: любой многочлен по определению равен линейной комбинации таких многочленов.

Это же множество является линейно независимым, так как у нулевого многочлена все коэффициенты равны 0. □

Пример 8.9. Рассмотрим поле действительных чисел $\mathbb{R}$ как векторное пространство над полем $\mathbb{Q}$. Это пространство бесконечномерное. Любое доказательство этого факта требует использования каких-то свойств рациональных и действительных чисел.

Самое простое — воспользоваться счётностью рациональных чисел и несчётностью действительных. Обычными средствами наивной теории множеств нетрудно показать, что множество линейных комбинаций с рациональными коэффициентами конечного множества действительных чисел счётно (так как множество конечных последовательностей элементов счётного множества счётно). Поэтому такое множество не совпадает с несчётным множеством всех действительных чисел. □

Проверим стандартные свойства введённых понятий и докажем основную теорему о базисах векторного пространства. Монотонность по включению порождающих и независимых множеств очевидна из определения и простого наблюдения: добавление в линейную комбинацию вектора $0 \cdot v$ с нулевым коэффициентом не меняет линейной комбинации.

Утверждение 8.10. *Если добавить к порождающему множеству вектор, то получится порождающее множество.*

Если удалить из независимого множества вектор, то получится независимое множество.

Ещё одно очевидное наблюдение.

Утверждение 8.11. *Любое независимое множество не содержит нулевого вектора.*

Доказательство. Если $0 \in B$, то нетривиальная линейная комбинация

$$1 \cdot 0 + \sum_{b \in B \setminus \{0\}} 0 \cdot b$$

равна 0. □

Другими словами это утверждение означает, что множество, состоящее из одного 0, является линейно зависимым.

Утверждение 8.12. *Каждый элемент конечномерного векторного пространства однозначно представляется в виде линейной комбинации элементов базиса.*

Доказательство. Действительно, если есть два представления в виде линейной комбинации векторов из базиса

$$v = \sum_{i=1}^{n} \alpha_i v_i = \sum_{i=1}^{n} \beta_i v_i,$$

то, вычитая одну из другой, получаем

$$0 = v - v = \sum_{i=1}^{n} \alpha_i v_i - \sum_{i=1}^{n} \beta_i v_i = \sum_{i=1} (\alpha_i v - \beta_i v) = \sum_{i=1}^{n} (\alpha_i - \beta_i) v.$$

По определению линейной независимости отсюда следует, что для всех i выполняется равенство $\alpha_i - \beta_i = 0$, то есть $\alpha_i = \beta_i$. □

Итак, каждому вектору v векторного пространства с конечным базисом $e_1, \ldots, e_n$ взаимно однозначно соответствует набор коэффициентов $(\alpha_1, \ldots, \alpha_n)$ разложения вектора v по этому базису.

Следствие 8.13. *Собственное подмножество базиса не является базисом.*

Доказательство. Коэффициенты разложения базисного вектора b по базису равны 1 для вектора b и 0 для остальных базисных векторов. В силу утверждения 8.12 это разложение единственно, поэтому ни один базисный вектор не является линейной комбинацией других базисных векторов. □

Определение 8.14. Векторные пространства V_1 и V_2 над полем F (линейно) *изоморфны*, если существует биекция $A\colon V_1 \to V_2$, сохраняющая операции, то есть $A(x+y) = A(x) + A(y)$; $A(\lambda \cdot x) = \lambda \cdot A(x)$.

Выше мы фактически доказали утверждение.

Утверждение 8.15. *Разложение по базису задаёт линейный изоморфизм между конечномерным пространством V над полем F и координатным пространством F^n, где n — количество векторов в базисе.*

Доказательство. Биекция уже задана выше: вектору из пространства V сопоставляется набор коэффициентов разложения вектора по базису. Это последовательность элементов поля F длины n, то есть как раз вектор координатного пространства F^n.

Проверка сохранения операций прямолинейная: если

$$v_1 = \sum_i \lambda_i b_i, \quad v_2 = \sum_i \mu_i b_i, \quad \lambda \in F,$$

то

$$v_1 + v_2 = \sum_i \lambda_i b_i + \sum_i \mu_i b_i = \sum_i \big(\lambda_i b_i + \mu_i b_i\big) = \sum_i (\lambda_i + \mu_i) b_i,$$
$$\lambda \cdot v_1 = \lambda \cdot \sum_i \lambda_i b_i = \sum_i \big(\lambda \lambda_i\big) b_i.$$

Однозначность, как уже говорилось, следует из утверждения 8.12. □

Утверждение 8.16. *В каждом конечномерном пространстве есть конечный базис.*

Доказательство. Выберем конечное порождающее множество векторов V'. Среди подмножеств V', которые являются порождающими для V, выберем минимальное по включению подмножество B и упорядочим его. Это и будет базис. Действительно, если

$$\sum_{i=1}^{k} \alpha_i b_i = 0, \quad \alpha_i \in F,\ b_i \in B,$$

и среди коэффициентов линейной комбинации есть ненулевой, скажем $\alpha_j \neq 0$, то b_j выражается через остальные векторы из B:

$$b_j = -\alpha_j^{-1} \sum_{i \neq j} \alpha_i b_i$$

(здесь существенно, что у любого ненулевого элемента поля есть обратный). Поэтому множество $B \setminus \{b_j\}$ также порождающее: для любого $v \in V$ имеем

$$v = \sum_{i=1}^{k} \beta_i b_i = \sum_{i \neq j} \beta_i b_i - \beta_j \alpha_j^{-1} \sum_{i \neq j} \alpha_i b_i = \sum_{i \neq j} \left(\beta_i - \alpha_i \beta_j \alpha_j^{-1} \right) b_i.$$

Это противоречит минимальности B по включению. □

Из доказанных утверждений следует, что каждое конечномерное пространство над полем F изоморфно координатному пространству F^n. Однако мы пока не выяснили, различаются ли пространства F^n и F^k при $n \neq k$ (с точностью до изоморфизма). Ответ окажется положительным, но доказательство этого утверждения не вполне прямолинейно.

Замечание 8.17. Если поле конечно, то неизоморфизм F^n и F^k при $n \neq k$ доказывается очень просто: достаточно подсчитать количество элементов координатного пространства. По комбинаторному правилу произведения получаем $|F^n| = |F|^n$. Поэтому при $n \neq k$ координатные пространства F^n и F^k содержат разное количество элементов, так что между ними нет никакой биекции, а не только биекции, сохраняющей операции.

Нас в основном интересуют именно конечные поля. Уже из утверждения 8.16 получаем важную теорему о конечных полях.

Теорема 8.18. *В любом конечном поле p^n элементов, где p — простое, n — натуральное.*

Доказательство. Характеристика конечного поля F конечна, и это простое число. Как мы знаем из леммы 7.36, поле F содержит подполе $\mathbb{Z}/(p)$.

Как векторное пространство над полем $\mathbb{Z}/(p)$ поле F конечномерно. Например, конечное порождающее множество образуют все элементы поля (любой элемент $v \in F$ выражается линейной комбинацией $1 \cdot v$). Поэтому в нём есть базис, так что количество элементов в поле F равно p^n, где n — количество элементов в базисе. □

Для полноты изложения приведём также доказательство теоремы о базисах в общем случае.

Лемма 8.19 (лемма о замене)**.** *Пусть $B = (b_1, \ldots, b_n)$ — базис векторного пространства V над полем F. Если в разложении вектора a по базису B коэффициент при $b_s \in B$ не равен нулю, то система векторов $(a, b_1, \ldots, b_{s-1}, b_{s+1}, \ldots, b_n)$ также является базисом.*

Доказательство. Пусть

$$a = \sum_{i=1}^{n} \mu_i b_i = 0, \quad \mu_s \neq 0.$$

Проверим, что система векторов $B' = (a, b_1, \ldots, b_{s-1}, b_{s+1}, \ldots, b_n)$ линейно независима. Действительно, линейная комбинация векторов из B' выражается как линейная комбинация векторов из B:

$$\alpha a + \sum_{i \neq s} \gamma_i b_i = \alpha \cdot \left(\sum_i \mu_i b_i \right) + \sum_{i \neq s} \gamma_i b_i = \alpha \mu_s b_s + \sum_{i \neq s} (\gamma_i + \alpha \mu_i) b_i.$$

Если левая часть равенства равна нулю, но не все её коэффициенты равны нулю, то $\alpha \neq 0$ (иначе получаем нетривиальную линейную комбинацию векторов из B, равную нулю). Но тогда равна нулю и правая часть равенства, причём коэффициент $\alpha\mu_s$ при b_s не равен нулю (в поле делителей нуля нет). Из независимости системы векторов B получаем независимость системы векторов B'.

Проверим, что множество B' порождающее. Вектор b_s выражается как линейная комбинация векторов из этого множества:

$$b_s = \mu_s^{-1} a - \sum_{i \neq s} \mu_s^{-1} \mu_i b_i.$$

Поэтому любая линейная комбинация векторов из B (то есть любой вектор пространства) является также и линейной комбинацией векторов из B':

$$\sum_i \gamma_i b_i = \gamma_s b_s + \sum_{i \neq s} \gamma_i b_i = \gamma_s \mu_s^{-1} a + \sum_{i \neq s} (\gamma_i - \gamma_s \mu_s^{-1} \mu_i) b_i.$$

Итак, мы доказали, что B' базис. □

Используя лемму о замене, можно доказать, что любое независимое множество можно расширить до базиса, добавляя в него элементы из некоторого заданного базиса.

Лемма 8.20 (продолжение до базиса)**.** *Пусть B — базис конечномерного векторного пространства V над полем F, A — конечное независимое множество. Тогда существует базис A', который содержит все элементы A и, быть может, какие-то элементы B. Количество элементов в A' такое же, как в B.*

Доказательство. Индукция по количеству k элементов в $B \setminus A$.

База $k = 0$: в этом случае $B \subseteq A$. По следствию 8.13 собственное подмножество базисом не является, так что $A = B$. Лемма выполняется тривиальным образом.

Индуктивный переход: предположим, что утверждение леммы доказано для всех независимых множеств A и базисов B, для которых $|B \setminus A| < k$.

Рассмотрим пару A, B, для которой $|B \setminus A| = k > 0$. Если $A \subset B$, то в качестве A' опять можно взять сам базис B. Иначе есть вектор $a \in A \setminus B$, причём он не равен 0, так как A независимое. В разложении a по базису B коэффициент при каком-то

векторе $b_s \in B \setminus A$ не равен 0, так как в противном случае из разложения

$$a = \sum_{i:b_i \in A} \mu_i b_i$$

получаем равную нулю нетривиальную линейную комбинацию векторов из A:

$$a - \sum_{i:b_i \in A} \mu_i b_i = 0.$$

По лемме о замене система векторов $B' = (a, b_1, \ldots, b_{s-1}, b_{s+1}, \ldots, b_n)$ является базисом с тем же количеством элементов, что и в базисе B. Для нового базиса B' выполняется неравенство $|B' \setminus A| = |B \setminus A| - 1 < k$. По индуктивному предположению существует базис A', который состоит из векторов A и ещё, быть может, каких-то векторов из B', причём количество элементов в этом базисе равно $|B'| = |B|$.

Индуктивный переход доказан, справедливость леммы теперь следует из принципа математической индукции. □

Теперь уже легко доказать основную теорему о базисах.

Теорема 8.21. *Все базисы в конечномерном векторном пространстве содержат одинаковое количество элементов.*

Доказательство. Рассмотрим два базиса A, B и применим лемму о продолжении базиса к базису A как независимому множеству и к базису B как базису. Получим базис A', который содержит не меньше элементов, чем базис A (содержит все элементы A) и столько же элементов, сколько есть в базисе B. Но собственное подмножество базиса базисом не является. Поэтому $A' = A$ и $|A| = |A'| = |B|$. □

Определение 8.22. Количество элементов в базисе конечномерного векторного пространства V называется *размерностью пространства*, обозначение $\dim V$.

Пример 8.23. Чему равна размерность координатного пространства F^n? Построенные в примере 8.7 векторы e_i, $1 \leqslant i \leqslant n$, образуют и порождающее, и независимое множество. То есть это базис, в котором n элементов. Поэтому $\dim F^n = n$. □

Контрольный вопрос 8.24. Докажите, что в n-мерном векторном пространстве V над полем из q элементов содержится ровно $q^{\dim V}$ векторов.

Пример 8.25. Какие значения может принимать размерность векторного пространства? Как следует из примера координатного пространства F^k, размерность может принимать любые целые положительные значения.

Однако это не все возможные случаи. Существует вырожденное векторное пространство, состоящее из одного нулевого вектора.

Контрольный вопрос 8.26. Проверьте выполнение аксиом векторного пространства в этом случае.

В этом случае базисом нужно считать пустое множество векторов (оно всегда линейно независимо, как может проверить знакомый с формальной логикой читатель). В пустом множестве 0 элементов, так что векторное пространство, состоящее из одного вектора 0, имеет размерность 0. □

Следствие 8.27 (теоремы 8.21)**.** *В любой независимой системе векторов конечномерного векторного пространства V над полем F не больше элементов, чем размерность V.*

Доказательство. Пусть d — размерность пространства V, а B — базис, в нём как раз d векторов.

Предположим противное: нашлась независимая система A, которая содержит больше d векторов. Применим лемму о продолжении до базиса к конечному независимому $A' \subseteq A$, в котором $d+1$ вектор, и к базису B. Получим базис A'', который содержит A', то есть не менее $d+1$ вектора. Это противоречит теореме о базисе. Значит, предположение ложно, а утверждение следствия истинно. □

Следствие 8.28. *Если $n = \dim V$ — размерность пространства V, то любая линейно независимая система из n векторов является базисом.*

Доказательство. Пусть $b_1, \ldots, b_k$ — линейно независимая система векторов, которая не порождает пространство V. Это означает, что некоторый вектор v не является линейной комбинацией b_i. Но тогда система $v, b_1, \ldots, b_k$ также линейно независима: из

$$\lambda v + \sum_i \mu_i b_i = 0$$

следует $\lambda = 0$ (иначе v выражается как $-\lambda^{-1} \sum_i \mu_i b_i$), а в силу линейной независимости b_1, …, b_k и $\mu_i = 0$ для всех i.

Из следствия 8.27 получаем неравенство $k + 1 < n$, которое не выполняется при $k = n$. Таким образом, любая линейно независимая система из $\dim V$ векторов порождает пространство V и потому является базисом. □

Пример 8.29. Многочлены степени не выше d с коэффициентами из поля F образуют векторное пространство над полем F размерности $d+1$, так как многочлены 1, x, x^2, …, x^d образуют базис в этом пространстве.

Но можно указать и другие интересные базисы в этом пространстве. Например, если в поле F больше d элементов, то для любого набора $a_0, \ldots, a_d$ из различных элементов поля существуют такие многочлены $f_0, \ldots, f_d$, что $f_i(a_i) = 1$ и $f_i(a_j) = 0$ при $i \neq j$. Многочлены $f_0, \ldots, f_d$ образуют базис. Действительно, многочлен $f = \sum_{i=0}^{d} \alpha_i f_i = 0$ принимает в каждой точке a_i значение α_i. Поэтому линейная зависимость между такими многочленами $\alpha_0 f_0 + \alpha_1 f_1 + \cdots + \alpha_d f_d = 0$ означает, что все α_i равны 0 (значение нулевого многочлена в любой точке).

Выпишем явно выражения для многочленов f_i:

$$f_0 = \frac{(x-a_1)(x-a_2)\dots(x-a_d)}{(a_0-a_1)(a_0-a_2)\dots(a_0-a_d)},$$
$$f_1 = \frac{(x-a_0)(x-a_2)\dots(x-a_d)}{(a_1-a_0)(a_1-a_2)\dots(a_1-a_d)},$$
$$\dots,$$
$$f_d = \frac{(x-a_0)(x-a_1)\dots(x-a_{d-1})}{(a_d-a_0)(a_d-a_1)\dots(a_d-a_{d-1})}$$

Коэффициенты разложения многочлена f по этому базису равны значениям многочлена в точках $a_0, \dots, a_d$, а само разложение по этому базису называется *интерполяционной формулой Лагранжа*:

$$f = \sum_{i=0}^{d} f(a_i) f_i \,. \tag{8.1}$$

□

Пример 8.30. В случае конечного поля F интерполяционная формула Лагранжа даёт следующий факт: для любой функции $\varphi\colon F \to F$ найдётся такой многочлен $f \in F[x]$, что $f(a) = \varphi(a)$ для любого $a \in F$.

В данном случае нужно записать интерполяционную формулу Лагранжа по всем точкам поля. □

8.2 Подпространства

Определение 8.31. *Подпространством* векторного пространства V называется подмножество, которое является векторным пространством над тем же полем относительно тех же операций, что и в V.

Утверждение 8.32. *Множество линейных комбинаций векторов из данного множества S векторов векторного пространства V является подпространством V.* Название: подпространство, порождённое множеством S; обозначение $F\langle S\rangle$.

Доказательство. Так как $-v = (-1)\cdot v$, достаточно проверить, что множество линейных комбинаций векторов из S замкнуто относительно сложения и умножения на элемент поля. Это проверяется вычислениями, аналогичными доказательству утверждения 8.15. Если

$$v_1 = \sum_i \lambda_i u_i, \quad v_2 = \sum_i \mu_i u_i, \quad \lambda \in F, \quad u_i \in S,$$

то

$$v_1 + v_2 = \sum_i \lambda_i u_i + \sum_i \mu_i u_i = \sum_i \big(\lambda_i u_i + \mu_i u_i\big) = \sum_i (\lambda_i + \mu_i) u_i,$$
$$\lambda \cdot v_1 = \lambda \cdot \sum_i \lambda_i u_i = \sum_i \big(\lambda\lambda_i\big) u_i,$$

что и требуется. □

Теорема 8.33. *Любое подпространство L конечномерного пространства V конечномерно и его размерность не превосходит размерность пространства V.*

Доказательство. Выберем какой-нибудь базис B в пространстве V и будем его перестраивать, пользуясь леммой о замене. Предположим, что в L лежат k базисных векторов, обозначим составленное из них множество через B'. Если векторы из B' порождают L, то B' — базис L (так как подмножество базиса независимо).

В противном случае найдётся вектор $a \in L$, который не является линейной комбинацией векторов из B'. Однако этот вектор раскладывается по базису B:

$$a = \sum_{b \in B} \mu_b b.$$

Поскольку a не является линейной комбинацией векторов из B', найдётся такой $b \in B \setminus B'$, что $\mu_b \neq 0$. Применяя к векторам a, b и базису B лемму о замене, получаем базис $B' = \{a\} \cup (B \setminus \{b\})$, в котором уже $k+1$ вектор лежит в подпространстве L.

Продолжаем этот процесс пока возможно. Из описания процесса видно, что он остановится лишь в том случае, когда текущее множество $B' = B \cap L$ порождает L. Но процесс когда-то останавливается, так как k не больше размерности V. Значит, результатом процесса будет такой базис V, часть которого является базисом L. Отсюда и получаем оценку на размерность. □

Фактически мы доказали больше. А именно, из доказательства теоремы 8.33 немедленно извлекается такое следствие:

Лемма 8.34. *Для любого подпространства L конечномерного пространства V существует базис V, начальная часть которого является базисом L.*

Пример 8.35. Рассмотрим важный для теории графов пример подпространства пространства над полем из двух элементов $\mathbb{F}_2 = \{0, 1\}$.

Пусть $G(V, E)$ — неориентированный граф без петель (параллельные рёбра разрешаются), в котором n вершин. Векторное пространство $\mathbb{F}_2^E$ — это координатное пространство, координаты в котором индексированы не числами, а рёбрами графа (если это непонятно, то представьте, что рёбра графа занумерованы, тогда $\mathbb{F}_2^E$ в точности соответствует данному выше определению координатного пространства).

Каждому ребру $e \in E$ отвечает координатный базисный вектор в этом пространстве: координата e у такого вектора равна 1, а остальные координаты равны 0. Для наглядности будем обозначать этот вектор точно так же: e.

Общий вектор $x \in \mathbb{F}_2^E$ в этом пространстве имеет координаты 0 и 1 и потому однозначно задаётся подмножеством $X \subseteq E$ рёбер графа. Используя введённые выше обозначения для базисных координатных векторов запишем этот вектор как

$$x = \sum_{e \in X} e.$$

Чтобы подчеркнуть связь между векторами и подмножествами будем обозначать такой вектор как $\mathbb{1}_X$.

Сложению векторов отвечает симметрическая разность множеств:

$$\mathbb{1}_X + \mathbb{1}_Y = \mathbb{1}_{X \triangle Y},$$

так как симметрическая разность множеств X и Y состоит из тех $e \in E$, которые входят в точности в одно из этих множеств, а координата e равна 1 в $\mathbb{1}_X + \mathbb{1}_Y$ в точности тогда, когда ровно одна из координат e в векторах $\mathbb{1}_X$ и $\mathbb{1}_Y$ равна 1.

Пока вершины графа никак не участвовали в происходящем. Теперь определим векторы-окрестности вершин $g_v = \mathbb{1}_{E(v)}$, где $E(v)$ — множество тех рёбер, одним из концов которых является вершина v.

Пространством разрезов $\mathrm{Cut}(G)$ назовём подпространство $\mathbb{F}_2\langle g_v : v \in V\rangle$ пространства $\mathbb{F}_2^E$, порождённое векторами-окрестностями вершин. Чтобы понять, откуда взялось такое название, посмотрим на элемент этого пространства, который равен сумме некоторых векторов-окрестностей

$$c(S) = \sum_{v \in S \subseteq V} g_v = \mathbb{1}_C$$

для некоторого множества C (напомним, что в поле всего два элемента, так что каждый ненулевой коэффициент в линейной комбинации равен 1). У ребра два конца, поэтому в множество C войдут те рёбра, у которых ровно один конец лежит в множестве S, а другой — в дополнении $V \setminus S$ к этому множеству. Такое множество рёбер в теории графов называется *разрезом*, задаваемым множеством вершин S.

Нетрудно видеть, что $c(V) = 0$ (в этом случае ни одно ребро не имеет координату, равную 1). Это означает линейную зависимость между векторами-окрестностями:

$$\sum_{v \in V} g_v = 0. \tag{8.2}$$

Поэтому размерность $\mathrm{Cut}(G)$ меньше n: как мы проверили в доказательстве утверждения 8.16, каждое порождающее множество содержит базис; а $\mathrm{Cut}(G)$ порождается n векторами g_v, $v \in V$.

Проверим, что $\dim \mathrm{Cut}(G) = n - 1$ для связного графа G. Пусть

$$0 = \sum_{v \in V} \lambda_v g_v = c(S) = \mathbb{1}_\varnothing, \qquad S \neq \varnothing,$$

нетривиальная линейная зависимость (опять используем то обстоятельство, что в поле из двух элементов любой ненулевой элемент равен 1).

Чтобы в разрезе, задаваемом множеством S, не было ни одного ребра, каждое ребро должно иметь оба конца или в S, или вне S. Но если $S \neq V$, это противоречит связности графа: и в S, и вне S в таком случае есть хотя бы одна вершина, но нет путей в графе, соединяющих эту пару вершин.

Выберем вершину w в графе. Из наших рассуждений вытекает, что множество векторов $\{g_v : v \in V \setminus \{w\}\}$ независимое. Но оно же и порождающее для $\mathrm{Cut}(G)$, так как из (8.2) следует

$$g_w = -\sum_{v \in V \setminus \{w\}} g_v,$$

а через векторы g_v, $v \in V$, любой вектор из пространства разрезов выражается по определению. □

Упражнение 8.36. Докажите, что в общем случае $\dim \mathrm{Cut}(G) = n - k$, где k — количество компонент связности графа G.

8.3 Линейные преобразования и матрицы

Определение 8.37. Отображение $A\colon V \to U$ одного векторного пространства над полем F в другое над тем же полем называется *линейным*, если для любых $v_1, v_2 \in V$ и $\lambda_1, \lambda_2 \in F$ выполняется равенство

$$A(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 A(v_1) + \lambda_2 A(v_2).$$

Линейное отображение является гомоморфизмом векторных пространств как групп по сложению: если выбрать в определении $\lambda_1 = \lambda_2 = 1$, то получим как раз условие сохранения групповой операции.

Линейные отображения обладают свойствами, похожими на свойства гомоморфизма.

Ядро и образ линейного отображения определяются так же, как для гомоморфизмов групп:

$$\mathrm{Ker}\, A = \{v \in V : A(x) = 0\}, \qquad \mathrm{Im}\, A = \{u \in U : u = A(x),\ x \in V\}.$$

Утверждение 8.38. *Ядро и образ линейного отображения $A\colon V \to U$ являются подпространствами векторных пространств V и U соответственно.*

Доказательство. Ядро: достаточно проверить, что для любых $v_1, v_2 \in \mathrm{Ker}\, A$ и $\lambda_1, \lambda_2 \in F$ выполняется $\lambda_1 u_1 + \lambda_2 u_2$. Это верно, так как по определению из $A(v_1) = = A(v_2) = 0$ следует

$$A(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1 A(v_1) + \lambda_2 A(v_2) = \lambda_1 \cdot 0 + \lambda_2 \cdot 0 = 0.$$

Аналогично для образа: если $u_1 = A(v_1) \in \mathrm{Im}\, A$, $u_2 = A(v_2) \in \mathrm{Im}\, A$, то

$$\lambda_1 u_1 + \lambda_2 u_2 = \lambda_1 A(v_1) + \lambda_2 A(v_2) = A(\lambda_1 v_1 + \lambda_2 v_2)$$

для любых $v_1, v_2 \in V$. □

Как мы знаем, для любого гомоморфизма конечных групп $\varphi\colon G \to H$ выполняется равенство $|G| = |\mathrm{Ker}\, \varphi| \cdot |\mathrm{Im}\, \varphi|$ (лемма 4.16). Если V, U — конечномерные векторные пространства над конечным полем из q элементов, то $|V| = q^{\dim V}$, $|\mathrm{Ker}\, A| = = q^{\dim \mathrm{Ker}\, A}$, $|\mathrm{Im}\, A| = q^{\dim \mathrm{Im}\, A}$ для любого линейного отображения $A\colon V \to U$.

Логарифмируя равенство $|V| = |\mathrm{Ker}\, A| \cdot |\mathrm{Im}\, A|$, получаем соотношение между размерностями

$$\dim V = \dim \mathrm{Ker}\, A + \dim \mathrm{Im}\, A. \tag{8.3}$$

То же самое соотношение выполняется и для любого линейного отображения конечномерных векторных пространств.

Теорема 8.39. *Пусть V, U — конечномерные векторные пространства (над произвольным полем), $A\colon V \to U$ — линейное отображение. Тогда для размерностей V, ядра и образа A выполняется равенство* (8.3).

Доказательство. По лемме 8.34 в пространстве V существует такой базис B, часть которого (быть может пустая, если ядро нулевое) является базисом ядра $\operatorname{Ker} A$.

Разобьём этот базис B на две части: $B = B_K \sqcup B_I$, где B_K — базис ядра, а $B_I = B \setminus B_K$ — остальные базисные векторы.

Докажем, что $A(B_I) = \{u : u = A(b),\ b \in B_I\}$ является базисом в пространстве $\operatorname{Im} A$ и $|A(B_I)| = |B_I|$ (базисные векторы из B_I отображаются в разные векторы). Отсюда следует теорема, так как

$$\dim V = |B|, \quad \dim \operatorname{Ker} A = |B_K|, \quad \dim \operatorname{Im} A = |A(B_I)| = |B_I|, \quad |B| = |B_K| + |B_I|.$$

Пусть $b', b'' \in B_I$ и $A(b') = A(b'')$. Тогда из линейности отображения получаем

$$A(b' - b'') = A(b') - A(b'') = 0,$$

то есть $b' - b'' \in \operatorname{Ker} A$ и потому является линейной комбинацией векторов из B_K вида

$$\sum_{b_k \in B_K} \lambda_k b_k = b' - b''.$$

Но тогда получаем линейную зависимость между векторами базиса B:

$$b' - b'' - \sum_{b_k \in B_K} \lambda_k b_k = 0,$$

что возможно лишь при $b' = b''$ и $\lambda_k = 0$ для всех $b_k \in B_K$.

Итак, образы всех векторов из множества B_I базисных векторов B_I различны. Поэтому $|A(B_I)| = |B_I|$.

Аналогично докажем, что $A(B_I)$ линейно независимое. Рассмотрим равную нулю линейную комбинацию векторов из $A(B_I)$:

$$0 = \sum_{b \in B_I} \mu_b A(b) = A\left(\sum_{b \in B_I} \mu_b b\right).$$

Во втором равенстве мы воспользовались линейностью отображения. Вектор, который стоит в скобках в последней части этих равенств, принадлежит ядру. Поэтому он разлагается в линейную комбинацию векторов B_K, то есть

$$\sum_{b \in B_I} \mu_b b - \sum_{b \in B_K} \lambda_b b = 0$$

для каких-то коэффициентов λ_b, $b \in B_K$. Получили линейную комбинацию векторов из базиса B, равную 0. Поэтому $\mu_b = 0$, $b \in B_I$, что и означает линейную независимость векторов из $A(B_I)$ (здесь важно, что множества B_I и B_K не пересекаются).

Докажем, что $A(B_I)$ порождает $\operatorname{Im} A$. Возьмём вектор $u \in \operatorname{Im} A$, по определению он является образом некоторого вектора $v \in V$: $u = A(v)$. Разложим вектор v по

базису B:

$$v = \sum_{b \in B_K} \mu_b b + \sum_{b \in B_I} \lambda_b b.$$

Тогда из линейности отображения A получаем

$$u = A(v) = A\left(\sum_{b \in B_K} \mu_b b + \sum_{b \in B_I} \lambda_b b\right) = \sum_{b \in B_K} \mu_b A(b) + \sum_{b \in B_I} \lambda_b A(b) = \sum_{b \in B_I} \lambda_b A(b),$$

так как $A(b) = 0$, если $b \in B_K$ (базис ядра). Полученное разложение показывает, что $A(B_I)$ порождает $\operatorname{Im} A$.

Итак, $A(B_I)$ линейно независимое и порождает $\operatorname{Im} A$. Значит, это базис в $\operatorname{Im} A$. □

Пусть в пространствах V и U заданы базисы $e_1, \ldots, e_n$ и $f_1, \ldots, f_m$ соответственно. Линейное отображение $A\colon V \to U$ однозначно задаётся образами базисных векторов

$$a_i = A(e_i) = \sum_{k=1}^{m} A_{ki} f_k, \quad A_{ki} \in F,$$

а именно

$$A(v) = \sum_{i=1}^{n} \lambda_i a_i, \qquad \text{где } v = \sum_i \lambda_i e_i,$$

то есть λ_i — это коэффициенты разложения вектора v по базису $e_1, \ldots, e_n$.

В обратную сторону: любая матрица A размеров $m \times n$, матричные элементы которой принадлежат полю F, задаёт линейное отображение

$$A : v = \sum_{i=1}^{n} \lambda_i e_i \mapsto \sum_{k=1}^{m} (A_{ki}\lambda_i)\, f_k.$$

Контрольный вопрос 8.40. Проверьте, что эта формула и впрямь корректно определяет линейное отображение.

Пример 8.41. Рассмотрим тождественное отображение $\mathrm{id}\colon V \to V$, задаваемое правилом $\mathrm{id}\, v = v$. Если выбрать один и тот же базис для векторов и их образов, то матрица I тождественного отображения единичная: диагональные элементы I_{kk} равны 1, а внедиагональные равны 0: $I_{jk} = 0$. □

Контрольный вопрос 8.42. Проверьте это утверждение непосредственно из определений.

Таким образом, между матрицами размеров $m \times n$ с элементами из поля F и линейными отображениями векторных пространств над полем F с фиксированными упорядоченными базисами имеется взаимно однозначное соответствие.

Замечание 8.43. При замене базиса матрицы линейных отображений изменяются. Нам это не понадобится, поэтому мы не обсуждаем правила замены (это линейные отображения пространства матриц). Заинтересованный читатель может найти нужную информацию в любом учебнике по линейной алгебре. От того, над каким полем рассматриваются пространства, формулы преобразований не зависят.

Если A — матрица, задающая линейное отображение, то её столбцы — это в точности образы базисных векторов. *Столбцовым рангом* матрицы называется размерность пространства, порождённого столбцами матрицы, то есть образа линейного отображения, задаваемого матрицей A.

Множество решений системы линейных уравнений $Ax = 0$ — это в точности ядро линейного отображения, задаваемого матрицей A.

Теорема 8.39 показывает, что количество столбцов матрицы равно столбцовому рангу матрицы плюс размерность решений системы $Ax = 0$.

8.4 Двойственность

В этом разделе мы для простоты обсуждаем только конечномерные векторные пространства.

Подпространство векторного пространства V можно задать системой векторов, порождающих это пространство. Пусть в системе n векторов $g_1, \ldots, g_n$, а размерность всего пространства равна m. С этой системой естественно связать линейное отображение координатного пространства F^n в V:

$$Gv = \sum_{i=1}^{n} v_i g_i, \quad \text{где } v = (v_1, \ldots, v_n).$$

Если и само пространство V координатное или в нём выбран базис, то этому линейному отображению соответствует матрица, которую мы для наглядности обозначим точно так же:

$$G = \begin{pmatrix} g_{11} & \cdots & g_{1n} \\ \cdots & \cdots & \cdots \\ g_{m1} & \cdots & g_{mn} \end{pmatrix}.$$

Столбцы этой матрицы как раз векторы g_i. Интересующее нас подпространство является образом линейного отображения G.

Матрица G называется *порождающей матрицей* подпространства.

Как мы уже видели, ядро линейного отображения также является подпространством. Если в пространстве задан базис, то подпространство можно выделить системой уравнений $Hx = 0$, то есть охарактеризовать его как ядро некоторого линейного отображения. В этом случае матрица H называется *проверочной матрицей* подпространства. Её строки задают коэффициенты однородных линейных уравнений, которые выполняются на этом пространстве, причём на любом векторе не из подпространства хотя бы одно из уравнений нарушается.

Задание подпространства проверочной матрицей также универсально.

Утверждение 8.44. *Любое подпространство — ядро некоторого отображения.*

Доказательство. Напомним, что мы ограничиваемся лишь случаем конечномерных пространств (хотя утверждение верно и в общем случае).

Пусть L — d-мерное подпространство n-мерного пространства V. Лемма 8.34 говорит, что существует существует базис $b_1, \ldots, b_n$ пространства V, первые d векторов которого образуют базис подпространства L.

Зададим линейное отображение $H\colon V \to F^{n-d}$ правилом

$$H\colon v \mapsto (v_{d+1}, \ldots, v_n),$$

где v_i — координаты разложения вектора v по базису.

У векторов подпространства L в разложении по базису все координаты, начиная с $(d+1)$-й, равны 0 (эти векторы являются линейными комбинациями первых d базисных векторов).

Верно и обратное: если для координат вектора v выполняются равенства

$$v_{d+1} = v_{d+2} = \cdots = v_n = 0,$$

то этот вектор является линейной комбинацией первых d базисных векторов и потому принадлежит подпространству L.

Таким образом, $\operatorname{Ker} H = L$, что и требовалось доказать. □

Эти два способа задания подпространств можно описать более общим образом.

Частным случаем линейных отображений являются линейные отображения векторного пространства V в поле F, над которым определено это пространство, более точно, в 1-мерное координатное пространство над полем F. Такие линейные отображения называются *линейными функционалами*. Если в пространстве задан базис $b_1, \ldots, b_n$, то линейный функционал φ однозначно задаётся значениями на базисных векторах. Если

$$\varphi(b_1) = \varphi_1,\ \varphi(b_2) = \varphi_2,\ \ldots,\ \varphi(b_n) = \varphi_n, \quad n = \dim V,$$

то из свойства линейности функционала получаем

$$\varphi(v) = \varphi\left(\sum_i v_i b_i\right) = \sum_i v_i \varphi(b_i) = \sum_i v_i \varphi_i. \tag{8.4}$$

Здесь v_i — координаты разложения вектора v по базису $b_1, \ldots, b_n$.

Таким образом, линейный функционал является однородной линейной функцией от координат в базисе.[7)]

Линейные функционалы образуют векторное пространство над тем же полем. Операции задаются как

$$(\varphi + \psi)(x) = \varphi(x) + \psi(x),$$
$$(\lambda \cdot \varphi)(x) = \lambda \cdot \varphi(x).$$

Контрольный вопрос 8.45. Проверьте, что из свойства линейности функционалов следует выполнение аксиом векторного пространства для этих операций.

Пространство линейных функционалов обозначается V^* и называется пространством, *сопряжённым* пространству V (или *двойственным* к пространству V).

Утверждение 8.46. *Если V — конечномерное пространство размерности n, то сопряжённое пространство также конечномерно и его размерность также равна n.*

[7)] Термин «функционал» сокращает выражение «однородная функция».

Доказательство. Выберем в V базис $b_1, \ldots, b_n$. Определим функционалы δ_i на базисных векторах правилом:

$$\delta_i(b_j) = \begin{cases} 1, & \text{если } i = j; \\ 0, & \text{в противном случае.} \end{cases}$$

Эти функционалы порождают сопряжённое пространство V^*: если значения линейного функционала φ на базисных векторах $b_1, \ldots, b_n$ равны $\varphi_1, \ldots, \varphi_n$ соответственно, то

$$\varphi_i = \sum_j \varphi_j \delta_j(b_i)$$

по определению функционалов δ_i. Поэтому из уравнения (8.4) и линейности функционала получаем

$$\varphi\left(\sum_i v_i b_i\right) = \sum_i v_i \varphi_i = \sum_{i,j} v_i \varphi_j \delta_j(b_i) = \sum_j \varphi_j \delta_j \left(\sum_i v_i b_i\right)$$

для любых коэффициентов разложения по базису. То есть

$$\varphi = \sum_j \varphi_j \delta_j,$$

что и даёт разложение функционала φ в линейную комбинацию функционалов δ_j.

Докажем, что функционалы δ_j линейно независимы. Пусть

$$\varphi = \sum_j \varphi_j \delta_j = 0.$$

Нулевой функционал принимает значение 0 на любом векторе. Поэтому $0 = \varphi(b_i) = \varphi_i$ для всех i. Это и означает линейную независимость. □

Определение 8.47. Базис δ_j, построенный в доказательстве утверждения 8.46 по базису b_i, называется *двойственным к базису* b_i.

Замечание 8.48. Зафиксируем в пространстве базис, а в сопряжённом — двойственный базис. Матрица линейного функционала имеет размер $1 \times n$, где $n = \dim V$ — размерность пространства V, то есть это строка.

Векторы самого пространства тогда удобно представлять как столбцы. Тогда значение функционала на векторе получается по общему правилу матричного умножения матрицы $1 \times n$ (строка) на матрицу $n \times 1$ (столбец). Результат этого умножения — матрица 1×1, то есть элемент поля F.

В силу утверждения 8.46 пространство и сопряжённое к нему линейно изоморфны. Однако для задания изоморфизма требуется указание базиса. Если взять сопряжённое к сопряжённому пространство, то изоморфизм будет *естественным*. А именно, каждый вектор v векторного пространства V задаёт линейный функционал v^{**} на сопряжённом пространстве V^*. Значение этого функционала на элементе

$\varphi \in V^*$, то есть на линейном функционале на пространстве V, равно $\varphi(v)$, формально $v^{**}(\varphi) = \varphi(v)$. Это линейный функционал на V^* потому что мы так определили операции с линейными функционалами, чтобы выполнялось равенство

$$(\lambda_1\varphi_1 + \lambda_2\varphi_2)(v) = \lambda_1\varphi_1(v) + \lambda_2\varphi_2(v),$$

которое и означает, что $v^{**}(\lambda_1\varphi_1 + \lambda_2\varphi_2) = \lambda_1 v^{**}(\varphi_1) + \lambda_2 v^{**}(\varphi_2)$.

Контрольный вопрос 8.49. Пусть в V выбран базис $b_1, \ldots, b_n$. Найдите значение функционала, задаваемого базисным вектором b_i, на базисных функционалах δ_j, построенных выше в доказательстве утверждения 8.46.

Построенное соответствие задаёт отображение $V \to V^{**}$ из пространства в сопряжённое к сопряжённому. Это отображение линейное, так как равенство

$$\varphi(\lambda_1 v_1 + \lambda_2 v_2) = \lambda_1\varphi(v_1) + \lambda_2\varphi(v_2),$$

которое выполняется для любого линейного функционала $\varphi \in V^*$, любых элементов поля $\lambda_1, \lambda_2 \in F$ и любых векторов $v_1, v_2 \in V$, по определению означает равенство

$$(\lambda_1 v_1 + \lambda_2 v_2)^{**} = \lambda_1 v_1^{**} + \lambda_2 v_2^{**}.$$

Утверждение 8.50. *Отображение $V \to V^{**}$ является линейным изоморфизмом векторных пространств.*

Доказательство. Проверим, что ядро отображения нулевое.

Для ненулевого вектора v рассмотрим 1-мерное подпространство $F\langle v\rangle$, порождённое вектором v и применим к нему утверждение 8.34. Получим базис пространства V, содержащий вектор v. Вектор v по построению стоит на первом месте. Значение функционала δ_1 из двойственного базиса равно 1 по определению. Раз $v^{**}(\delta_1) = 1 \neq 0$, то $v^{**} \neq 0$. Это и значит, что ядро нулевое.

Из формулы (8.3) для суммы размерностей ядра и образа получаем, что размерность образа отображения равна $\dim V - 0 = \dim V$. Но и размерность V^{**} равна $\dim V$, как следует из утверждения 8.46. Поэтому отображение сюръективно. □

В силу этого утверждения мы дальше отождествляем пространства V и V^{**}, пользуясь биекцией $v \mapsto v^{**}$.

Итак, естественное соответствие между векторами удаётся установить только для пространства и второго сопряжённого к нему. Однако между подпространствами векторного пространства V и сопряжённого ему пространства V^* имеется естественное соответствие двойственности.

Определение 8.51. *Аннулятором* множества $S \subseteq V$, где V — векторное пространство, называется множество

$$\{\varphi \in V^* : \varphi(v) = 0,\ v \in S\}$$

(обозначение $\operatorname{Ann} S$).

Лемма 8.52. *Пусть V — векторное пространство. Для любого множества S векторов пространства V аннулятор* $\operatorname{Ann} S$ *является подпространством сопряжённого пространства. Кроме того, для подпространства L дополнительно выполняется равенство* $\operatorname{Ann}(\operatorname{Ann} L) = L$ *(здесь V и V^{**} отождествляются как указано выше).*

Доказательство. Пусть $\varphi_1, \varphi_2 \in \operatorname{Ann} A$, $\lambda_1, \lambda_2 \in F$. Тогда для любого $v \in S$ выполняется равенство

$$(\lambda_1\varphi_1 + \lambda_2\varphi_2)(v) = \lambda_1\varphi_1(v) + \lambda_2\varphi_2(v) = 0,$$

то есть $\lambda_1\varphi_1 + \lambda_2\varphi_2 \in \operatorname{Ann} S$. Поэтому $\operatorname{Ann} S$ — подпространство.

Если $v \in L$, то по определению аннулятора $\varphi(v) = 0$ для всех $\varphi \in \operatorname{Ann} L$. Но $v^{**}(\varphi) = \varphi(v)$, то есть $v^{**} \in \operatorname{Ann}(\operatorname{Ann} L)$. Значит, $L \subseteq \operatorname{Ann}(\operatorname{Ann} L)$.

Докажем обратное включение $\operatorname{Ann}(\operatorname{Ann} L) \subseteq L$.

Пусть $v^{**} \in \operatorname{Ann}(\operatorname{Ann} L)$. Это означает, что $v^{**}(\varphi) = 0$ для всех $\varphi \in \operatorname{Ann} L$, то есть $\varphi(v) = 0$ при всех $\varphi \in \operatorname{Ann} L$. Нужно доказать, что отсюда следует $v \in L$.

Предположим обратное.

Выберем базис $b_1, \ldots, b_n$ пространства V, частью которого является базис b_1, $\ldots$, b_d подпространства L. Такой базис существует по утверждению 8.34. Обозначим через δ_j двойственный базис в пространстве V^*.

Заметим, что $\delta_j(b_i) = 0$ при $j > d$, $i \leqslant d$. Поэтому $\delta_j \in \operatorname{Ann} L$ при $j > d$.

Вектор v не принадлежит L тогда и только тогда, когда среди коэффициентов v_j разложения вектора по базису есть не равный нулю коэффициент при $j > d$. Тогда $\delta_j(v) \neq 0$, то есть не выполняется условие $\varphi(v) = 0$ при всех $\varphi \in \operatorname{Ann} L$. □

Замечание 8.53. Соответствие между пространством и двойственным ему аннулятором *антимонотонно*: если $L_1 \subseteq L_2$, то $\operatorname{Ann} L_1 \supseteq \operatorname{Ann} L_2$ (всякий функционал, обращающийся в 0 на подпространстве, заведомо обращается в 0 на меньшем подпространстве).

Более того, размерности подпространства и аннулятора связаны простым соотношением.

Лемма 8.54. *Пусть L — подпространство конечномерного пространства V. Тогда* $\dim L + \dim \operatorname{Ann} L = \dim V$.

Доказательство. Как и выше, выберем базис $b_1, \ldots, b_n$ пространства V, частью которого является базис b_1, $\ldots$, b_d подпространства L. Такой базис существует по утверждению 8.34. Обозначим через δ_j двойственный базис в пространстве V^*.

Выше уже было доказано, что $\{\delta_{d+1}, \ldots, \delta_n\} \subseteq \operatorname{Ann} L$ (проверьте это ещё раз прямым вычислением на бумажке). С другой стороны, $\delta_i \notin \operatorname{Ann} L$ при $i \leqslant d$: ведь $\delta_i(b_i) = 1 \neq 0$, а $b_i \in L$ в этом случае.

Поэтому $\operatorname{Ann} L$ в точности совпадает с подпространством, порождённым системой векторов $(\delta_{d+1}, \ldots, \delta_n)$ и эта система векторов является базисом $\operatorname{Ann} L$. Так как $d = \dim L$, $n = \dim V$, а теперь ещё установлено, что $\dim \operatorname{Ann} L = n - d$, получаем равенство леммы. □

По линейному отображению $A\colon V \to U$ векторных пространств определим *сопряжённое отображение* $A^*\colon U^* \to V^*$ сопряжённых пространств (обратите внимание, что порядок пространств другой) правилом

$$A^*(\varphi)(v) = \varphi(A(v)), \quad \varphi \in U^*,\ v \in V.$$

Правая часть этого равенства является линейным функционалом на V, так как

$$\varphi(A(\lambda_1 v_1 + \lambda_2 v_2)) = \varphi(\lambda_1 A(v_1) + \lambda_2 A(v_2)) = \lambda_1 \varphi(A(v_1)) + \lambda_2 \varphi(A(v_2))$$

(мы несколько раз пользовались линейностью A и φ).

Утверждение 8.55. *Сопряжённое отображение линейно.*

Доказательство. Для любого вектора v в силу линейности выполняются равенства

$$\begin{aligned} A^*(\lambda_1\varphi_1 + \lambda_2\varphi_2)(v) = \\ = (\lambda_1\varphi_1 + \lambda_2\varphi_2)(A(v)) = \lambda_1\varphi_1(A(v)) + \lambda_2\varphi_2(A(v)) = \\ = \lambda_1 A^*(\varphi_1)(v) + \lambda_2 A^*(\varphi_2)(v), \end{aligned}$$

то есть $A^*(\lambda_1\varphi_1 + \lambda_2\varphi_2) = \lambda_1 A^*(\varphi_1) + \lambda_2 A^*(\varphi_2)$, что и означает линейность сопряжённого отображения. □

Упражнение 8.56. Проверьте, что если задать линейное отображение матрицей в двух фиксированных базисах, то сопряжённое отображение в двойственных базисах задаётся транспонированной матрицей.

Между размерностями ядер и образов сопряжённых отображений имеется связь.

Лемма 8.57. *Пусть $A\colon V \to U$ — линейное отображение векторных пространств. Тогда*

$$\dim \operatorname{Ker} A + \dim \operatorname{Im} A^* = \dim V, \quad \dim \operatorname{Im} A + \dim \operatorname{Ker} A^* = \dim U$$

и $\operatorname{Im} A^* = \operatorname{Ann} \operatorname{Ker} A$, $\operatorname{Im} A = \operatorname{Ann} \operatorname{Ker} A^*$.

Доказательство. Проверим, что $\operatorname{Im} A^* \subseteq \operatorname{Ann} \operatorname{Ker} A$. Действительно, если $v \in \operatorname{Ker} A$, то $A^*(\varphi)(v) = \varphi(A(v)) = \varphi(0) = 0$ для любого $\varphi \in U^*$.

Из симметрии в определении сопряжённого отображения получаем аналогичное включение $\operatorname{Im} A \subseteq \operatorname{Ann} \operatorname{Ker} A^*$ (проверьте прямым вычислением!).

Включение подпространств влечёт неравенство на размерности этих подпространств. Используя теорему 8.39 и лемму 8.54, получаем такую цепочку неравенств:

$$\begin{aligned} \dim \operatorname{Im} A^* \leqslant \dim \operatorname{Ann} \operatorname{Ker} A = \dim V - \dim \operatorname{Ker} A = \dim \operatorname{Im} A \leqslant \\ \leqslant \dim \operatorname{Ann} \operatorname{Ker} A^* = \dim U - \dim \operatorname{Ker} A^* = \dim \operatorname{Im} A^*. \end{aligned}$$

Цепочка начинается и заканчивается одним и тем же числом. Поэтому все неравенства внутри цепочки обращаются в равенства. Из них следуют равенства в лемме (при равенства размерностей включения $\operatorname{Im} A^* \subseteq \operatorname{Ann} \operatorname{Ker} A$, $\operatorname{Im} A \subseteq \operatorname{Ann} \operatorname{Ker} A^*$ обращаются в равенства). □

Пример 8.58 (продолжение примера 8.35)**.** Как и в том примере, $G(V, E)$ — неориентированный граф без петель, но, возможно, с параллельными рёбрами. В примере 8.35 мы рассматривали векторное пространство $\mathbb{F}_2^E$ — координатное пространство, координаты в котором индексированы рёбрами графа. Теперь рассмотрим также и пространство $\mathbb{F}_2^V$, координаты в котором индексированы вершинами графа.

Пусть линейное отображение $A\colon \mathbb{F}_2^V \to \mathbb{F}_2^E$ задаётся правилом

$$A(\mathbb{1}_S) = \sum_{v \in S} g_v,$$

здесь $\mathbb{1}_S$, как и раньше, — вектор, у которого единичные координаты образуют множество S, а нулевые — дополнение к S.

Тогда, как нетрудно видеть, $\operatorname{Im} A = \operatorname{Cut}(G)$.

Как устроено сопряжённое отображение $A^*\colon (\mathbb{F}_2^E)^* \to (\mathbb{F}_2^V)^*$?

Элементы базиса, двойственного к базису E, обозначим ε_e, $e \in E$; элементы базиса, двойственного V, обозначим ν_v, $v \in V$. То есть

$$\varepsilon_e(e') = \begin{cases} 1, & \text{если } e = e', \\ 0, & \text{если } e \neq e', \end{cases} \qquad \nu_v(v') = \begin{cases} 1, & \text{если } v = v', \\ 0, & \text{если } v \neq v'. \end{cases}$$

По определению сопряжённого отображения $A^*(\varepsilon_e)(v) = \varepsilon_e(A(v))$. Это выражение равно 1 тогда и только тогда, когда v — один из концов ребра e. Для ребра $e = \{u, v\}$ получаем $A^*(\varepsilon_e) = \nu_u + \nu_v$.

Обозначим $\mathbb{1}_X^* = \sum_{e \in X} \varepsilon_e$. Этот вектор принадлежит ядру A^* тогда и только тогда, когда в каждой вершине графа сходится чётное количество рёбер из множества X. Такое множество рёбер называется *циклом*.[8)] Поэтому пространство $\operatorname{Ker} A^*$ называют *пространством циклов* и обозначают $\operatorname{Cyc}(G)$.

Из леммы 8.57 мы знаем, что $\operatorname{Im} A = \operatorname{Ann} \operatorname{Ker} A^*$, то есть $\operatorname{Cut}(G) = \operatorname{Ann} \operatorname{Cyc}(G)$. Видим, что пространства циклов и разрезов двойственны друг другу.

Отсюда, в частности, следует формула для размерности пространства циклов связного графа G: $\dim \operatorname{Cyc}(G) = |E| - \dim \operatorname{Cut}(G) = |E| - |V| + 1$. □

Контрольный вопрос 8.59. Докажите, что пересечение цикла и разреза состоит из чётного количества рёбер.

8.5 Детерминант

Детерминант (или определитель) — это функция от квадратной матрицы. Мы будем считать, что матричные элементы принадлежат некоторому полю F.

Определение 8.60. Многочлен, задаваемый формулой

$$\det A = \sum_{\pi \in S_n} \operatorname{sgn} \pi \prod_{s=1}^{n} a_{\pi(s)s}, \tag{8.5}$$

[8)] Это одно из трёх различных определений цикла в теории графов. Терминология в этой области прискорбно запутана.

называется *детерминантом матрицы* A. Здесь $\operatorname{sgn}\pi$ — знак перестановки π: он равен $+1$ для чётных перестановок и -1 для нечётных.

Мы приведём несколько основных свойств детерминанта, которые используются в этом курсе. Для формулировок иногда бывает удобно представлять квадратную матрицу A порядка n как набор столбцов, то есть последовательность $a_1, a_2, \ldots, a_n$ векторов координатного пространства F^n.

Лемма 8.61. *Функция* $\det A = \det(a_1, \ldots, a_n)$ *(детерминант матрицы* A*) удовлетворяет следующим свойствам.*

– *Линейность: для любого столбца* j *выполняется равенство*

$$\det(a_1, \ldots, a_{j-1}, a'_j + a''_j, a_{j+1}, \ldots, a_n) = \\ \det(a_1, \ldots, a_{j-1}, a'_j, a_{j+1}, \ldots, a_n) + \det(a_1, \ldots, a_{j-1}, a''_j, a_{j+1}, \ldots, a_n).$$

– *Однородность: для любого столбца* j *и любого* $\lambda \in F$ *выполняется равенство*

$$\det(a_1, \ldots, a_{j-1}, \lambda a_j, a_{j+1}, \ldots, a_n) = \lambda \det(a_1, \ldots, a_{j-1}, a_j, a_{j+1}, \ldots, a_n).$$

– *Антисимметричность: для любых* $i < j$ *выполняется равенство*

$$\det(a_1, \ldots, a_{i-1}, a_i, a_{i+1}, \ldots, a_{j-1}, a_j, a_{j+1}, \ldots, a_n) = \\ = -\det(a_1, \ldots, a_{i-1}, a_j, a_{i+1}, \ldots, a_{j-1}, a_i, a_{j+1}, \ldots, a_n)$$

(детерминант меняет знак при перестановке любых двух столбцов матрицы).

– *Нормировка: детерминант единичной матрицы равен* 1.

Доказательство. Линейность и однородность очевидны, так как в каждом слагаемом есть ровно один элемент столбца j:

$$\sum_{\pi \in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} a_{\pi(s)s} = \sum_{i=1}^{n} a_{ij} P_i^{(j)}(A), \tag{8.6}$$

многочлен $P_i^{(j)}(A)$ получается из мономов (8.5) группировкой слагаемых, в которые входит множитель a_{ij}.

Условие нормировки выполняется потому, что для единичной матрицы в сумме (8.5) отлично от нуля только одно слагаемое (отвечающее тождественной перестановке), это слагаемое равно произведению 1.

Осталось проверить условие антисимметричности. Пусть в матрице A переставили j-й и k-й столбцы и получили матрицу B, матричные элементы которой выражаются через матричные элементы исходной матрицы как

$$b_{rs} = \begin{cases} a_{rs}, & \text{если } s \neq j,\ s \neq k, \\ a_{rk}, & \text{если } s = j, \\ a_{rj}, & \text{если } s = k. \end{cases}$$

Поэтому для любой перестановки π выполняется равенство

$$\prod_{s=1}^{n} b_{\pi(s)s} = \prod_{s=1}^{n} a_{\sigma(s)s},$$

где $\sigma = \pi \circ (j\ k)$. Отсюда получаем

$$\sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} b_{\pi(s)s} = \sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} a_{(\pi\circ(j\ k))(s)s} =$$

$$= \sum_{\sigma\in S_n} \operatorname{sgn}(\sigma\circ(j\ k) \prod_{s=1}^{n} a_{\sigma(s)s} = -\sum_{\sigma\in S_n} \operatorname{sgn}\sigma \prod_{s=1}^{n} a_{\sigma(s)s},$$

так как умножение на транспозицию меняет знак перестановки. Это и есть условие антисимметричности. □

Аналогичные свойства выполняются также, если рассматривать детерминант как функцию от строк матрицы. Достаточно проверить, что детерминант матрицы равен детерминанту транспонированной матрицы (при транспонировании столбцы становятся строками и наоборот).

Лемма 8.62. $\det A = \det A^T$.

Доказательство. Напомним, что знаки перестановки и обратной к ней совпадают (следствие 4.21). Перепишем формулу (8.5), заменяя суммирование по π на суммирование по π^{-1}:

$$\sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} a_{\pi(s)s} = \sum_{\pi\in S_n} \operatorname{sgn}\pi^{-1} \prod_{k=1}^{n} a_{k\pi^{-1}(k)} = \sum_{\sigma\in S_n} \operatorname{sgn}\sigma \prod_{k=1}^{n} a_{k\sigma(k)}.$$

Правая часть этого равенства совпадает с $\det A^T$. □

Многочлены $P_i^j(A)$, возникающие в разложении детерминанта по j-му столбцу матрицы, выражаются через детерминанты матриц меньших размеров. Обозначим через $A_i^{(j)}$ матрицу, которая получается из матрицы A вычёркиванием i-й строки и j-го столбца.

Лемма 8.63. $\det A = \sum_{i=1}^{n} a_{ij}(-1)^{i+j} \det A_i^{(j)}$.

Выполняется также аналогичное разложение по строкам.

Доказательство. Выделим в формуле (8.5) детерминанта множители a_{ij}, отвечающие j-му столбцу:

$$\det A = \sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} a_{\pi(s)s} = \sum_{\pi\in S_n} a_{\pi(j)j} \operatorname{sgn}\pi \prod_{s\neq j} a_{\pi(s)s} = \sum_{i=1}^{n} a_{ij} \sum_{\pi:\pi(j)=i} \operatorname{sgn}\pi \prod_{s\neq j} a_{\pi(s)s}.$$

Ограничение π на столбцы без j-го и строки без i-й задаёт однозначно определяет перестановку π_{ij} на множестве строк и столбцов матрицы $A_i^{(j)}$, если нумеровать их числами от 1 до $n-1$, пропуская j в столбцах и i в строках.

Докажем, что $\operatorname{sgn}\pi_{ij} = (-1)^{i+j}\operatorname{sgn}\pi$, откуда из формулы для детерминанта следует равенство леммы.

Продолжим π_{ij} до перестановки на множестве $\{1,\dots,n\}$, полагая $\pi'_{ij}(k) = \pi_{ij}(k)$, если $k < n$; $\pi'_{ij}(n) = n$. Тогда перестановка π представляется как композиция

$$\pi = (i\ (i+1)\ \dots\ n) \circ \pi'_{ij} \circ (n\ (n-1)\ \dots\ j).$$

Действительно, перестановка $(n\ \ (n-1)\ \ \dots\ \ j)$ переводит j в n, а номера всех чисел, бóльших n, уменьшаются на 1. Это соответствует нумерации столбцов с пропущенным j-м столбцом. Перестановка π'_{ij} переводит столбцы в их π-образы, если нумеровать строки с пропущенной i-строкой. Применение третьей перестановки $(i\ \ (i+1)\ \ \dots\ \ n)$ восстанавливает исходную нумерацию строк и переводит n в i. А композиция двух предыдущих перестановок переводит j в n. Так что в точности получается перестановка π.

Легко проверить, что

$$\operatorname{sgn}\pi'_{ij} = \operatorname{sgn}\pi_{ij};\quad \operatorname{sgn}(n\ (n-1)\ \dots\ j) = (-1)^{n-j};\quad \operatorname{sgn}(i\ (i+1)\ \dots\ n) = (-1)^{n-i}$$

(выполните эту проверку самостоятельно, это очень важное и полезное упражнение).

Так как знак перестановки — это гомоморфизм в группу $\{\pm1,\cdot\}$, то получаем $\operatorname{sgn}\pi = (-1)^{n-i}\operatorname{sgn}\pi_{ij}(-1)^{n-j} = (-1)^{2n-i-j}\operatorname{sgn}\pi_{ij}$. Чётность $2n-i-j$ и $i+j$ одинакова. Поэтому последнее выражение равно $(-1)^{i+j}\operatorname{sgn}\pi$, что и требовалось доказать. □

Следствие 8.64. *Если в матрице есть столбец (или строка), состоящий из нулевых элементов, то определитель матрицы равен нулю.*

Доказательство. Пусть j-й столбец состоит только из нулевых элементов. Тогда в разложении по этому столбцу (лемма 8.63) все слагаемые нулевые. □

Из условия антисимметричности легко следует, что детерминант матрицы, в которой есть два одинаковых столбца, равен нулю:

$$\det(\dots,a,\dots,a,\dots) = -\det(\dots,a,\dots,a,\dots) = 0$$

(на месте многоточий стоят одинаковые последовательности столбцов в первом и втором детерминанте).

Из этого наблюдения следует важное свойство детерминанта.

Лемма 8.65. *Если к столбцу матрицы прибавить линейную комбинацию других столбцов, детерминант матрицы не изменится. (Аналогично для строк.)*

Доказательство. Пусть к j-му столбцу добавляется линейная комбинация остальных столбцов:

$$a'_j = a_j + \sum_{k\neq j} \lambda_k a_k.$$

Из свойств линейности и однородности получаем

$$\det(a_1,\dots,a'_j,\dots,a_n)=\det(a_1,\dots,a_j,\dots,a_n)+\sum_{k\neq j}\lambda_k\det(a_1,\dots,a_k,\dots,a_n).$$

В правой части помимо детерминанта исходной матрицы стоит линейная комбинация детерминантов матриц, у которых хотя бы два столбца совпадают. Поэтому все слагаемые равны 0. □

Теперь приведём один из известных примеров вычисления детерминанта, который потребуется нам в дальнейшем.

Пример 8.66 (определитель Вандермонда)**.** Обозначим через $W(x_1,\dots,x_n)$ матрицу порядка n с матричными элементами $w_{ij}=x_i^{j-1}$.

Докажем, что

$$\det W(x_1,\dots,x_n)=\prod_{1\leqslant i<j\leqslant n}(x_j-x_i) \tag{8.7}$$

индукцией по порядку матрицы.

База индукции: $W(x)=x^{1-1}=1$ (произведение пустого количества сомножителей равно 1).

Индуктивный переход. В силу леммы 8.65 детерминант матрицы не изменяется, если к столбцу прибавить кратное другого столбца. Выполним с матрицей $W(x_1,\dots,x_n)$ следующие преобразования: прибавим ко второму столбцу первый, умноженный на x_1, к третьему — второй, умноженный на x_1 и т.д. Получим матрицу

$$A=\begin{pmatrix}1 & 0 & 0 & \dots & 0\\ 1 & (x_2-x_1) & x_2(x_2-x_1) & \dots & x_2^{n-1}(x_2-x_1)\\ \dots & \dots & \dots & \dots & \dots\\ 1 & (x_n-x_1) & x_n(x_n-x_1) & \dots & x_n^{n-1}(x_n-x_1)\end{pmatrix},$$

детерминант которой равен детерминанту $W(x_1,\dots,x_n)$. Разлагая по первой строке, в которой лишь один ненулевой элемент, получаем

$$\det W(x_1,\dots,x_n)=\det A=\det\begin{pmatrix}(x_2-x_1) & x_2(x_2-x_1) & \dots & x_2^{n-1}(x_2-x_1)\\ \dots & \dots & \dots & \dots\\ (x_n-x_1) & x_n(x_n-x_1) & \dots & x_n^{n-1}(x_n-x_1)\end{pmatrix}.$$

Используя однородность детерминанта по строкам, получаем

$$\det W(x_1,\dots,x_n)=\prod_{i=2}^{n}(x_i-x_1)\det W(x_2,\dots,x_n).$$

Теперь осталось применить индуктивное предположение для $\det W(x_2,\dots,x_n)$ и получить формулу (8.7) для матрицы $W(x_1,\dots,x_n)$. □

Одним из важнейших свойств детерминанта, не столь очевидных непосредственно из определения, является мультипликативность.

Лемма 8.67. $\det(BC)=\det B\cdot\det C$.

Доказательство. Обозначим $A = BC$ и запишем детерминант матрицы A, используя определение произведения матриц:

$$\det A = \sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} a_{\pi(s)s} = \sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n}\sum_{k=1}^{n} b_{\pi(s)k}c_{ks} =$$

$$= \sum_{\pi\in S_n} \operatorname{sgn}\pi \sum_{k_1,\ldots,k_n} \prod_{s=1}^{n} b_{\pi(s)k_s}c_{k_s s} = \sum_{k_1,\ldots,k_n} \sum_{\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} b_{\pi(s)k_s}c_{k_s s}$$

(в предпоследнем равенства переставлены произведение и сумма).

В последней формуле внешняя сумма производится по n^n слагаемым — всем возможным функциям из $\{1,\ldots,n\}$ в $\{1,\ldots,n\}$. Однако большинство слагаемых в этой сумме нулевые. Более точно, не равными нулю могут оказаться только слагаемые отвечающие перестановкам.

Действительно, пусть $k_j = k_r$. Тогда произведения под внутренней суммой для перестановок π и $\sigma = \pi\circ(j\ r)$ одинаковы:

$$\prod_{s=1}^{n} b_{\sigma(s)k_s}c_{k_s s} = b_{\pi(r)k_j}c_{k_j r}b_{\pi(j)k_r}c_{k_r j} \prod_{s\neq j,\ s\neq r} b_{\pi(s)k_s}c_{k_s s} = \prod_{s=1}^{n} b_{\pi(s)k_s}c_{k_s s}.$$

А знаки у перестановок π и $\sigma = \pi\circ(j\ r)$ противоположные.

Итак, остаются лишь те слагаемые, для которых функция $s\mapsto k_s$ биективна, то есть является перестановкой. Перепишем последнее полученное выше выражение для $\det A$ как

$$\det A = \sum_{\sigma,\pi\in S_n} \operatorname{sgn}\pi \prod_{s=1}^{n} b_{\pi(s)\sigma(s)}c_{\sigma(s)s} = \sum_{\pi_1,\pi_2\in S_n} \operatorname{sgn}(\pi_1\circ\pi_2) \prod_{s=1}^{n} b_{\pi_1(s)s}c_{\pi_2(s)s},$$

где $\pi_1 = \pi\circ\sigma^{-1}$, $\pi_2 = \sigma$, и потому $\pi = \pi_1\circ\pi_2$.

Поскольку $\operatorname{sgn}(\pi_1\circ\pi_2) = \operatorname{sgn}(\pi_1)\cdot\operatorname{sgn}(\pi_2)$, последняя сумма раскладывается в произведение двух сумм

$$\det A = \sum_{\pi_1,\pi_2\in S_n} \operatorname{sgn}(\pi_1\circ\pi_2) \prod_{s=1}^{n} b_{\pi_1(s)s}c_{\pi_2(s)s} =$$

$$= \left(\sum_{\pi_1\in S_n} \operatorname{sgn}(\pi_1) \prod_{s=1}^{n} b_{\pi_1(s)s}\right) \cdot \left(\sum_{\pi_2\in S_n} \operatorname{sgn}(\pi_2) \prod_{s=1}^{n} c_{\pi_2(s)s}\right),$$

которые равны $\det B$ и $\det C$ соответственно. □

Квадратная матрица порядка n задаёт линейное отображение координатного пространства F^n в себя. Это отображение является линейным изоморфизмом (то есть биективно) тогда и только тогда, когда столбцы матрицы (то есть образы базисных векторов) линейно независимы. Действительно, линейная зависимость между

столбцами

$$\sum_{j=1}^{n} \lambda_j a_j = 0$$

означает, что $A(\lambda_1, \ldots, \lambda_n) = 0$ и наоборот. Поэтому ядро отображения A нулевое тогда и только тогда, когда нет нетривиальных линейных зависимостей между столбцами A.

Из этого наблюдения получается критерий равенства нулю детерминанта матрицы.

Лемма 8.68. $\det A = 0$ *тогда и только тогда, когда столбцы матрицы* A *линейно зависимы.*

Доказательство. Пусть $a_1, \ldots, a_n$ — столбцы матрицы A, а

$$\sum_{j=1}^{n} \lambda_j a_j$$

нетривиальная линейная комбинация столбцов, равная нулю.

Пусть $\lambda_k \neq 0$. Тогда

$$a_k = -\sum_{j \neq k} \frac{\lambda_j}{\lambda_k} a_j.$$

Прибавляя к k-му столбцу линейную комбинацию остальных столбцов с коэффициентами λ_j/λ_k, получаем матрицу с тем же детерминантом, что и A (лемма 8.65), и с нулевым столбцом. Детерминант этой матрицы (как и детерминант исходной) равен нулю по следствию 8.64.

Пусть теперь столбцы матрицы A линейно независимы. Это означает, что у матрицы A есть обратная. Из соотношения $AA^{-1} = I$ и леммы 8.67 получаем

$$1 = \det I = \det(AA^{-1}) = \det A \cdot \det A^{-1},$$

то есть $\det A \neq 0$. □

9 Гомоморфизмы колец и кольца вычетов

9.1 Определения гомоморфизмов и изоморфизмов колец

Определение гомоморфизма колец аналогично определению гомоморфизма групп. Поскольку в кольце есть две операции, гомоморфизм обязан сохранять обе.

Определение 9.1. *Гомоморфизмом* кольца кольцо R с операциями $+$, $\cdot$ в кольцо R' с операциями $\oplus$, $\otimes$ называется такое отображение $\varphi\colon R \to R'$, что:

$$\varphi(r_1 + r_2) = \varphi(r_1) \oplus \varphi(r_2),$$
$$\varphi(r_1 \cdot r_2) = \varphi(r_1) \otimes \varphi(r_2).$$

Контрольный вопрос 9.2. Проверьте из определения, что композиция гомоморфизмов $\varphi\colon R_1 \to R_2$ и $\psi\colon R_2 \to R_3$ является гомоморфизмом $\psi \circ \varphi\colon R_1 \to R_3$.

Пример 9.3. Рассмотрим отображение $\delta\colon \mathbb{Z} \to 2\mathbb{Z}$ из кольца целых чисел в кольцо чётных чисел, задаваемое умножение на 2: $\delta(x) = 2x$. Является ли оно гомоморфизмом? Первое условие из определения гомоморфизма выполнено

$$\delta(x + y) = 2(x + y) = 2x + 2y = \delta(x) + \delta(y).$$

Другими словами, δ — гомоморфизм аддитивных групп колец $\mathbb{Z}$ и $2\mathbb{Z}$.

Но второе условие гомоморфизма колец не выполнено:

$$\delta(2 \cdot 2) = 8 \neq 4 \cdot 4 = \delta(2) \cdot \delta(2).$$ □

Пример 9.4. Рассмотрим отображение $\varphi_n\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$, которое ставит в соответствие целому числу $x \in \mathbb{Z}$ класс вычетов по модулю n, содержащий число x (напомним, класс вычетов — это множество тех целых чисел, которые имеют по модулю n тот же остаток, что и x).

Отображение $\varphi_n\colon \mathbb{Z} \to \mathbb{Z}/n\mathbb{Z}$ является гомоморфизмом колец.

По сути это было доказано, когда определялось кольцо $\mathbb{Z}/n\mathbb{Z}$ вычетов по модулю n. Напомним, что через $[x]$ мы обозначаем класс вычетов, содержащий x. Операции в кольце $\mathbb{Z}/n\mathbb{Z}$ были определены как

$$[x] + [y] = [x + y],$$
$$[x] \cdot [y] = [xy].$$

Эти равенства как раз и есть свойства сохранения операций при отображении φ_n. □

Контрольный вопрос 9.5. Непосредственно из определения φ_n проверьте включения (определите, какие истинны, а какие — нет):

$17 \in \varphi_5(-3)$, $16 \in \varphi_6(-2)$, $15 \in \varphi_7(-1)$,
$19 \in \varphi_8(9)$, $20 \in \varphi_9(11)$, $21 \in \varphi_{10}(-11)$.

Ниже мы обобщим пример 9.4 и он станет частным случаем общей конструкции, аналогичной конструкции факторгрупп.

Пример 9.6. Пусть задан элемент $a \in R$ коммутативного кольца. Многочлену $f \in R[x]$ мы сопоставили его значение в точке a. Отображение (*гомоморфизм значения*)

$$\mathrm{Ev}_a \colon f \mapsto f(a)$$

является гомоморфизмом колец. Требуемые в определении гомоморфизма равенства уже были проверены, когда определялось кольцо многочленов. □

Гомоморфизм значения можно расширить. Если кольцо коэффициентов Q вложено в бóльшее кольцо R, то значение многочлена $f \in Q[x]$ можно определить и в точке $a \in R$:

$$f(a) = \sum_{i=0}^{d} f_i a^i \in R.$$

Контрольный вопрос 9.7. Проверьте, что и в этом случае отображение значения является гомоморфизмом.

Аналогично случаю гомоморфизмов групп для гомоморфизмов колец выполняется следующее простое свойство.

Лемма 9.8. *Гомоморфный образ кольца есть кольцо.*

Доказательство. По сложению все рассуждения повторяются дословно — ведь гомоморфизм колец является также и гомоморфизмом их аддитивных групп.

Ассоциативность умножения проверяется так же, как и ассоциативность сложения.

Проверка дистрибутивности осуществляется прямым вычислением:

$$\begin{aligned}\varphi(a) \otimes (\varphi(b) \oplus \varphi(c)) &= \varphi(a) \otimes \varphi(b+c) = \\ &= \varphi(a \cdot (b+c)) = \varphi(a \cdot b + a \cdot c) = \varphi(a \cdot b) \oplus \varphi(a \cdot c) = \\ &= \varphi(a) \otimes \varphi(b) \oplus \varphi(a) \otimes \varphi(c).\end{aligned}$$

□

Поскольку гомоморфизм колец является также и гомоморфизмом их аддитивных групп (первое условие в определении гомоморфизма), гомоморфный образ нуля (т.е. нейтрального элемента аддитивной группы) обязательно равен нулю.

Если в кольце есть единица (нейтральный элемент по умножению), то гомоморфный образ единицы уже не обязан равняться единице в кольце-образе. Например, в кольце-образе может просто не быть нейтрального элемента по умножению.

Пример 9.9. Рассмотрим кольцо $\mathbb{Z} \oplus (2\mathbb{Z})$. Элементы этого кольца — пары (целое число, чётное число), операции выполняются покомпонентно. Единицы в этом кольце нет: условие $(a, b) \cdot (x, y) = (x, y)$ равносильно двум равенствам $ax = x$ и $by = y$. Второе из них выполняется для чётных чисел только при $y = 0$.

С другой стороны, отображение $x \mapsto (x, 0)$ является гомоморфизмом $\mathbb{Z} \to \mathbb{Z} \oplus (2\mathbb{Z})$.

Контрольный вопрос 9.10. Проверьте последнее утверждение.

При таком гомоморфизме 1 переходит в $(1,0)$, то есть не в нейтральный элемент кольца-образа. □

Но даже если в кольце-образе есть единичный элемент, гомоморфный образ единицы кольца-прообраза не обязан с ним совпадать.

Пример 9.11. Рассмотрим отображение $\mathbb{Z} \to \mathbb{Z} \oplus \mathbb{Z}$, задаваемое как $x \mapsto (x, 0)$. Это гомоморфизм колец, как и в предыдущем примере. Но образ 1, то есть $(1,0)$ не совпадает с единичным элементом в $\mathbb{Z} \oplus \mathbb{Z}$ (это $(1,1)$). □

Эти примеры противоречат интуитивному пониманию гомоморфизма, как отображения, сохраняющего алгебраические свойства. Некоторые алгебраические свойства, как видим, могут теряться при гомоморфизме колец. Гораздо лучше дело обстоит с сюръективными гомоморфизмами (и позже мы выясним причину этого). Пока предлагаем выполнить несложное упражнение.

Контрольный вопрос 9.12. Докажите, что если $\varphi\colon R_1 \to R_2$ — сюръективный гомоморфизм и кольцо R_1 содержит 1 (нейтральный элемент по умножению), то $\varphi(1)$ — нейтральный элемент по умножению в кольце R_2.

Определение 9.13. *Изоморфизм* колец — это взаимно однозначный гомоморфизм. Будем обозначать изоморфизм как $R_1 \cong R_2$.

Так же, как и в случае групп, изоморфные кольца «одинаковы» с алгебраической точки зрения.

Пример 9.14. Рассмотрим важнейший для комбинаторики и теоретической информатики пример двух изоморфных колец.

Первое кольцо P_n состоит из всех подмножеств n-элементного множества (для удобства далее предполагаем, что это множество $\{1, 2, \ldots, n\}$). Операции в этом кольце такие: сложение — это симметрическая разность множеств, умножение — пересечение множеств. (Проверку того, что эти операции удовлетворяют свойствам кольца, пока отложим.)

Второе кольцо — это кольцо функций $\{0,1\}^n$ из n-элементного множества в множество $\{0,1\}$. Такая функция задаётся таблицей значений, то есть двоичной строкой из нулей и единиц. Скажем, строка $(1, 0, \ldots, 0)$ задаёт функцию, которая равна 1 в точке 1, а в точках 2, 3, …, n равна 0. Операции в этом кольце — это поточечные сложение и умножение по модулю 2.

Другими словами, кольцо $\{0,1\}^n$ — это прямая сумма колец вычетов по модулю 2:

$$\underbrace{(\mathbb{Z}/2\mathbb{Z}) \oplus (\mathbb{Z}/2\mathbb{Z}) \oplus \cdots \oplus (\mathbb{Z}/2\mathbb{Z})}_{n \text{ копий кольца } \mathbb{Z}/2\mathbb{Z}}.$$

Мы уже проверяли, что прямая сумма колец является кольцом.

Теперь установим биекцию между P_n и $\{0,1\}^n$ по естественному правилу: множеству S сопоставим его индикаторную функцию

$$\chi_S(x) = \begin{cases} 1, & \text{если } x \in S, \\ 0, & \text{если } x \notin S. \end{cases}$$

Проверим, что это соответствие сохраняет операции.

Действительно, $\chi_{S\triangle T}(x) = 1$ равносильному тому, что $x \in S \triangle T$. То есть, либо $x \in S$ и $x \notin T$, либо $x \notin S$ и $x \in T$. Первое условие равносильно тому, что $\chi_S(x) = 1$, $\chi_T(x) = 0$; а второе — тому, что $\chi_S(x) = 0$, $\chi_T(x) = 1$. Но это равносильно тому, что $\chi_S(x) + \chi_T(x) \equiv 1 \pmod 2$. Таким образом,

$$\chi_{S\triangle T}(x) \equiv \chi_S(x) + \chi_T(x) \pmod 2,$$

то есть соответствие сохраняет «сложение» (в случае первого кольца это симметрическая разность множеств).

Аналогично проверяется сохранение умножения.

Контрольный вопрос 9.15. Проверьте, что $\chi_{S\cap T}(x) = \chi_S(x) \cdot \chi_T(x)$ (это равенство выполняется даже не по модулю 2, а просто в целых числах).

После того, как установлен изоморфизм, доказывать, что P_n является кольцом, уже не нужно: это следует из того, что $\{0,1\}^n$ является кольцом. □

Пример 9.16. Рассмотрим образы R_0, $R_{\sqrt{2}}$ и R_ϕ кольца $\mathbb{Q}[x]$ многочленов с рациональными коэффициентами при гомоморфизмах значения в точках 0, $\sqrt{2}$ и $\phi = (1+\sqrt{5})/2$ соответственно (напомним, что последнее число называется *золотым сечением*).

Изоморфны эти кольца или нет?

Во-первых, легко увидеть изоморфизм между R_0 и $\mathbb{Q}$. Действительно, значение многочлена f в точке 0 равно его свободному члену (т.е. коэффициенту при нулевой степени) f_0. Таким образом $R_0 = \mathrm{Ev}_0(\mathbb{Q}[x]) = \mathbb{Q}$.

С двумя другими кольцами сложнее. Это подкольца кольца действительных чисел и они не совпадают с рациональными числами.

Контрольный вопрос 9.17. Проверьте, что $\sqrt{2} \in R_{\sqrt{2}}$, $1 \in R_{\sqrt{2}}$.

Утверждение 9.18. *$R_0 \ncong R_{\sqrt{2}}$; $R_0 \ncong R_\phi$.*

Доказательство. Поскольку $\sqrt{2} \in R_{\sqrt{2}}$, $1 \in R_{\sqrt{2}}$ в кольце $R_{\sqrt{2}}$ уравнение $x^2 - 2 = x^2 - (1+1) = 0$ имеет решения. А в кольце $R_0 \cong \mathbb{Q}$ это уравнение решений не имеет. (Напомним, как это доказывается: если $(p/q)^2 = 2$, то $p^2 = 2q^2$; для показателей α и β, с которыми 2 входит в разложение на простые чисел p и q, получаем равенство $2\alpha = 2\beta + 1$, которое ложно для всех целых α и β, пришли к противоречию.)

Это и означает, что кольца R_0 и $R_{\sqrt{2}}$ неизоморфны: поскольку изоморфизм сохраняет операции, то образ $\sqrt{2}$ при изоморфизме с R_0 обязан дать рациональное число, квадрат которого равен 2.

Обратите внимание, что в предыдущем абзаце пропущено одно тонкое место. Давайте развернём это рассуждение, следуя определениям. Пусть $\varphi\colon R_{\sqrt{2}} \to R_0$ — изоморфизм. В первом кольце $(\sqrt{2})^2 = 1 + 1$. Значит, во втором кольце $\varphi(\sqrt{2})^2 = \varphi(1) + \varphi(1)$. Выше мы заменили $\varphi(1) + \varphi(1)$ на $1 + 1 = 2$. Вообще говоря, как мы видели, гомоморфный образ единицы не всегда единица. Однако здесь мы используем то, что изоморфизм является сюръективным отображением, а при сюръективном гомоморфизме единица переходит в единицу.

Аналогично проверяется и второе утверждение. Для этого заметим, что в кольце R_ϕ содержится $\sqrt{5}$: это образ многочлена $2x-1$. Поэтому уравнение $x^2-5=0$ имеет решения в кольце R_ϕ. Но в кольце R_0 оно решений не имеет. (Доказательство аналогично предыдущему, нужно только рассмотреть показатели 5, а не 2 в разложении числителя и знаменателя на простые множители.) □

Утверждение 9.19. $R_{\sqrt{2}} \ncong R_\phi$.

Доказательство. Рассуждаем аналогично предыдущему. Докажем, что в кольце $R_{\sqrt{2}}$ уравнение $x^2 - 5 = 0$ не имеет решений (а в кольце R_ϕ, как мы видели, имеет). Из этого следует неизоморфизм колец аналогично предыдущим рассуждениям.

Прежде всего заметим, что все элементы кольца $R_{\sqrt{2}}$ имеют вид $a+b\sqrt{2}$, $a, b \in \mathbb{Q}$. Действительно, прямо из определения следует, что они имеют вид

$$\sum_{i=0}^{d} a_i 2^{i/2}, \qquad a_i \in \mathbb{Q}.$$

Но положительные степени $\sqrt{2}$ либо целые (для чётных показателей), либо целые кратные $\sqrt{2}$ (для нечётных показателей). Суммируя рациональные слагаемые и слагаемые вида «рациональное число, умноженное на $\sqrt{2}$», получаем искомое представление.

Теперь предположим, что для каких-то рациональных a, b выполняется равенство

$$(a + b\sqrt{2})^2 = 5.$$

Тогда

$$\sqrt{2} = \frac{5 - a^2 - b^2}{2ab} \in \mathbb{Q},$$

а мы уже установили, что $\sqrt{2}$ не является рациональным числом. □

Итак, кольца R_0, $R_{\sqrt{2}}$ и R_ϕ попарно неизоморфны. □

Пример 9.20. Если F — поле с бесконечным числом элементов, то кольцо $F[x]$ изоморфно кольцу функций $F \to F$, которые представляются многочленами. Гомоморфизм поставляет многочлену функцию, которая представляется этим многочленом. По построению это сюръективный гомоморфизм. Инъективность следует из леммы о числе корней многочлена. Если $f_1 \neq f_2$, то $f_1 - f_2 \neq 0$ и количество точек, в которых совпадают функции, представленные многочленами f_1, f_2 не превосходит

максимальной степени этих многочленов. Поэтому функции, представленные этими многочленами, различны. □

Приведём ещё один полезный пример изоморфизма кольца многочленов на себя (как обычно, такие изоморфизмы называются *автоморфизмами*).

Пример 9.21. Выберем в коммутативном кольце R с единицей какой-нибудь элемент a. По этому элементу определим отображение $\varphi\colon R[x] \to R[x]$ «сдвига переменной», задаваемое правилом

$$f(x) \mapsto f(x+a).$$

Такое отображение очевидным образом определяется для функций, заданных многочленами. Но, как мы уже отмечали, нас интересуют сами кольца многочленов, а не только задаваемые ими функции. Поэтому нужно уточнить определение отображения φ.

Зададим образы многочленов нулевой степени и x:

$$\varphi\colon b \mapsto b, \quad \deg b = 0; \qquad \varphi\colon x \mapsto x + a.$$

Если потребовать, чтобы φ был гомоморфизмом $R[x] \to R[x]$ (такие гомоморфизмы обычно называют *эндоморфизмами*), то эти два условия определяют φ однозначно.

Действительно, из сохранения умножения получаем

$$\varphi(bx^n) = \varphi(b)\varphi(x) \cdot \varphi(x) \cdot \ldots \cdot \varphi(x) = b(x+a)^n.$$

Теперь из сохранения сложения получаем формулу для образа любого многочлена:

$$\varphi\left(\sum_{i=0}^{d} b_i x^i\right) = \sum_{i=0}^{d} b_i (x+a)^i.$$

Мы показали, что φ определён однозначно, если он является гомоморфизмом. Как всегда в таких случаях, нужно теперь проверить, что построенное отображение и впрямь гомоморфизм. Проверка состоит в рутинных вычислениях, использующих общие свойства колец и свойство коммутативности умножения:

$$\varphi(f+g) = \sum_i (f+g)_i (x+a)^i = \sum_i f_i (x+a)^i + \sum_i g_i (x+a)^i = \varphi(f) + \varphi(g),$$

$$\begin{aligned}
\varphi(fg) &= \sum_i (fg)_i (x+a)^i = \sum_{i=0}^{d} \sum_{j+k=i} f_j g_k (x+a)^i = \\
&= \sum_{j,k} f_j (x+a)^j g_k (x+a)^k = \left(\sum_j f_j (x+a)^j\right) \cdot \left(\sum_k g_k (x+a)^k\right) = \\
&= \varphi(f) \cdot \varphi(g).
\end{aligned}$$

Итак, эндоморфизм сдвига определён корректно. Осталось понять, почему это автоморфизм, то есть почему отображение φ взаимно однозначно.

Для этого определим обратное отображение:

$$\psi\left(\sum_{i=0}^{d} b_i x^i\right) = \sum_{i=0}^{d} b_i (x-a)^i.$$

Проверим корректность такого определения. Композиция этих отображений имеет вид

$$\varphi \circ \psi\left(\sum_{i=0}^{d} b_i x^i\right) = \sum_{i=0}^{d} b_i \varphi((x-a)^i).$$

Нужно проверить, что $\varphi \circ \psi(f) = f$. Как видно из предыдущей формулы, для этого достаточно проверить, что $\varphi((x-a)^i) = x^i$. Используя свойства гомоморфизма, получаем

$$\varphi((x-a)^i) = (\varphi(x-a))^i = (\varphi(x) - \varphi(a)))^i = ((x+a)-a))^i = x^i.$$

Аналогично проверяется, что $\psi \circ \varphi(f) = f$. По теореме об обратной функции это означает, что φ — биекция и $\psi = \varphi^{-1}$. □

9.2 Ядра гомоморфизмов и идеалы

Как и в случае гомоморфизмов групп, для гомоморфизмов колец определено ядро.

Определение 9.22. *Ядром* гомоморфизма $\varphi\colon R_1 \to R_2$ называется множество элементов R_1, отображающихся в 0 (нейтральный элемент по сложению кольца R_2):

$$\operatorname{Ker}\varphi = \{x \in R_1 : \varphi(x) = 0\}.$$

Поскольку гомоморфизм колец является также гомоморфизмом их аддитивных групп, ядро гомоморфизма является подгруппой аддитивной группы кольца-прообраза. Но поскольку в любом кольце выполняются равенства $0 \cdot x = x \cdot 0 = 0$, ядро гомоморфизма колец обладает ещё и дополнительным свойством «втягивания».

Утверждение 9.23. *Пусть $\varphi\colon R_1 \to R_2$ — гомоморфизм колец. Для любого $a \in R_1$ и любого $i \in \operatorname{Ker}\varphi$ выполняется*

$$ai \in \operatorname{Ker}\varphi, \quad ia \in \operatorname{Ker}\varphi.$$

Для свойства втягивания нужны два равенства потому, что умножение кольца не обязательно коммутативно.

Доказательство. Из определений гомоморфизма и ядра получаем равенства

$$\varphi(ai) = \varphi(a) \cdot \varphi(i) = \varphi(a) \cdot 0 = 0, \quad \varphi(ia) = \varphi(i) \cdot \varphi(a) = 0 \cdot \varphi(a) = 0.$$

Эти равенства и означают, что произведение элемента из ядра на любой элемент кольца принадлежит ядру. □

Теперь забудем временно про гомоморфизмы и посмотрим на те подгруппы аддитивной группы кольца, которые удовлетворяют свойству втягивания. Для таких подгрупп есть специальное название — идеалы.

Определение 9.24. (Двусторонним) *идеалом* I кольца R называется такое множество, для которого выполняются два свойства:

1. I — подгруппа аддитивной группы кольца;
2. для любого $a \in R$ и любого $i \in I$ выполняются включения $ai \in I$, $ia \in I$.

Если в этом определении оставить только первое включение (втягивание при умножении слева), то соответствующее множество будет называться *левым идеалом*. Если оставить только второе включение (втягивание при умножении справа), то получаем определение *правого идеала*.

Разумеется, для коммутативных колец разницы между двусторонними, левыми и правыми идеалами нет — эти понятия совпадают.

Замечание 9.25. При изучении групп мы доказали удобный критерий того, что множество является подгруппой: замкнутость относительно операции xy^{-1} ($x - y$ в аддитивной записи). Поэтому в определении идеалов первое свойство можно заменить на такое: для любых $i_1 \in I$, $i_2 \in I$ выполняется включение $i_1 - i_2 \in I$.

Пример 9.26. Возьмём кольцо целых чисел $\mathbb{Z}$. Выберем в нем фиксированный элемент n, и рассмотрим все его кратные, то есть множество $n\mathbb{Z} = \{rn : r \in \mathbb{Z}\}$. Это множество — идеал, что легко проверить из определения.

Более того, других идеалов в $\mathbb{Z}$ нет. Действительно, пусть $I \subseteq \mathbb{Z}$ — идеал. Если $I = \{0\}$, то $I = 0\mathbb{Z}$. В противном случае I содержит положительные числа (так как идеал — подгруппа аддитивной группы $\mathbb{Z}$). Пусть n — наименьшее положительное число, принадлежащее I. Докажем, что $I = n\mathbb{Z}$. В силу определения идеала $I \supseteq n\mathbb{Z}$. С другой стороны, если a принадлежит идеалу I, то и остаток r от деления a на n принадлежит I, так как $r = a - qn$. Но тогда, если a не принадлежит идеалу $n\mathbb{Z}$, то $r \neq 0$ и $r < n$, что противоречит выбору n. □

Это рассуждение из примера 9.26 можно значительно обобщить. Это будет сделано далее, в разделе об евклидовых кольцах. Пока приведём ещё один пример.

Пример 9.27 (продолжение примера 9.16)**.** Какое ядро у гомоморфизма значения $\mathrm{Ev}_\phi \colon \mathbb{Q}[x] \to \mathbb{R}$?

Для начала найдём хотя бы какой-нибудь ненулевой элемент ядра $\operatorname{Ker} \mathrm{Ev}_\phi$. Это должен быть такой многочлен с рациональными коэффициентами, что ϕ является его корнем. Прямо из определения золотого сечения получаем, что $(2\phi - 1)^2 = 5$. Поэтому $4x^2 - 4x - 4 \in \operatorname{Ker} \mathrm{Ev}_\phi$. При умножении на ненулевую константу (многочлен степени 0) корни многочлена не меняются. Поэтому также $x^2 - x - 1 \in \operatorname{Ker} \mathrm{Ev}_\phi$.

Поскольку ядро обладает свойством втягивания, то для любого многочлена $f \in \mathbb{Q}[x]$ многочлен $f(x)(x^2 - x - 1)$ также принадлежит ядру Ev_ϕ. (Попросту говоря, если многочлен делится на $x^2 - x - 1$, то ϕ — его корень.)

Есть ли ещё какие-нибудь многочлены в $\operatorname{Ker} \mathrm{Ev}_\phi$? Ответ отрицательный: мы уже нашли всё ядро. Доказательство этого утверждения похоже на рассуждение в примере 9.26.

Пусть $f \in \operatorname{Ker} \mathrm{Ev}_\phi$. Разделим f на $x^2 - x - 1$ с остатком. Если остаток равен нулю, то многочлен f кратен $x^2 - x - 1$, этот случай мы уже описали выше. Если же остаток $r(x) \neq 0$, то его степень меньше 2. Из равенства

$$0 = f(\phi) = q(\phi)(\phi^2 - \phi - 1) + r(\phi)$$

заключаем, что $r(\phi) = 0$. Но тогда ϕ — рациональное число: если $r(x) = a + bx$, то $\phi = -a/b$. Но если ϕ рациональное, то и $\sqrt{5}$ — рациональное, а это не так (см. набросок доказательства выше). Таким образом, случай $r(x) \neq 0$ невозможен. Все многочлены из ядра Ev_ϕ кратны $x^2 - x - 1$. □

Замечание 9.28. Бывают ли такие числа $a \in \mathbb{R}$, для которых $\operatorname{Ker} \mathrm{Ev}_a = (0)$? Другими словами, такое число a не является корнем никакого многочлена с рациональными коэффициентами (или, равносильно, многочлена с целыми коэффициентами). Числа с таким свойством называются *трансцендентными*. (А корни многочленов с целыми коэффициентами называются *алгебраическими числами*.)

Существование трансцендентных чисел легко доказать из общих соображений теории множеств. Множество алгебраических чисел счётно, а множество действительных чисел — нет. Поэтому трансцендентные числа существуют. На самом деле, почти все действительные числа трансцендентные. Легко привести конкретные примеры трансцендентных чисел: это хорошо известные числа e и π. Однако доказательство их трансцендентности требует значительных усилий, здесь мы его не приводим.

В обоих разобранных примерах ядра гомоморфизмов (идеалы) имели очень простой вид: $\{ra : r \in R\}$, то есть множество кратных некоторого элемента кольца. И кольцо $\mathbb{Z}$, и кольцо $\mathbb{Q}[x]$ — коммутативные кольца с единицей. Поэтому другими словами множество кратных некоторого элемента можно описать как наименьший идеал, содержащий a. Действительно, если $a \in I$, то $\{ra : r \in R\} \subseteq I$ из-за свойства втягивания. С другой стороны, в коммутативном кольце с единицей множество $\{ra : r \in R\}$ содержит a, так как $a = 1 \cdot a$ (вот тут и понадобилась единица кольца) и уже является идеалом. Свойство втягивания мгновенно следует из ассоциативности умножения: $r'(ra) = (r'r)a \in \{ra : r \in R\}$. Чтобы проверить, что кратные a образуют подгруппу аддитивной группы кольца, воспользуемся полезной леммой, упомянутой в замечании 9.25, и проверим замкнутость этого множества относительно вычитания:

$$r'a - r''a = (r' - r'')a \in \{ra : r \in R\}$$

(использовали дистрибутивность вычитания в кольце).

Обобщим конструкцию множества кратных. Для этого уточним слово «наименьший».

Утверждение 9.29. *Пересечение идеалов — идеал.*

Доказательство. Мы уже доказывали, что пересечение подгрупп является подгруппой. Свойство втягивания выполняется по очевидным причинам. Пусть $r \in R$, а i принадлежит пересечению идеалов. Но тогда ri также принадлежит пересечению:

по свойству втягивания этот элемент принадлежит каждому из рассматриваемых идеалов. □

Определение 9.30. Пусть S — подмножество кольца R. *Идеал, порождённый множеством* S (обозначается (S)), — это пересечение всех идеалов, содержащих S.

Определение 9.31. Идеал называется *главным*, если он порождён одним элементом. Обозначается идеал как (a).

Идеалы (главные) в примерах выше можно обозначить как (n) (кратные числа n) и $(x^2 - x - 1)$ (кратные многочлена $x^2 - x - 1$). Далее мы будем пользоваться этим удобным и коротким обозначением. В случае конечных множеств порождающих будем перечислять порождающие через запятую и окружать их скобками, как и в случае главных идеалов,.

Упражнение 9.32. Найдите $(12, 15) \subset \mathbb{Z}$.

Ответ: (3).

Утверждение 9.33. *В коммутативном кольце с единицей идеал* (S) *состоит в точности из всех конечных сумм вида*

$$\sum_\alpha r_\alpha s_\alpha, \quad r_\alpha \in R,\ s_\alpha \in S \tag{9.1}$$

(линейные комбинации элементов S*).*

Доказательство. Нужно повторить рассуждение для множества кратных, которое приведено выше.

Поскольку $s = 1 \cdot s$, все элементы S принадлежат множеству линейных комбинаций.

Любой идеал, содержащий S, обязан содержать и все линейные комбинации: из-за свойства втягивания $rs \in (S)$ для любых $s \in S$, $r \in R$; а поскольку идеал — подгруппа аддитивной группы, он замкнут относительно сложения.

Осталось проверить, что множество линейных комбинаций является идеалом:

$$\begin{aligned} &\sum_\alpha r_\alpha s_\alpha - \sum_\alpha r'_\alpha s_\alpha = \sum_\alpha (r_\alpha - r'_\alpha) s_\alpha; \\ &r \sum_\alpha r_\alpha s_\alpha = \sum_\alpha (r r_\alpha) s_\alpha \end{aligned} \tag{9.2}$$

(мы опять используем полезную характеризацию подгрупп как множеств, замкнутых относительно вычитания). □

Контрольный вопрос 9.34. В первом равенстве (9.2) суммирование в линейной комбинации записано по одному и тому же множеству элементов множества порождающих S. Объясните, почему мы ничего не теряем при таком предположении (в определении линейные комбинации берутся по всем возможным конечным подмножествам S).

Обратите внимание, что существование единицы существенно для утверждения 9.33. Иначе пришлось бы записывать более громоздкие выражения из сумм, в которые помимо кратных порождающих входят и сами порождающие.

Контрольный вопрос 9.35. Проверьте, что $(a) = \{ra + na \mid r \in R, n \in \mathbb{Z}\}$ в любом коммутативном кольце R.

Пример 9.36. Приведём пример, когда возникает разница между линейными комбинациями и более общим представлением, включающем суммы порождающих без коэффициентов.

Рассмотрим кольцо многочленов с чётными целыми коэффициентами, в наших обозначениях это $(2\mathbb{Z})[x]$.

Из каких элементов состоит идеал $(2x)$? Конечно, он содержит все кратные многочлена $2x$.

Контрольный вопрос 9.37. Проверьте, что кратные $2x$ в кольце $(2\mathbb{Z})[x]$ — это многочлены с нулевым свободным членом и коэффициентами, кратными 4.

Сам многочлен $2x$ не входит в множество кратных $2x$. Поэтому многочлен из идеала $(2x)$ — это либо кратное $2x$, либо кратное плюс $2x$. □

Контрольный вопрос 9.38. Докажите последнее утверждение.

Определение 9.39. Кольцо, в котором все идеалы, отличные от самого кольца, — главные, называется *кольцом главных идеалов*.

В примере 9.26 мы проверили, что кольцо целых чисел является кольцом главных идеалов. Рассуждение из примера 9.27 не очень трудно расширить до доказательства, что и кольцо многочленов с коэффициентами в поле является кольцом главных идеалов. Именно эти примеры колец и будут для нас самыми важными.

Тем не менее, мы докажем ниже аналогичные утверждения в несколько большей общности (рассуждения труднее не станут).

9.3 Кольца вычетов и теорема о гомоморфизмах колец

Аналогично группам, сюръективный гомоморфизм колец однозначно (с точностью до симметрий) определяется своим ядром, а любой двусторонний идеал является ядром некоторого гомоморфизма. В этом разделе мы фактически повторим рассуждения и конструкции из раздела 4.2 для случая гомоморфизмов колец.

Хотя рассуждения те же самые по существу, терминология в случае колец отличается (по историческим причинам).

Определение 9.40. Пусть I — двусторонний идеал кольца R.

Смежные классы по I как подгруппе аддитивной группы кольца называются *классами вычетов*.

Кольцо классов вычетов R/I по модулю идеала I состоит из классов вычетов, на которых определены операции

$$[a] + [b] = [a + b],$$
$$[a] \cdot [b] = [ab].$$

В этих формулах использовано краткое обозначение смежного класса $[a] = a + I$, которым мы будем пользоваться и в дальнейшем.

Поскольку теория идеалов в кольцах во многом получилась как обобщение теории сравнимости целых чисел по модулю, будем также использовать следующее определение.

Определение 9.41. Пусть I — двусторонний идеал кольца R.

Если элементы a, b кольца R принадлежат одному классу вычетов по модулю двустороннего идеала I, то такие элементы называются *сравнимыми* по модулю идеала I. Отношение сравнимости обозначается $a \equiv b \pmod{I}$ (иногда для краткости I не пишут), читается «a сравнимо с b по модулю идеала I».

В кратких обозначениях для классов вычетов отношение сравнимости выражается как $a \in [b]$. Эта запись формально несимметрична, но, как мы знаем, симметрична по существу: $a \in [b]$ равносильно $b \in [a]$. (Принадлежность одному смежному классу по подгруппе является отношением эквивалентности, так как несовпадающие смежные классы не пересекаются.)

Как и всегда в случае определения операций на классах эквивалентности через действия с представителями этих классов, необходимо проверить корректность введённых операций, то есть, что результат не зависит от выбора представителей из операндов (классов эквивалентности).

По сути мы повторим рассуждение, которое было проведено раньше для вычетов по модулю целого числа.

Утверждение 9.42. *Если $a_1 \equiv a_2$, $b_1 \equiv b_2$, то $a_1 + b_1 \equiv a_2 + b_2$ и $a_1 b_1 \equiv a_2 b_2$.*

Доказательство. Как мы проверяли, два элемента принадлежат одному классу смежности по подгруппе, если их разность принадлежит подгруппе (лемма 2.38 в случае аддитивной записи групповой операции).

Поэтому условие утверждения равносильно тому, что

$$i = a_1 - a_2 \in I,$$
$$j = b_1 - b_2 \in I.$$

Нужно доказать, что тогда

$$(a_1 + b_1) - (a_2 + b_2) \in I,$$
$$a_1 b_1 - a_2 b_2 \in I.$$

Оба включения проверяются прямым вычислением. Первое:

$$(a_1 + b_1) - (a_2 + b_2) = (a_1 - a_2) + (b_1 - b_2) = i + j \in I$$

(сложение в кольце коммутативно, идеал — подгруппа аддитивной группы и потому замкнут относительно сложения).

Второе проверяется чуть более сложной выкладкой

$$a_1b_1 - a_2b_2 = (a_2 + i)(b_2 + j) - a_2b_2 = ib_2 + a_2j + ij \in I,$$

так как каждое из слагаемых в последней формуле принадлежит идеалу: первое из-за свойства левого втягивания, второе — из-за свойства правого втягивания, третье — по любому из этих свойств.

Как видим, в этом рассуждении существенно, что идеал — двусторонний. □

Лемма 9.43. *Для двустороннего идеала множество классов вычетов с операциями, заданными в определении* 9.40*, является кольцом.*

Доказательство. Поскольку идеал является подгруппой аддитивной группы, то R/I является факторгруппой по операции сложения и эта факторгруппа абелева.

Ассоциативность умножения и дистрибутивность легко следуют из определения 9.40 операций с классами вычетов и тех же свойств в самом кольце R:

$$[a] \cdot ([b] \cdot [c]) = [a] \cdot [bc] = [a(bc)] = [(ab)c] = [ab] \cdot [c] = ([a] \cdot [b]) \cdot [c],$$
$$[a] \cdot ([b] + [c]) = [a] \cdot [b + c] = [a(b + c)] = [ab + ac] = [a] \cdot [b] + [a] \cdot [c],$$
$$([a] + [b]) \cdot [c] = [a + b] \cdot [c] = [(a + b)c)] = [ac + bc] = [a] \cdot [c] + [b] \cdot [c],$$

здесь мы используем корректность определения операций через представителей, доказанную в утверждении 9.42. □

Определение 9.44. Отображение $\varphi\colon R \to R/I$, задаваемое правилом $\varphi\colon a \mapsto [a]$, называется *каноническим гомоморфизмом* кольца на кольцо вычетов по модулю идеала.

То, что это гомоморфизм, ясно из определения 9.40 операций с классами вычетов.

Теорема 9.45 (теорема о гомоморфизмах колец)**.** *Пусть $\varphi\colon R_1 \to R_2$ — гомоморфизм колец. Тогда кольцо классов вычетов по модулю ядра гомоморфизма изоморфно гомоморфному образу кольца*: $R_1/\operatorname{Ker}\varphi \cong \varphi(R_1)$.

Доказательство. Изоморфизм устанавливается естественным образом:

$$\alpha\colon [r] \mapsto \varphi(r). \tag{9.3}$$

Здесь $r \in R_1$, $[r]$ — класс вычетов по модулю $I = \operatorname{Ker}\varphi$.

Нужно проверить, что отображение (9.3) определено корректно. Сравнимость элементов по модулю ядра $r_1 \equiv r_2 \pmod I$ влечёт равенство $\varphi(r_1 - r_2) = 0$, то есть $\varphi(r_1) = \varphi(r_2)$. Это и означает корректность определения (9.3) отображения α.

Проверим свойства гомоморфизма:

$$\alpha([r_1] + [r_2]) = \alpha([r_1 + r_2]) = \varphi(r_1 + r_2) = \varphi(r_1) + \varphi(r_2) = \\ = \alpha([r_1]) + \alpha([r_2]),$$
$$\alpha([r_1] \cdot [r_2]) = \alpha([r_1 \cdot r_2]) = \varphi(r_1 \cdot r_2) = \varphi(r_1) \cdot \varphi(r_2) = \\ = \alpha([r_1]) \cdot \alpha([r_2]).$$

Ядро этого гомоморфизма нулевое, так как $\varphi(r) = 0$ означает по определению, что $r \in I$, то есть $[r]$ — нулевой элемент класса вычетов (ядро гомоморфизма). Значит, α — инъективно.

Очевидно также, что α сюръективно (оно переводит в $\varphi(r)$ класс вычетов $[r]$).

Итак, все свойства изоморфизма для α выполнены. □

Рассмотрим несколько примеров колец вычетов и гомоморфизмов.

Пример 9.46. Рассмотрим кольцо многочленов $\mathbb{Z}[x]$ с целыми коэффициентами и идеал (x), порождённый многочленом x. Докажем, что $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$.

Рассмотрим отображение

$$\varphi\colon \mathbb{Z}[x] \to \mathbb{Z}, \qquad \varphi\colon a_0 + a_1 x + \cdots + a_d x^d \mapsto a_0,$$

которое переводит многочлен в его свободный член. Это гомоморфизм, так как по правилам арифметических операций с многочленами их свободные члены складываются и перемножаются. С другой стороны, это сюръективный гомоморфизм, поскольку любое целое число является свободным членом некоторого многочлена (например, некоторой константы: подумайте, какой).

По теореме о гомоморфизмах $\mathbb{Z}[x]/\operatorname{Ker}\varphi \cong \mathbb{Z}$. Осталось проверить, что $\operatorname{Ker}\varphi = (x)$. По определению $\varphi(x) = 0$. В обратную сторону также очень просто: если $a_0 = 0$, то

$$a_1 x + \cdots + a_d x^d = (a_1 + \ldots a_d x^{d-1})x.$$ □

Пример 9.47. Рассмотрим теперь кольцо вычетов $\mathbb{Z}[x]/(x^2)$. Аналогично предыдущему примеру проверяется, что класс вычетов однозначно задаётся первыми двумя коэффициентами a_0 и a_1.

Однако $\mathbb{Z}[x]/(x^2)$ не изоморфно $\mathbb{Z} \oplus \mathbb{Z}$. Хотя сложение коэффициентов покомпонентное, умножение многочленов и, соответственно, умножение классов вычетов в кольце $\mathbb{Z}[x]/(x^2)$ устроено иначе. В частности, в кольце $\mathbb{Z}[x]/(x^2)$ есть нильпотентный элемент (это класс $[x]$: $[x] \cdot [x] = [x^2] = [0]$), а в кольце $\mathbb{Z} \oplus \mathbb{Z}$ такого элемента нет: $(a, b)^n = (a^n, b^n) \neq (0, 0)$, если $a \neq 0$ или $b \neq 0$. □

Пример 9.48. Пусть двусторонний идеал I кольца R порождён всеми элементами вида x^2, где $x \in R$. Докажем, что тогда кольцо классов вычетов R/I антикоммутативно, то есть

$$ab = -ba \quad \text{для всех } a, b \in R/I.$$

Из построения идеала следует, что $x^2 \in I$, $y^2 \in I$, $(x+y)^2 \in I$ для любых $x, y \in R$. Поэтому

$$(x + y)^2 - x^2 - y^2 = xy + yx \in I$$

(поскольку идеал является подгруппой аддитивной группы кольца).

Но это означает, что для любых $x, y \in R$ выполняется равенство с классами вычетов $[xy + yx] = 0$, что равносильно $[x] \cdot [y] = -[y] \cdot [x]$. Поскольку канонический гомоморфизм сюръективен, это и означает антикоммутативность умножения в кольце вычетов R/I. □

Пример 9.49. Кольцо $C_0[0,1]$ состоит из непрерывных на отрезке $[0,1]$ функций со значениями в $\mathbb{R}$, операции — поточечное сложение и умножение.

Рассмотрим множество I тех функций f из кольца $C_0[0,1]$, которые обращаются в 0 на отрезке $[1/3, 2/3]$.

Легко видеть, что I — идеал: сумма двух функций, которые равны нулю на отрезке $[1/3, 2/3]$, также равна нулю на этом отрезке; произведение такой функции на любую другую равно 0 на отрезке $[1/3, 2/3]$.

Как устроено кольцо вычетов C_0/I? Докажем, что оно изоморфно $C_0[0,1]$.

Для начала докажем, что $C_0/I \cong C_0[1/3, 2/3]$. Если $f \equiv g \pmod I$, то $f - g \in I$, то есть равна 0 на отрезке $[1/3, 2/3]$. Поэтому для любой функции из класса вычетов $[f]$ ограничение на $[1/3, 2/3]$ совпадает с ограничением функции f на $[1/3, 2/3]$.

Таким образом, отображение $\alpha\colon [f] \to f\big|_{[1/3,2/3]}$ корректно определено. Оно задаёт гомоморфизм $\alpha\colon C_0[0,1]/I \cong C_0[1/3,2/3]$. Ядро этого гомоморфизма нулевое, так как функция, которая равна нулю на всём отрезке $[1/3, 2/3]$, принадлежит идеалу I. Очевидно также, что оно сюръективное: прообразом функции $g\colon [1/3,2/3] \to \mathbb{R}$ будет функция

$$f(x) = \begin{cases} g(1/3), & \text{если } 0 \leqslant x \leqslant 1/3, \\ g(x), & \text{если } 1/3 \leqslant x \leqslant 2/3, \\ g(2/3), & \text{если } 2/3 \leqslant x \leqslant 1, \end{cases}$$

которая непрерывна, как легко проверить.

Осталось построить изоморфизм $C_0[1/3,2/3]$ и $C_0[0,1]$. Например, годится такое отображение $g(x) \mapsto g(3x-1)$. Если $1/3 \leqslant x \leqslant 2/3$, то $0 \leqslant 3x - 1 \leqslant 1$, так что это и впрямь отображение $C_0[1/3,2/3]$ и $C_0[0,1]$. Это отображение взаимно однозначное, обратное отображение задаётся как $f(x) \mapsto f((x+1)/3)$. Сохранение операций очевидно. □

Пример 9.50 (продолжение примера 9.27)**.** В том примере мы нашли ядро гомоморфизма значения $\mathrm{Ev}_\phi\colon \mathbb{Q}[x] \to \mathbb{R}$, где $\phi = (1+\sqrt{5})/2$ — золотое сечение. Напомним, что это главный идеал, порождённый многочленом $x^2 - x - 1$.

Как описать кольцо вычетов по модулю ядра этого многочлена? По теореме о гомоморфизме, оно изоморфно образу кольца многочленов при этом гомоморфизме, то есть кольцу тех действительных чисел, которые представляются в виде

$$a_0 + a_1\phi + \cdots + a_d\varphi^d.$$

Обозначим это кольцо $\mathbb{Q}(\phi)$. Его можно определить короче: это минимальное подкольцо действительных чисел, содержащее все рациональные числа и ϕ. Действительно, все элементы из $\mathbb{Q}(\phi)$ обязаны принадлежать любому подкольцу действительных чисел, содержащему $\mathbb{Q}$ и ϕ, в силу замкнутости подкольца относительно умножения и сложения.

Есть и другой способ описать это кольцо. Оно состоит из действительных чисел, которые представляются в виде $a + b\phi$, где a и b — рациональные.

Ясно, что такое множество является векторным пространством над $\mathbb{Q}$, то есть замкнуто относительно линейных комбинаций с рациональными коэффициентами.

Поэтому достаточно проверить, что все степени φ представляются в таком виде. Для $1 = \varphi^0 = 1 + 0 \cdot \varphi$ и $\phi = 0 \cdot 1 + 1 \cdot \phi$ это очевидно. Далее рассуждаем по индукции.

Если

$$\phi^k = a + b\phi, \qquad a, b \in \mathbb{Q},$$

то

$$\phi^{k+1} = \phi(a + b\phi) = a\phi + b\phi^2 = b + (a + b)\phi,$$

так как $\varphi^2 = 1 + \phi$. □

9.4 Теорема о максимальном идеале

Далее мы будем строить поля как кольца вычетов по модулю некоторого идеала. В этом разделе приведём критерий того, что кольцо вычетов коммутативного кольца с $1 \neq 0$ по модулю идеала является полем.

Для начала разберёмся, как ведут себя идеалы при гомоморфизмах. При точной формулировке это вопрос разбивается на два. Пусть $\varphi\colon R_1 \to R_2$ — гомоморфизм колец.

Первый вопрос: если $I \subset R_1$ — идеал, верно ли, что его образ $\varphi(I)$ тоже идеал?

Второй вопрос: если $J \subset R_2$ — идеал. Верно ли, что его полный прообраз $\varphi^{-1}(J)$ также идеал?

Ответ на первый вопрос в общем случае отрицательный.

Пример 9.51. Рассмотрим гомоморфизм кольца R с единицей, $0 \neq 1$, в кольцо многочленов с коэффициентами в R, который переводит элемент r в константу r, то есть многочлен нулевой степени, нулевой коэффициент которого равен r, если $r \neq 0$, и нулевой многочлен в случае $r = 0$.

Если вспомнить, как умножаются многочлены, то станет ясно, что это гомоморфизм. Его образ — множество констант в кольце многочленов $R[x]$. Оно не является идеалом, так как не удовлетворяет свойству втягивания: $x \cdot 1 = x$ и это уже не константа. □

Однако такие патологии невозможны для сюръективных гомоморфизмов.

Лемма 9.52. *Пусть $\varphi\colon R_1 \to R_2$ — сюръективный гомоморфизм колец, а $I \subset R_1$ — (левый, правый, двусторонний) идеал в кольце R_1. Тогда $\varphi(I)$ — (левый, правый, двусторонний) идеал в кольце R_2.*

Доказательство. Идеал является подгруппой аддитивной группы кольца, а гомоморфизм колец является гомоморфизмом аддитивных групп. Поэтому $\varphi(I)$ — подгруппа аддитивной группы кольца R_2 (см. утверждение 4.9).

Осталось проверить свойства втягивания. Пусть $y = \varphi(i)$, $i \in I$, где I — левый идеал. Так как гомоморфизм сюръективный, то для любого $r \in R_2$ найдётся такой $x \in R_1$, что $\varphi(x) = r$.

Сохранение операций при гомоморфизме влечёт равенства $ry = \varphi(x) \cdot \varphi(i) = \varphi(xi)$. Свойство (левого) втягивания для идеала I влечёт $xi \in I$. Поэтому $ry = \varphi(xi) \in \varphi(I)$, то есть для образа I также выполняется свойство левого втягивания.

Проверка свойства правого втягивания для образа правого идеала при сюръективном гомоморфизме аналогична. □

На второй вопрос ответ всегда положительный.

Лемма 9.53. *Пусть $\varphi\colon R_1 \to R_2$ — гомоморфизм колец, а $J \subset R_1$ — (левый, правый, двусторонний) идеал в кольце R_2. Тогда $\varphi^{-1}(J)$ — (левый, правый, двусторонний) идеал в кольце R_1.*

Доказательство. Идеал является подгруппой аддитивной группы кольца, а гомоморфизм колец является гомоморфизмом аддитивных групп. Поэтому $\varphi^{-1}(J)$ является подгруппой аддитивной группы кольца R_1 в силу утверждения 4.10.

Осталось проверить свойства втягивания. Пусть J — левый идеал, а $\varphi(x) = j \in J$, то есть $x \in \varphi^{-1}(J)$. Тогда $\varphi(rx) = \varphi(r) \cdot \varphi(j) \in J$ для любого элемента r кольца R в силу свойства левого втягивания для идеала J. Это и означает выполнение свойства левого втягивания для $\varphi^{-1}(J)$.

Проверка свойства правого втягивания для полного прообраза правого идеала аналогична. □

В любом кольце есть по крайней мере два идеала: само кольцо и нулевой идеал, который содержит только 0. Эти идеалы называются *несобственными*, а любой другой идеал называется *собственным*.

Теорема 9.54. *Коммутативное кольцо с $1 \neq 0$ является полем тогда и только тогда, когда оно не содержит собственных идеалов.*

Доказательство. Пусть $(0) \subset I \subseteq F$ — ненулевой идеал поля F. Выберем в нём какой-нибудь элемент $a \neq 0$. По свойству втягивания $1 = a^{-1} \cdot a \in I$, а потому и любой элемент поля $x = x \cdot 1$ принадлежит идеалу. То есть $I = F$ и собственных идеалов в поле нет.

В обратную сторону. Пусть коммутативное кольцо R с единицей не содержит собственных идеалов. Выберем в нём какой-нибудь элемент $a \neq 0$ и рассмотрим идеал (a), порождённый этим элементом. Поскольку этот идеал не совпадает с нулевым, он должен совпадать со всем кольцом R. Но тогда $1 \in (a)$, то есть $1 = \tilde{a} \cdot a$, то есть a обратим.

Итак, любой ненулевой элемент обратим, то есть ненулевые элементы кольца образуют группу относительно умножения. Это и означает, что кольцо R является полем. □

Комбинируя теорему 9.54 с двумя предыдущими леммами об образах и прообразах идеалов, получаем искомый критерий.

Определение 9.55. Идеал I кольца R называется *максимальным*, если он не содержится строго ни в каком собственном идеале, то есть из $I \subset J$, где J — идеал, следует, что $J = R$.

Теорема 9.56 (теорема о максимальном идеале)**.** *Кольцо вычетов R/I коммутативного кольца R с $1 \neq 0$ является полем тогда и только тогда, когда идеал I максимальный.*

Доказательство. Пусть R/I поле. Тогда в нём нет собственных идеалов по теореме 9.54. Если J — идеал кольца R, который строго содержит I, то образ J при каноническом гомоморфизме в кольцо вычетов R/I ненулевой (какие-то элементы идеала J не принадлежат I и потому лежат в классах вычетов, отличных от I). Поскольку канонический гомоморфизм сюръективный, этот образ является идеалом в поле R/I и потому совпадает со всем кольцом вычетов. В частности, какой-то элемент $j \in J$ лежит в классе вычетов $[1]$, содержащем 1. То есть, $j = 1 + i$, где $i \in I$. Но тогда $1 = j - i \in J$ (так как мы предполагаем, что $I \subset J$). Поэтому $R = (1) \subseteq J$. Это и означает, что идеал I максимальный.

В обратную сторону: пусть I — максимальный идеал кольца R, а $J \subseteq R/I$ — ненулевой идеал кольца вычетов. Полный прообраз J при каноническом гомоморфизме строго содержит I (так как идеал J содержит нулевой класс вычетов, то есть идеал I, но не совпадает с ним). Значит, полный прообраз J совпадает со всем кольцом R и, в частности, содержит 1. Это означает, что $[1] \in J$ и потому $J = R/I$ (кратные единицы дают все кольцо). Значит, в кольце R/I нет собственных идеалов. По теореме 9.54 кольцо R/I является полем. □

Пример 9.57. Найдём все идеалы в кольце $\mathbb{F}_3^n = \underbrace{\mathbb{F}_3 \oplus \mathbb{F}_3 \oplus \cdots \oplus \mathbb{F}_3}_{n \text{ раз}}$ и выделим среди них максимальные. Здесь $\mathbb{F}_3 = \mathbb{Z}/3\mathbb{Z}$ — поле из трёх элементов.

Пусть $I \subset \mathbb{F}_3^n$ — идеал. Поделим координаты на группы: N_0 содержит те координаты, которые равны 0 для любого элемента идеала I, N_* — оставшиеся.

Обозначим $e_i = (\underbrace{0, 0, \ldots, 0}_{i-1 \text{ нулей}}, 1, \underbrace{0, 0, \ldots, 0}_{n-i-1 \text{ нулей}})$.

Рассмотрим координату $i \in N_*$. Пусть $a \in I$, причём $a_i \neq 0$. Тогда по свойству втягивания $a \cdot (\alpha a_i^{-1} e_i) = \alpha e_i \in I$ для любого $\alpha \in \mathbb{F}_3$.

Поскольку I замкнут также относительно сложения, то любой элемент кольца $\mathbb{F}_3^n$, у которого нули стоят во всех координатах из N_0 принадлежит идеалу I.

Итак, любой идеал кольца $\mathbb{F}_3^n$ обязан иметь вид $I_S = \{x \in \mathbb{F}_3^n : x_i = 0,\ i \in S\}$, где $S \subseteq \{1, 2, \ldots, n\}$ — некоторое множество координат. Непосредственной проверкой убеждаемся, что I_S — идеал для любого S (выполните эту проверку самостоятельно!).

Если $S \subseteq T$, то из определения получаем включение $I_S \supseteq I_T$. Это означает, что все максимальные идеалы в кольце $\mathbb{F}_3^n$ имеют вид $I_k = \{x \in \mathbb{F}_3^n : x_k = 0\}$. □

10 Теория делимости в евклидовых кольцах

Описание идеалов (и, тем самым, возможных гомоморфизмов) колец многочленов с коэффициентами в поле очень похоже на описание идеалов в кольце целых чисел (пример 9.26). Эта аналогия прослеживается и для более общего случая так называемых *евклидовых колец.* Доказательства для многочленов не сильно проще, чем для общих евклидовых колец. Поэтому мы рассмотрим общую теорию.

Так же, как и в случае кольца целых чисел, описание идеалов в евклидовых кольцах тесно связано с теорией делимости в этих кольцах. Поэтому мы начнём с обсуждения общих вопросов делимости.

10.1 Делимость элементов колец и идеалы

В этом разделе мы ограничиваемся лишь случаем коммутативных колец с $1 \neq 0$ и без делителей нуля (области целостности). Это условие нужно добавлять в формулировки всех утверждений этого раздела.

В этом разделе мы увидим, что теория делимости для таких колец в общем случае намного сложнее теории делимости целых чисел.

Отношение делимости вводится в общем кольце R так же, как и для целых чисел: $x \mid y$ означает по определению, что $y = qx$, $q \in R$; x называется *делителем* y; y называется *кратным* x.

В любом кольце $0 = 0 \cdot r$, то есть 0 делится на любой элемент кольца. Ничего больше сказать о нуле в теории делимости нельзя и поэтому дальнейший анализ ограничивается ненулевыми элементами.

В кольце целых чисел особую роль играют ± 1: это делители любого целого числа. Нетрудно найти все такие элементы в общем случае.

Утверждение 10.1. *Обратимые элементы кольца и только они являются делителями любого элемента кольца.*

Доказательство. Пусть $\tilde{\varepsilon} \cdot \varepsilon = 1$. Тогда для любого $r \in R$ выполняется равенство $r = (r\tilde{\varepsilon})\varepsilon$, которое и означает, что $\varepsilon \mid r$.

И обратно: если $\varepsilon \mid r$ для любого $r \in R$, то, в частности, $\varepsilon \mid 1$, то есть $1 = \tilde{\varepsilon} \cdot \varepsilon$. Поэтому такой элемент ε обратим. □

Замечание 10.2. Обратимые элементы кольца называют также *делителями единицы.* Они образуют группу по умножению, и эта группа обычно называется *группой единиц кольца.*

Определение 10.3. Два ненулевых элемента x, y кольца R называются *ассоциированными*, если $x \mid y$ и $y \mid x$. Обозначать ассоциированность будем $x \sim y$.

Ассоциированность — это отношение эквивалентности.

Контрольный вопрос 10.4. Проверьте, что отношение ассоциированности обладает свойствами рефлексивности, симметричности и транзитивности. (То есть, что оно является отношением эквивалентности.)

Утверждение 10.5. *$x \sim y$ равносильно тому, что $y = \varepsilon x$, где ε — делитель единицы (обратимый элемент).*

Доказательство. Пусть $y = q'x$, $x = q''y$. Тогда $y = y \cdot 1 = q'q''y$. Поскольку в R нет делителей нуля (это общее предположение для всего раздела), то в R выполняется закон сокращения (лемма 7.16), то есть $1 = q'q''$. Значит, q' — делитель единицы.

В обратную сторону совсем просто: если $y = \varepsilon x$, где ε — делитель единицы, то $x = \varepsilon^{-1}y$. □

С точки зрения теории делимости ассоциированные элементы ведут себя одинаково. Поэтому, в частности, никакой интересной теории делимости для полей нет.

У любого элемента r в любом кольце есть очевидные делители: сам элемент r и обратимые элементы.

Определение 10.6. *Собственным делителем* называется необратимый делитель элемента r, не совпадающий с r.

Если у необратимого элемента p нет собственных делителей, то такой элемент кольца называется *простым*.

Это определение обобщает определение простых целых чисел: они являются простыми элементами кольца $\mathbb{Z}$ (с точностью до ассоциированности: -2 также является простым элементом $\mathbb{Z}$).

Для кольца целых чисел выполняется основная теорема арифметики: каждое целое число разлагается в произведение обратимого элемента и простых элементов кольца $\mathbb{Z}$ и это разложение единственно с точностью до перестановок множителей и ассоциированности. Последнее означает, что два разных разложения различаются разве что на обратимые множители, как в примере

$$15 = 3 \cdot 5 = (-3) \cdot (-5).$$

В общем случае это не так. Во-первых, простых элементов может вовсе не быть. Пример довольно сложный и требует использования фактов, которые мы пока не обсуждали. (Заметим также, что этот пример нигде дальше не потребуется.)

Пример 10.7. Рассмотрим множество D действительных чисел, которые представляются конечными суммами вида

$$\sum_{k=1}^{n} a_k 2^{r_k/2^{q_k}}, \qquad a_k, r_k, q_k \in \mathbb{Z}, \quad r_k \geqslant 0,\ q_k \geqslant 0. \tag{10.1}$$

Это множество является кольцом относительно обычных сложения и умножения действительных чисел. Свойства операций наследуются из поля $\mathbb{R}$. Замкнутость относительно сложения и умножения вполне очевидна: при сложении складываются коэффициенты a_k, а при умножении складываются двоично-рациональные показатели слагаемых. Существенно, что мы разрешаем суммы из любого числа слагаемых, поскольку при арифметических операциях количество слагаемых, вообще говоря, возрастает.

В кольце D нет делителей нуля, потому что их нет в объемлющем кольце $\mathbb{R}$.

В отличие от $\mathbb{R}$ кольцо D не является полем. Например, любое число $2^{1/2^q}$, $q > 0$, необратимо в кольце D, то есть $2^{-1/2^q}$ не принадлежит кольцу D. Доказать это не так уж просто, для этого нужно предварительно разобраться с теорией делимости многочленов. Попробуйте вернуться к этому утверждению после прочтения данной главы (используйте лемму Гаусса, лемма 10.81, и пример 10.84).

Любой необратимый элемент в этом кольце не является простым, так как делится на $2^{1/2^q}$ при достаточно большом q (бóльшем, чем все q_k в представлении (10.1) в виде суммы двоично-рациональных степеней двойки). □

Во-вторых, в кольце могут простые элементы, но не все элементы кольца разлагаются в произведение простых. Пример получается незначительной модификацией предыдущего.

Пример 10.8. В предыдущем примере в основании степени заменим 2 на какое-нибудь трансцендентное число α. Это по-прежнему будет подкольцо действительных чисел. Но в таком кольце уже есть простые, например, простые целые числа. При этом α не представляется в виде произведения простых.

Конечно, эти два утверждения нуждаются в доказательстве. Предлагаем заинтересованному читателю придумать эти доказательства самостоятельно. □

Наконец, возможна и такая ситуация, когда все элементы разлагаются в произведение простых, но это разложение не единственно (даже с учётом ассоциированности).

Пример 10.9. Рассмотрим множество $\mathbb{Z}[\sqrt{-6}] = \{x + y\sqrt{-6} : x, y \in \mathbb{Z}\}$. Легко проверить, что это множество замкнуто относительно операций сложения и умножения комплексных чисел:

$$(x_1 + y_1\sqrt{-6}) + (x_2 + y_2\sqrt{-6}) = (x_1 + x_2) + (y_1 + y_2)\sqrt{-6},$$
$$(x_1 + y_1\sqrt{-6}) \cdot (x_2 + y_2\sqrt{-6}) = (x_1x_2 - 6y_1y_2) + (x_1y_2 + y_1x_2)\sqrt{-6}.$$

Поэтому оно является подкольцом поля комплексных чисел. Единица входит в это подкольцо: $1 = 1 + 0 \cdot \sqrt{-6}$.

Докажем, что любой элемент этого кольца разлагается в произведение простых.

Заметим, что квадрат модуля любого комплексного числа из этого кольца целый:

$$|x + y\sqrt{-6}|^2 = x^2 + 6y^2.$$

Модуль произведения комплексных чисел равен произведению модулей. Поэтому обратимые элементы кольца $\mathbb{Z}[\sqrt{-6}]$ должны иметь модуль 1. Легко видеть, что такие элементы — в точности ± 1.

Если число z из $\mathbb{Z}[\sqrt{-6}]$ раскладывается в произведение собственных делителей, их модули строго меньше модуля z. Поэтому рано или поздно в разложении z в произведение останутся только простые множители.

Однако такое разложение не единственно, как показывает пример

$$\sqrt{-6} \cdot \sqrt{-6} = (-2) \cdot 3.$$

Докажем, что $\sqrt{-6}$, -2 и 3 — простые элементы кольца $\mathbb{Z}[\sqrt{-6}]$. Для этого проверим, что в этом кольце нет элементов с квадратами модулей 2 и 3, то есть уравнения

$$x^2 + 6y^2 = 2, \qquad x^2 + 6y^2 = 3$$

не имеют решений в целых числах.

Квадраты модулей $\sqrt{-6}$, -2 и 3 равны соответственно 6, 4 и 9. Разложения этих чисел в произведение собственных делителей содержат только числа 2 и 3. Поскольку в кольце $\mathbb{Z}[\sqrt{-6}]$ нет элементов с квадратами модулей 2 и 3, все эти три элемента простые. □

Каждому элементу a кольца R отвечает главный идеал (a), то есть множество кратных a (напомним, что мы сейчас рассматриваем только кольца с 1). Делимости элементов соответствует включение главных идеалов.

Утверждение 10.10. *Если $a \mid b$, то $(b) \subseteq (a)$.*

Доказательство. Пусть $b = ca$, а $x = qb \in (b)$. Тогда $x = qca \in (a)$. □

Следствие 10.11. *Если $a \sim b$, то $(a) = (b)$.*

В кольце целых чисел простые элементы порождают максимальные идеалы. Действительно, как мы видели $\mathbb{Z}/(p)$ — это поле, так что по теореме о максимальном идеале идеал (p) максимальный. В общем случае это не так.

Пример 10.12. Как мы проверили, в кольце $\mathbb{Z}[\sqrt{-6}]$ число $\sqrt{-6}$ является простым элементом. Однако идеал $(\sqrt{-6})$ не является максимальным.

Действительно, в кольце вычетов $\mathbb{Z}[\sqrt{-6}]/(\sqrt{-6})$ выполняется равенство

$$[-2] \cdot [3] = [\sqrt{-6} \cdot \sqrt{-6}] = [0],$$

причём $[-2] \neq [0]$ и $[3] \neq [0]$, что равносильно $-2 \notin (\sqrt{-6})$, $3 \notin (\sqrt{-6})$. Проверить два последних утверждения легко: квадрат модуля любого элемента, кратного $\sqrt{-6}$, делится на 6, а $6 \nmid 2^2$ и $6 \nmid 3^2$.

Таким образом, в кольце $\mathbb{Z}[\sqrt{-6}]/(\sqrt{-6})$ есть делители нуля. Но тогда это кольцо не поле. По теореме о максимальном идеале, идеал $(\sqrt{-6})$ не максимальный.

Нетрудно указать идеал, который не совпадает со всем кольцом, но строго содержит $(\sqrt{-6})$. Это идеал $(\sqrt{-6}, 3)$, то есть элементы кольца, которые представляются в виде

$$z_1\sqrt{-6} + z_2 \cdot 3, \quad z_1, z_2 \in \mathbb{Z}[\sqrt{-6}].$$

Этот идеал строго содержит $(\sqrt{-6})$, так как $3 \notin (\sqrt{-6})$. Он не совпадает со всем кольцом, так как $1 \notin (\sqrt{-6}, 3)$. Действительно, предположим обратное:

$$1 = (a + b\sqrt{-6})\sqrt{-6} + (c + d\sqrt{-6})3 = (3c - 6b) + (a + d)\sqrt{-6}.$$

Это означает, что $3c - 6b = 1$, что невозможно — левая часть равенства делится на 3, а правая — нет. □

Определение 10.13. Идеал I кольца R называется *простым*, если кольцо вычетов R/I не содержит делителей нуля.

Пример 10.12 показывает, что бывают простые элементы, которые порождают не простой идеал. Обратное невозможно для главных идеалов.

Утверждение 10.14. *Если элемент a не простой, то идеал (a) также не простой.*

Доказательство. Пусть $a = bc$ — разложение a в произведение необратимых элементов кольца. Тогда в $R/(a)$ выполняется

$$[b] \cdot [c] = [a] = [0],$$

то есть в этом кольце есть делители нуля. □

Максимальные идеалы, конечно же, простые (кольцо вычетов не только не содержит делителей нуля, но даже является полем по теореме о максимальном идеале). Но и те, и другие не обязательно главные.

Пример 10.15. Проверим, что идеал $(\sqrt{-6}, 3)$ в кольце $\mathbb{Z}[\sqrt{-6}]$ является простым (и потому максимальным).

Элементы этого идеала имеют вид

$$3(a + b\sqrt{-6}) + \sqrt{-6}(c + d\sqrt{-6}) = (3a - 6d) + (3b + c)\sqrt{-6} = 3x + y\sqrt{-6}, \quad x, y \in \mathbb{Z}.$$

Поэтому есть ровно три класса вычетов по модулю этого идеала с представителями $0, 1, 2$ соответственно. Кольцо классов вычетов совпадает с кольцом классов вычетов целых чисел по модулю 3 (простого числа) и потому является полем. Значит, идеал максимальный по теореме о максимальном идеале.

С другой стороны, идеал $(\sqrt{-6}, 3)$ не является главным. В противном случае выполнялись бы равенства $\sqrt{-6} = q_1 a$ и $3 = q_2 a$. Но квадрат модуля общего делителя $\sqrt{-6}$ и 3 должен быть общим делителем 6 и 9. Это не 1, так как идеал (1) совпадает со всем кольцом. Значит, остаётся единственная возможность (с точностью до ассоциированности): $|a|^2 = 3$. Как мы уже проверяли выше, элементов с квадратом модуля 3 в кольце $\mathbb{Z}[\sqrt{-6}]$ нет. □

Простые идеалы не обязательно максимальные.

Пример 10.16 (продолжение примера 9.46)**.** Как мы видели в том примере, $\mathbb{Z}[x]/(x) \cong \mathbb{Z}$. В целых числах нет делителей нуля. Поэтому идеал (x) простой. Но он не максимальный (кольцо вычетов не является полем). Пример содержащего его идеала легко найти. Это, скажем, идеал $(2, x)$. □

Итак, теория делимости в общем случае коммутативного кольца с единицей устроена гораздо сложнее теории делимости целых чисел. Есть, однако, класс колец, в которых теория делимости очень похожа на делимость в целых числах.

10.2 Определение евклидова кольца, основные свойства

Определение 10.17. Коммутативное кольцо R называется *евклидовым*, если для него выполнены следующие свойства.

E1: Кольцо R — целостное (то есть, в нём нет делителей нуля: из $ab = 0$ следует, что $a = 0$ или $b = 0$).

E2: Для каждого ненулевого элемента кольца определена числовая характеристика — норма, которая принимает целые неотрицательные значения. То есть *норма* — это такое отображение $N\colon R \setminus \{0\} \to \mathbb{Z}$, что $N(r) \geqslant 0$.

E3: В R возможно деление с остатком. Это означает, что для любых элементов a, b кольца, $b \neq 0$, существуют такие q, r, что $a = qb + r$ и либо $r = 0$, либо $N(r) < N(b)$. Элемент r называется остатком при делении a на b.

Это основное свойство нормы. Собственно, отсюда и возник термин «евклидово». Дело в том, что в дошедших до нас рукописях термин «деление с остатком» впервые появляется в сочинениях Евклида.

E4: Норма произведения двух ненулевых сомножителей больше либо равна норме любого из сомножителей. Формально: для любых $a, b \in R$, $a \neq 0$, $b \neq 0$ выполнено $N(ab) \geqslant \max(N(a), N(b))$.

Пример 10.18. Кольцо целых чисел евклидово. Нормой числа x можно взять абсолютную величину $|x|$. Все свойства евклидова кольца уже проверялись ранее. □

Пример 10.19. Кольцо многочленов $F[x]$ с коэффициентами в поле F евклидово. Все свойства уже проверялись ранее. □

Эти два примера основные для нашего курса. Но они не исчерпывают всех возможностей.

Пример 10.20. Любое поле является евклидовым кольцом. Действительно, поле — это область целостности. Норму любого ненулевого элемента положим равной 0. Тогда свойства E3 и E4 тривиально выполняются: (E3) возможно деление без остатка (то есть с остатком, равным 0); (E4) получаем неравенство $0 \leqslant \max(0, 0)$. □

Есть и нетривиальные примеры, ограничимся простейшим.

Пример 10.21. Кольцо *гауссовых целых* $\mathbb{Z}[i]$ состоит из комплексных чисел с целой действительной и мнимой частью. Операции те же, что и обычно с комплексными числами.

Нетрудно видеть, что $\mathbb{Z}[i]$ замкнуто относительно вычитания и потому образует подгруппу аддитивной группы поля комплексных чисел. Оно также замкнуто относительно умножения. Поэтому $\mathbb{Z}[i]$ — кольцо. Это коммутативное кольцо с единицей и без делителей нуля (так как делителей нуля нет и в более широком поле комплексных чисел).

В качестве нормы выберем квадрат модуля комплексного числа:

$$N(a + bi) = |a + bi|^2 = a^2 + b^2.$$

Из свойств модуля комплексного числа следует, что нормы гауссовых чисел перемножаются.

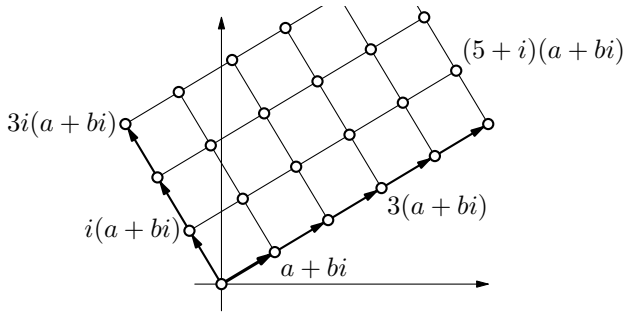


Рис. 29: Решётка гауссовых целых, которые кратны $a+bi$

Проверим возможность деления с остатком (остальные свойства евклидова кольца выполняются и для всего поля комплексных чисел).

Числа, кратные $a+bi$, образуют квадратную решётку на плоскости комплексных чисел, длина стороны квадрата равна модулю $a+bi$. Выберем тот квадрат, в который попадает $c+di$ и ближайшую к $c+di$ вершину квадрата $z(a+bi)$. Квадрат расстояния от $c+di$ до $z(a+bi)$ и есть норма разности $(c+di)-z(a+bi)$. По построению, она не превосходит $N(a+bi)/2$, то есть меньше $N(a+bi)$. □

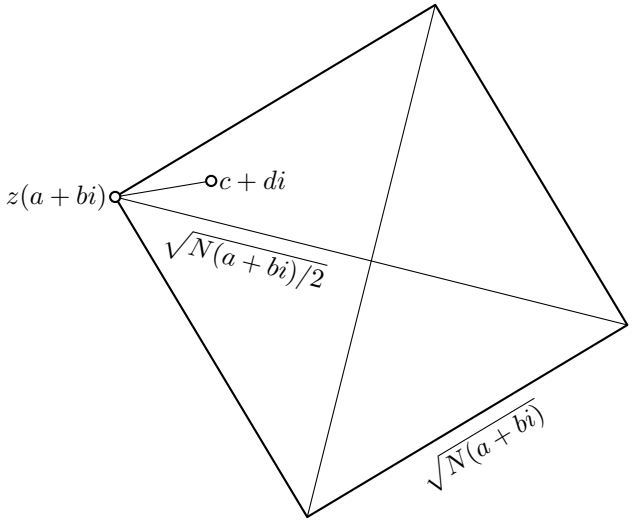


Рис. 30: Нахождение остатка при делении $c+di$ на $a+bi$

Замечание 10.22. Деление с остатком гауссовых целых, как мы его определили, неоднозначно. Эту неоднозначность легко устранить. Но обратите внимание, что однозначность деления с остатком в свойствах евклидова кольца не требуется.

Аналогично кольцу гауссовых целых можно строить и другие кольца *целых алгебраических чисел* (точное определение не приводим, оно нам не понадобится). Некоторые из них оказываются евклидовыми, другие — нет (см. примеры ниже).

Прямо из определения доказывать, что кольцо не является евклидовым, довольно трудно: ведь нужно доказать, что никакая функция из $R \setminus 0$ в $\mathbb{Z}_+$ не обладает свойствами нормы.

Однако у евклидовых колец есть много свойств. Если хотя бы одно из этих свойств не выполнено, кольцо не является евклидовым. Примеры таких доказательств появятся позже, когда мы установим достаточно много свойств евклидовых колец.

Начнём с самых простых и важных для дальнейшего свойств евклидовых колец. Поскольку норма — неотрицательное целое число, она принимает наименьшее значение. Давайте обозначим это значение $N_{\min}$.

Контрольный вопрос 10.23. Найдите $N_{\min}$ для всех приведённых выше примеров: кольца целых чисел; для кольца многочленов $F[x]$ с коэффициентами в поле F и для кольца целых гауссовых чисел.

Лемма 10.24. *Евклидово кольцо является кольцом с единицей.*

Доказательство. Выберем ненулевой элемент e' евклидова кольца R с минимально возможной нормой, $N(e') = N_{\min}$. Разделим произвольный элемент a на e' с остатком: $a = qe' + r$. По свойству E3 верно одно из двух: либо $N(r) < N(e')$, либо $r = 0$. Первое невозможно в силу минимальности нормы e'. Значит, $a = qe'$. Итак, все элементы кольца кратны e' (делятся без остатка). В частности, это верно и для самого e': $e' = ee'$. Но тогда для любого $a \in R$ имеем $ae' = aee'$, то есть $e'(a - ae) = 0$. Поскольку кольцо R целостное и $e' \neq 0$, получаем $a - ae = 0$. Значит, e является единицей кольца R. □

Лемма 10.25. *Обратимые элементы в евклидовом кольце R — это в точности элементы наименьшей нормы: $R^* = \{x \in R : N(x) = N_{\min}\}$.*

Здесь через R^* мы обозначили множество обратимых элементов кольца (*делителей единицы*, как их ещё называют).

Контрольный вопрос 10.26. Проверьте, что R^* образует группу по умножению.

Доказательство леммы 10.25. Из доказательства леммы 10.24 следует, что любой элемент e' наименьшей нормы обратим (1 является его кратным).

Норму единичного элемента (обозначаем его 1) можно оценить сверху наименьшей нормой, так как $e' = 1 \cdot e'$ и по свойству E4 имеем $N(1) \leqslant N(e') = N_{\min}$. Так как меньших значений нормы нет, то $N(1) = N_{\min}$.

Пусть u — обратимый элемент кольца R, то есть $\tilde{u}u = 1$. Тогда $N(u) \leqslant N(1) = N_{\min}$ по свойству E4. Это значит, что $N(u) = N_{\min}$. □

В дальнейшем нам понадобится усилить свойство E4 для произведений необратимых элементов кольца.

Утверждение 10.27. *Если b, c — необратимые ненулевые элементы евклидова кольца R, то их произведение $a = bc$ имеет бóльшую норму, чем сомножители:*

$$N(a) > N(b), \quad N(a) > N(c).$$

Доказательство. Разделим b на a с остатком: $b = qa + r$. Если $r = 0$, то $b = qa = qbc$ и $b(1 - qc) = 0$. В евклидовом кольце делителей нуля нет, поэтому $1 - qc = 0$, то есть c — обратим, что противоречит условию утверждения.

Значит, $r \neq 0$ и $N(r) < N(a)$. Но $a = bc$, так что $b = qcb + r$, то есть $r = b(1 - qc)$. Следовательно, $N(a) > N(r) \geqslant N(b)$ (норма остатка строго меньше нормы делителя, а норма произведения не меньше нормы ненулевого сомножителя). □

Важнейшие свойство евклидовых колец задаётся следующей теоремой. Мы уже проводили это рассуждение несколько раз в частных случаях. Теперь приведём общую формулировку.

Теорема 10.28. *Евклидовы кольца — это кольца главных идеалов.*

Доказательство. Пусть $I \subset R$ — идеал в евклидовом кольце R. Выберем ненулевой элемент $a \in I$ с наименьшей нормой среди всех ненулевых элементов идеала. Возьмём любой другой элемент идеала b и разделим его на a с остатком: $b = qa + r$, $N(r) < N(a)$. Но $r = b - qa \in I$, поэтому в силу минимальности нормы a получаем, что $r = 0$, то есть $b = qa$. Таким образом, $I = (a)$. □

Пример 10.29. Обратное к теореме 10.28 утверждение неверно. Простейший контрпример — это кольцо вычетов $\mathbb{Z}/(n)$ по любому составному модулю.

Оно не является евклидовым хотя бы потому, что в нём есть делители нуля. Но это кольцо главных идеалов, как показывает следующая лемма. □

Лемма 10.30. *Гомоморфный образ (коммутативного, с единицей) кольца главных идеалов является кольцом главных идеалов.*

Доказательство. Пусть $\varphi\colon R_1 \to R_2$ — сюръективный гомоморфизм и R_1 — кольцо главных идеалов.

Рассмотрим идеал $I \subset R_2$. Как уже было доказано, его прообраз $\varphi^{-1}(J)$ — собственный идеал в R_1, он порождён каким-то элементом a, так как R_1 — кольцо главных идеалов.

Проверим, что тогда $I = (\varphi(a))$. Ясно, что $I \supseteq (\varphi(a))$. Возьмём любой элемент $b \in I$. Из сюръективности он является прообразом какого-то элемента $\varphi^{-1}(b) \in J$. То есть $\varphi^{-1}(b) = qa$. Но тогда $\varphi(\varphi^{-1}(b)) = b = \varphi(q)\varphi(a)$ и $b \in (a)$. □

Замечание 10.31. Естественно поинтересоваться, есть ли области целостности, которые являются кольцами главных идеалов, но не являются евклидовыми кольцами? Ответ положительный.

Например, таково кольцо $\mathbb{Z}[\alpha]$, которое состоит из целочисленных линейных комбинаций степеней числа $\alpha = \frac{1}{2}(1 + \sqrt{-19})$. Делителей нуля в это кольце нет, так как это подкольцо поля комплексных чисел.

Доказательства обоих интересующих нас свойств (кольцо главных идеалов, но не евклидово кольцо) здесь опущены.

Теорема 10.28 даёт удобный способ доказательства того, что кольцо не является евклидовым. Достаточно найти в кольце не главный идеал.

Пример 10.32. Кольцо многочленов с целыми коэффициентами не является евклидовым. Действительно, идеал $(2, x)$ в этом кольце не является главным. Доказательство от противного: пусть $2 = q_1 f$ и $x = q_2 f$. Первое равенство означает, что $\deg q_1 = \deg f = 0$, то есть f константа. Возможные значения этой константы ± 1, ± 2. Второй случай невозможен, так как тогда все коэффициенты в $q_2 f$ чётные. Первый случай означает, что $(f) = \mathbb{Z}[x]$. Но идеал $(2, x)$ не совпадает со всем кольцом многочленов: например, $1 \notin (2, x)$. Действительно, в разложении

$$1 = q_1 \cdot 2 + q_2 \cdot x$$

в первом слагаемом все коэффициенты чётные, а во втором степени всех мономов больше нуля. Приходим к противоречию. □

Пример 10.33. Является ли кольцо $\mathbb{Z}[\sqrt{-3}]$ комплексных чисел вида $x + y\sqrt{-3}$, $x, y \in \mathbb{Z}$, кольцом главных идеалов?

Как и в примере 10.9 легко обнаружить неоднозначность разложения на простые:

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

Для проверки простоты $1 - \sqrt{3}$, $1 - \sqrt{3}$ и 2 в этом кольце опять используем тот факт, что квадрат модуля числа из этого кольца является целым (а модуль произведения равен произведению модулей). Как и в том примере, проверяем, что делители единицы — это ± 1.

Квадраты модулей всех этих трёх чисел равны 4, причём в кольце нет элементов с квадратом модуля 2 (проверка аналогична той, которую мы сделали в примере 10.9.

Из неоднозначного разложения извлекаем неглавный идеал $(1 + \sqrt{-3}, 2)$. Общие делители у двух разных простых — только делители единицы (обратимые элементы). Поэтому достаточно проверить, что $1 \notin (1 + \sqrt{-3}, 2)$. Пусть

$$1 = (a + b\sqrt{-3})(1 + \sqrt{-3}) + (c + d\sqrt{-3}) \cdot 2.$$

Тогда $1 = a - 3b + 2c$, $0 = a + b + 2d$. Из второго равенства получаем $b = -a - 2d$. Подставляя это в первое равенство, получаем противоречие

$$1 = a + 3a + 6d + 2c = 4a + 2c + 6d,$$

так как чётность двух частей равенства различна.

Итак, кольцо $\mathbb{Z}[\sqrt{-3}]$ не является кольцом главных идеалов. Значит, оно не евклидово. □

10.3 Основная теорема арифметики для евклидовых колец

Оказывается, связь между простыми элементами и максимальными идеалами в евклидовых кольцах точно такая же, как в кольце целых чисел.

Лемма 10.34. *Элемент p евклидова кольца прост тогда и только тогда, когда идеал (p) максимальный.*

Доказательство. Если $p = ab$ — разложение p в произведение необратимых элементов, то $p \in (a)$ и $(a) \supseteq (p)$. Чтобы показать, что идеал (p) не максимальный, нужно строгое включение $(a) \supset (p)$.

Предположим противное. Пусть $a \in (p)$. Тогда $a = qp = qba$ для некоторого $q \in R$ и закон сокращения даёт $1 = qb$, то есть b обратим, что противоречит сделанному выше предположению о разложении p в произведение необратимых элементов кольца.

В другую строну: пусть идеал (p) не максимальный. Тогда существует идеал (a), который строго содержит (p), но не совпадает со всем кольцом. Это значит, что $1 \notin (a)$ (так как $(1) = R$), то есть элемент a необратим. Поскольку $p \in (a)$, то $p = ab$ и b необратим (иначе $(p) = (a)$). Это означает, что элемент p не простой. □

Следствие 10.35. *Все простые идеалы в евклидовом кольце являются максимальными.*

Доказательство. Пусть идеал (p) не максимальный. Мы построили в доказательстве предыдущей леммы разложение $p = ab$ на необратимые элементы. Там же мы доказали, что в таком случае $(a) \supset (p)$ (аналогично $(b) \supset (p)$. Значит, $a \notin (p)$, $b \notin (p)$ и находим в кольце вычетов $R/(p)$ делители нуля $[a] \cdot [b] = [ab] = [p] = [0]$. □

Следствие 10.36. *Если p — простой элемент евклидова кольца и $p \mid ab$, то $p \mid a$ или $p \mid b$.*

Доказанное выше свойство убывания нормы при разложении в произведение необратимых (утверждение 10.27) позволяет доказать возможность разложения любого ненулевого элемента евклидова кольца в произведение простых.

Лемма 10.37. *Каждый ненулевой элемент евклидова кольца разлагается в произведение простых элементов и делителя единицы.*

Доказательство. Индукция по величине нормы. В качестве базы индукции рассмотрим утверждение леммы для элементов с минимальной нормой. Лемма 10.25 говорит, что все такие элементы обратимые.

Предположим, что утверждение теоремы верно для всех значений нормы, меньших либо равных некоторому числу m. Возьмём следующее значение нормы $m + s$, то есть наименьшее из значений нормы, бóльших m (существенное отличие от стандартной индукции здесь в том, что никто не утверждает, что значения норм идут подряд). Пусть $N(a) = m+s$. Если a — простой элемент, то утверждение выполняется. Если a — не простой, то он разлагается в произведение собственных делителей, норма которых меньше нормы a: $a = bc$, $N(b) < N(a)$, $N(c) < N(a)$ (утверждение 10.27). Для b, c утверждение справедливо в силу предположения индукции, значит оно справедливо и для a. □

Это разложение на простые, как и в случае целых чисел, единственно с точностью до ассоциированности и перестановок множителей. Доказательство также по сути повторяет доказательство основной теоремы арифметики.

Теорема 10.38 (основная теорема арифметики для евклидовых колец)**.** *Каждый элемент евклидова кольца однозначно разлагается в произведение простых элементов*

с точностью до делителей единицы и перестановок множителей. Это означает, что если есть два разложения в произведение простых

$$a = \varepsilon p_1 p_2 \dots p_n = \delta q_1 q_2 \dots q_m$$

ε, δ — делители единицы), то $n = m$ и существует такая перестановка индексов σ, что $p_i = \varepsilon_i q_{\sigma(i)}$, где ε_i — делители единицы.

Доказательство. Индукция по длине кратчайшего разложения в произведение простых.

База индукции: $n = 0$, делители единицы. Пусть делитель единицы ε является произведением простых элементов и делителя единицы: $\varepsilon = \delta q_1 q_2 \dots q_m$, δ — делитель единицы, q_i — простые. Тогда q_1 — делитель единицы, так как

$$1 = q_1 \cdot \varepsilon^{-1} \delta q_2 \dots q_m.$$

Но это противоречит определению простоты. Значит, $m = 0$.

Пусть утверждение теоремы выполнено для всех элементов кольца, которые разлагаются в произведение не более чем n простых. Рассмотрим два разложения на простые:

$$\varepsilon p_1 p_2 \dots p_n p_{n+1} = \delta q_1 q_2 \dots q_m, \quad m \geqslant n.$$

Поскольку в кольце вычетов по модулю (p_{n+1}) нет делителей нуля (это поле), то один из сомножителей в правой части должен делиться на p_{n+1}. С точностью до перенумерации элементов можно считать, что это q_m. Из простоты p_{n+1}, q_m вытекает, что $p_{n+1} = \varepsilon_{n+1} q_m$. Значит,

$$\varepsilon \varepsilon_{n+1} p_1 p_2 \dots p_n = \delta q_1 q_2 \dots q_{m-1}.$$

Осталось применить предположение индукции к этим разложениям. □

Замечание 10.39. Единственность разложения на простые выполняется не только для евклидовых колец. Например, единственно разложение на простые в кольцах многочленов $F[x_1, \dots, x_n]$ от нескольких переменных с коэффициентами в поле. Если переменных больше одной, такие кольца не являются кольцами главных идеалов и потому не являются евклидовыми.

Применение основной теоремы арифметики упрощает доказательство неевклидовости колец, см. примеры 10.9 и 10.33.

10.4 Наибольший общий делитель и алгоритм Евклида

Пусть a, b — два элемента евклидова кольца R. Рассмотрим идеал (a, b), порождённый этими элементами. По теореме 9.26 этот идеал — главный, то есть он порождён каким-то одним элементом d:

$$(a, b) = (d) \tag{10.2}$$

Этот элемент d, как мы сейчас увидим, является наибольшим общим делителем элементов a и b.

Определение 10.40. *Наибольшим общим делителем* a *и* b называют такой элемент d, что $a = qd$, $b = rd$, и для любого общего делителя d' ($a = q'd'$, $b = r'd'$) выполнено $d = d'd''$ для какого-то $d'' \in R$.

Определение 10.41. Элементы евклидова кольца R с единицей называются *взаимно простыми*, если единица — их наибольший общий делитель.

Эти определения применимы к любому коммутативному кольцу (в некоммутативном случае пришлось бы говорить о правых и левых делителях). Наибольших общих делителей в смысле данного определения может быть много.

Контрольный вопрос 10.42. Проверьте, что 5 и -5 являются наибольшими общими делителями чисел 10 и 15 в кольце $\mathbb{Z}$.

Однако различий между разными наибольшими общими делителями в смысле теории делимости нет.

Лемма 10.43. *Любые два наибольших общих делителя элементов* a *и* b *коммутативного кольца с единицей ассоциированны, то есть отличаются на множитель, который является делителем единицы. Произведение наибольшего общего делителя и обратимого элемента является наибольшим общим делителем.*

Доказательство. Пусть d_1 и d_2 — два наибольших общих делителя. Тогда по определению $d_1 = d'd_2 = d'd''d_1$, то есть $d'd'' = 1$.

В обратную сторону: если d — наибольший общий делитель, а ε — обратимый элемент (делитель единицы), то εd — также наибольший общий делитель. Действительно, если $a = qd$, $b = rd$, то $a = (q\varepsilon^{-1})(\varepsilon d)$, $b = (r\varepsilon^{-1})(\varepsilon d)$; если $d = d'd''$, то $\varepsilon d = (\varepsilon d')d''$. □

Иногда наибольшего общего делителя у двух элементов нет, даже если каждый элемент кольца разлагается в произведение простых.

Пример 10.44 (продолжение примера 10.9)**.** Как мы уже видели, в кольце $\mathbb{Z}[\sqrt{-6}] = \{x + y\sqrt{-6} : x, y \in \mathbb{Z}\}$ разложение на простые множители неоднозначно, например

$$\sqrt{-6} \cdot \sqrt{-6} = (-2) \cdot 3.$$

В частности, это означает, что у $3\sqrt{-6}$ и 6 в этом кольце нет наибольшего общего делителя. Действительно, $\sqrt{-6}$ и 3 — общие делители $3\sqrt{-6}$ и 6.

Любое кратное 3, отличное от 3, не является делителем $3\sqrt{-6}$: если $3\sqrt{-6} = (x + y\sqrt{-6}) \cdot 3$, то

$$x = 0, \quad 3y = 3,$$

откуда $y = 1$.

Поэтому наибольший общий делитель обязан иметь вид ± 3. Но эти числа не делятся на $\sqrt{-6}$: если $\pm 3 = (x + y\sqrt{-6})\sqrt{-6}$, то

$$\pm 3 = -6y, \quad x = 0,$$

что невозможно (y целое). □

В евклидовых кольцах, как мы уже говорили выше, такое невозможно. На самом деле это невозможно даже в кольцах главных идеалов (а все евклидовы кольца являются кольцами главных идеалов).

Утверждение 10.45. *Если $(a,b)=(d)$ в коммутативном кольце R с единицей, то d является наибольшим общим делителем элементов a и b.*

Доказательство. Из определений сразу следует, что d — общий делитель. Пусть d' — какой-то общий делитель, то есть $a=a'd'$, $b=b'd'$. Любой элемент идеала (a,b), в том числе и d, представляется как линейная комбинация порождающих. Поэтому $d=\tilde{a}a+\tilde{b}b=(\tilde{a}a'+\tilde{b}b')d'$. □

Далее мы используем традиционное, хотя и несколько двусмысленное обозначение: через (a,b) обозначаем наибольший общий делитель (любой из возможных). В доказательстве предыдущего утверждения мы уже отметили важный факт.

Теорема 10.46. *Наибольший общий делитель двух элементов евклидова кольца R можно представить как линейную комбинацию элементов a,b с коэффициентами из кольца*: $(a,b)=\tilde{a}a+\tilde{b}b$, $\tilde{a},\tilde{b}\in R$.

В евклидовых кольцах существует простой способ нахождения наибольшего общего делителя и решения уравнения $xa+yb=(a,b)$, который называется *алгоритмом Евклида*. Он основан на следующем простом наблюдении: наибольший общий делитель не изменяется, если вычесть из одного элемента кратное другого элемента.

Утверждение 10.47. *Для любых элементов a,b,q евклидова кольца выполняется равенство $(a,b)=(a-qb,b)$.*

Доказательство. Пусть d — общий делитель a и b, то есть $a=a'd$, $b=b'd$. Тогда $a-qb=a'd-qb'd=(a'-qb')d$, так что d является общим делителем $a-qb$ и b. И наоборот, если $a-qb=cd$, $b=b'd$, то $a=qb+cd=(qb'+c)d$. Значит, множество общих делителей для пар a,b и $a-qb,b$ одинаково. □

Алгоритм Евклида работает следующим образом. Вычисляется последовательность остатков

$$a_i=q_{i+1}a_{i+1}+a_{i+2},\quad a_0=a,\ a_1=b, \tag{10.3}$$

пока $a_{i+2}\neq 0$. Последние числа в этой последовательности $a_t=q_{t+1}a_{t+1}$, a_{t+1}, причём $a_{t+1}=(a_0,a_1)=(a,b)$.

Почему алгоритм Евклида работает конечное число шагов и даёт правильный ответ? Нормы элементов a_i уменьшаются при $i\geqslant 1$, поскольку каждый следующий элемент является остатком от деления на предыдущий. Значит, после конечного числа шагов остаток станет нулю. Корректность ответа следует из утверждения 10.47:

$$(a_0,a_1)=(a_1,a_2)=\cdots=(a_t,a_{t+1})=a_{t+1}\,.$$

Пример 10.48. Найдём НОД чисел 48 и 36 алгоритмом Евклида. Возникает такая последовательность:

$$\begin{aligned} & a_0 = 48, \\ & a_1 = 36, \\ 48 = 1 \cdot 36 + 12, \quad & a_2 = 12, \\ 36 = 3 \cdot 12 + 0, \quad & a_3 = 0. \end{aligned}$$

Таким образом, $(48, 36) = 12$. □

Пример 10.49. Найдём НОД многочленов $x^5 - 1$ и $x^3 - 1$ алгоритмом Евклида:

$$\begin{aligned} & f_0 = x^5 - 1, \\ & f_1 = x^3 - 1, \\ x^5 - 1 = x^2 \cdot (x^3 - 1) + x^2 - 1, \quad & a_2 = x^2 - 1, \\ x^3 - 1 = x \cdot (x^2 - 1) + x - 1, \quad & a_3 = x - 1, \\ x^2 - 1 = (x + 1) \cdot (x - 1) + 0, \quad & a_4 = 0. \end{aligned}$$

Значит, $(x^5 - 1, x^3 - 1) = x - 1$. □

Этот пример подсказывает общее утверждение (и его доказательство).

Теорема 10.50. *Для поля F в кольце многочленов $F[x]$ выполняется равенство*

$$(x^n - 1, x^m - 1) = x^{(n,m)} - 1.$$

Доказательство. Не ограничивая общности, считаем $n > m$. Из утверждения 10.47 получаем равенство

$$(x^n - 1, x^m - 1) = (x^n - 1 - x^{n-m}(x^m - 1), x^m - 1) = (x^{n-m} - 1, x^m - 1).$$

Повторяя его при необходимости несколько раз, получим равенство

$$(x^n - 1, x^m - 1) = (x^{n-qm} - 1, x^m - 1),$$

если $n - qm \geqslant 0$.

Продолжая такие вычисления, получаем цепочку равенств

$$(x^{a_0} - 1, x^{a_1} - 1) = (x^{a_1} - 1, x^{a_2} - 1),$$

где числа в показателях степени совпадают с результатами деления с остатком в алгоритме Евклида для пары исходных показателей n, m. Осталось заметить, что при $n = km$ выполняется

$$x^n - 1 = x^{mk} - 1 = (x^m)^k - 1 = (x^m - 1)((x^m)^{k-1} + (x^m)^{k-2} + \cdots + 1),$$

то есть в конце применения алгоритма Евклида к показателям получается пара многочленов, в которой один делится на другой. Этот последний многочлен имеет вид $x^{(n,m)} - 1$ и он является НОД исходной пары многочленов. □

Следствие 10.51. *Многочлен $x^n - 1$ делится на $x^m - 1$ тогда и только тогда, когда n делится на m.*

Часто оказывается нужным не только найти НОД двух элементов, но и его представление в виде линейной комбинации этих элементов, то есть найти в кольце решение уравнения

$$d = ax + by. \tag{10.4}$$

В кольце целых чисел такие уравнения называются диофантовыми. Мы сохраним это название и для общего случая евклидовых колец.

Для решения линейных диофантовых уравнения пригодна несложная модификация алгоритма Евклида, которая называется *расширенным алгоритмом Евклида*. Алгоритм по-прежнему вычисляет последовательность остатков (10.3). Все эти остатки лежат в идеале (a, b) и потому представляются как линейные комбинации исходной пары элементов a, b. Расширенный алгоритм вместе с остатком находит и его разложение в линейную комбинацию a, b. Коэффициенты этой линейной комбинации выражаются через коэффициенты предыдущих линейных комбинаций.

Опишем более точно порядок вычислений в расширенном алгоритме Евклида. Он рекуррентно вычисляет три такие последовательности чисел a_i, x_i, y_i, что для каждого i выполняется соотношение

$$a_i = x_i a + y_i b. \tag{10.5}$$

Начальные члены этих последовательностей такие:

$$a_0 = a, \quad x_0 = 1, \quad y_0 = 0,$$
$$a_1 = b, \quad x_1 = 0, \quad y_1 = 1,$$

для них указанное выше соотношение (10.5) выполняется очевидным образом.

Чтобы найти a_i, x_i, y_i при $i \geqslant 2$, делим a_{i-2} на a_{i-1} с остатком, это и есть a_i. Запишем его как $a_i = a_{i-2} - q_{i-1}a_{i-1}$. Остальные числа вычисляем по аналогичной формуле, используя найденное неполное частное q_{i-1}:

$$x_i = x_{i-2} - q_{i-1}x_{i-1},$$
$$y_i = y_{i-2} - q_{i-1}y_{i-1}.$$

Контрольный вопрос 10.52. Проверьте, что если $a_{i-2} = x_{i-2}a + y_{i-2}b$ и $a_{i-1} = x_{i-1}a + y_{i-1}b$, то в результате указанных вычислений получатся такие числа, что $a_i = x_i a + y_i b$.

Отсюда по индукции получаем, что для вычисленных последовательностей чисел выполняются соотношения (10.5).

Алгоритм заканчивает работу, получив нулевой остаток при делении с остатком. Коэффициенты x_i, y_i для этого нулевого остатка уже не вычисляются, но они и не нужны: последний ненулевой элемент в последовательности равен (a, b) и алгоритм уже нашёл для него представление в виде линейной комбинации исходной пары элементов.

Пример 10.53. Найдём представление 1 как целочисленной линейной комбинации (взаимно простых) чисел 12 и 17. Для это выполним вычисления по расширенному

алгоритму Евклида

$$\begin{array}{llll}
a_0 = 12, & x_0 = 1, & y_0 = 0, & \\
a_1 = 17, & x_1 = 0, & y_1 = 1, & \\
a_2 = 12, & x_2 = 1, & y_2 = 0, & \text{так как } 12 = 0 \cdot 17 + 12, \\
a_3 = 5, & x_3 = 0 - 1 \cdot 1 = -1, & y_3 = 1 - 1 \cdot 0 = 1, & \text{так как } 17 = 1 \cdot 12 + 5, \\
a_4 = 2, & x_4 = 1 - 2 \cdot (-1) = 3, & y_4 = 0 - 2 \cdot 1 = -2, & \text{так как } 12 = 2 \cdot 5 + 2, \\
a_5 = 1, & x_5 = -1 - 2 \cdot 3 = -7, & y_5 = 1 - 2 \cdot (-2) = 5, & \text{так как } 5 = 2 \cdot 2 + 1.
\end{array}$$

Итак, $1 = -7 \cdot 12 + 5 \cdot 17$, в чём несложно убедиться прямой проверкой.

Конечно, для таких маленьких чисел как 12 и 17 найти искомое представление можно и подбором. Важность алгоритма Евклида в том, что он работает для любых пар. □

Замечание 10.54. Вычисления в этом примере делают один лишний шаг, которого можно избежать, если начинать сразу с пары 17, 12. Как видим, никакой роли порядок чисел в паре не играет, как и должно быть.

Типичный случай применения алгоритма Евклида — вычисление обратного в кольце вычетов.

Пример 10.55. Найдём 7^{-1} в кольце вычетов $\mathbb{Z}/(19)$. По определению обратного должно выполняться $7^{-1} \cdot 7 - 1 \in (19)$, то есть нужно найти решение диофантова уравнения $7x + 19y = 1$. Найдём его алгоритмом Евклида:

$$\begin{array}{llll}
a_0 = 7, & x_0 = 1, & y_0 = 0, & \\
a_1 = 19, & x_1 = 0, & y_1 = 1, & \\
a_2 = 7, & x_2 = 1, & y_2 = 0, & \text{так как } 7 = 0 \cdot 19 + 7, \\
a_3 = 5, & x_3 = 0 - 2 \cdot 1 = -2, & y_3 = 1 - 2 \cdot 0 = 1, & \text{так как } 19 = 2 \cdot 7 + 5, \\
a_4 = 2, & x_4 = 1 - 1 \cdot (-2) = 3, & y_4 = 0 - 1 \cdot 1 = -1, & \text{так как } 7 = 1 \cdot 5 + 2, \\
a_5 = 1, & x_5 = -2 - 2 \cdot 3 = -8, & y_5 = 1 - 2 \cdot (-1) = 3, & \text{так как } 5 = 2 \cdot 2 + 1.
\end{array}$$

Решением диофантова уравнения является пара $(-8; 3)$. Поэтому $7^{-1} = -8 = 11 \bmod 19$. □

Расширенный алгоритм Евклида применим в любом евклидовом кольце, например, в кольце многочленов. Однако в этом случае вычисления (многократные деления с остатком) становятся более громоздкими. При ручных вычислениях бывает удобно применять другие приёмы, например, интерполяцию.

Пример 10.56. Найдём обратный к $x^2 + 1$ в кольце $\mathbb{Q}[x]/(x^2 - x)$. Заметим, что в любом классе вычетов этого кольца есть многочлен степени не выше первой. Будем искать представление обратного именно в виде $f(x) = ax + b$.

Из равенства $f(x)(x^2+1) - 1 = q(x)(x^2 - x)$ можно заметить, что левая его часть обращается в 0 в корнях многочлена $x^2 - x$. Таким образом, выполняются равенства

$$\begin{aligned}
b - 1 &= 0, \\
(a \cdot 1 + b) \cdot 2 - 1 &= 0,
\end{aligned}$$

откуда получаем $f(x) = -\frac{1}{2}x + 1$ без трудоёмких вычислений. □

Обсудим вопрос о решении линейных диофантовых уравнений общего вида $ax + by = c$. Во-первых, опишем решения *однородных* уравнений, когда правая часть равна 0.

Утверждение 10.57. *Решениями однородного линейного уравнения с двумя переменными $ax + by = 0$ с коэффициентами из евклидова кольца R ($a \neq 0$ и $b \neq 0$) являются в точности такие пары (x, y), что*

$$x = t\frac{b}{d},\ y = -t\frac{a}{d}, \quad d = (a, b),\ t \in R.$$

Доказательство. Разделив обе части уравнения на наибольший общий делитель (a, b), получим равносильное уравнение, коэффициенты которого взаимно просты. Поэтому достаточно решить уравнение $ax + by = 0$ со взаимно простыми коэффициентами.

Если $(a, b) = 1$, то для некоторых $u, v \in R$ выполнено $au + bv = 1$. Умножим равенство $ax + by = 0$ на u:

$$u(ax + by) = uax + uby = (1 - bv)x + uby = x + b(uy - vx) = 0.$$

Таким образом, $x = tb$ при некотором $t \in R$. Но тогда $by = -ax = -tab$ и $y = -ta$. С другой стороны, любая пара $(tb, -ta)$ является решением уравнения $ax+by = 0$. □

Рассмотрим теперь неоднородное уравнение $ax + by = c$. Необходимым условием его разрешимости является делимость c на $(a, b) = d$, то есть $c = kd$. Расширенный алгоритм Евклида даст решение уравнения $ax+by = d$. Решение уравнения $ax+by = = c = kd$ получится умножением обеих компонент на k.

Расширенный алгоритм Евклида даёт лишь частное решение диофантова уравнения. Иногда бывают нужны все решения. Их несложно получить из найденного частного решения и общего решения однородного уравнения.

Теорема 10.58. *Если c кратно (a, b), то решениями неоднородного уравнения $ax + by = c$ являются в точности суммы любого частного решения (x_0, y_0) этого уравнения и решений однородного уравнения $ax + by = 0$, то есть решениями являются те и только те пары (x, y), для которых*

$$x = x_0 + t\frac{b}{d},\ y = y_0 - t\frac{a}{d} \quad d = (a, b),\ t \in R.$$

Доказательство. Пусть пары (x_0, y_0), (x_1, y_1) являются решениями неоднородного уравнения. Тогда пара $(x_0 - x_1, y_0 - y_1)$ является решением однородного уравнения, так как

$$0 = c - c = (ax_1 + by_1) - (ax_0 + by_0) = a(x_1 - x_0) + b(y_1 - y_0).$$

В обратную сторону рассуждение аналогично: прибавление решения (x, y) однородного уравнения к решению (x_0, y_0) неоднородного даёт решение неоднородного уравнения:

$$a(x_0 + x) + b(y_0 + y) = (ax_0 + by_0) + (ax + by) = c + 0 = c.$$ □

Замечание 10.59. Эта связь между решениями однородных и неоднородных линейных уравнений имеет место для гораздо более широкого круга задач.

10.5 Китайская теорема об остатках для евклидовых колец

Из теоремы о максимальном идеале мы знаем, что кольцо вычетов евклидова кольца по модулю идеала, порождённого простым элементом, является полем. Китайская теорема показывает устройство колец вычетов по идеалам, порождённым произведением взаимно простых элементов.

Теорема 10.60. *Для взаимно простых элементов p_1, p_2 евклидова кольца R имеет место изоморфизм колец $R/(p_1p_2) \cong R/(p_1) \oplus R/(p_2)$.*

Доказательство. Рассмотрим гомоморфизм

$$\varphi\colon R \to R/(p_1) \oplus R/(p_2), \quad \varphi\colon r \mapsto ([r]_1, [r]_2),$$

где $[r]_i$ обозначает класс вычетов по модулю p_i.

Теорема 10.46 утверждает, что в кольце найдутся такие элементы x_1, x_2, для которых $x_1p_1 + x_2p_2 = (p_1, p_2) = 1$. Из этого равенства следует, что $[p_1]_2$ является делителем единицы в $R/(p_2)$, а $[p_2]_1$ является делителем единицы в $R/(p_1)$.

Пусть $r \in \operatorname{Ker}\varphi$, то есть $[r]_1 = 0$ и $[r]_2 = 0$. В кольце R это означает, что r кратно как p_1, так и p_2: $r = y_1p_1 = y_2p_2$. Так как p_1, p_2 — просты элементы, из основной теоремы арифметики получаем, что $p_1 \mid y_2$. Тогда $[y_2p_2]_1 = 0$ в $R/(p_1)$. Поскольку $[p_2]_1$ — делитель единицы (обратимый элемент), то $[y_2]_1 = 0$, то есть $y_2 = ap_1$. Значит, $r = ap_1p_2$. Поэтому ядро гомоморфизма φ совпадает с (p_1p_2).

При этом гомоморфизм φ сюръективен, так как для любой пары классов вычетов $([r_1]_1, [r_2]_2)$ найдётся элемент кольца $s = r_1 + (r_2 - r_1)x_1p_1 = r_2 - (r_2 - r_1)x_2p_2$, который отображается в эту пару при гомоморфизме φ. (Напомним, что $x_1p_1 + x_2p_2 = 1$.)

Осталось применить теорему о гомоморфизме колец. □

Для приложений важно, что изоморфизм, задаваемый китайской теоремой, легко вычислим в обе стороны. В прямую сторону достаточно уметь делить с остатком. Чтобы найти φ^{-1}, требуется применить алгоритм Евклида. Действительно, решение системы сравнений

$$\begin{aligned} x &\equiv r_1 \pmod{p_1}, \\ x &\equiv r_2 \pmod{p_2} \end{aligned}$$

по модулю идеалов (p_1) и (p_2) получается из решения уравнений в исходном кольце R: $x = r_1 + yp_1$, $x = r_2 + zp_2$.

Для того, чтобы найти какое-нибудь решение x достаточно решить линейное диофантово уравнение $r_1 + yp_1 = r_2 + zp_2$ и из его решения найти x.

Пример 10.61. Решим систему сравнений

$$\begin{cases} x \equiv 1 \pmod{29} \\ x \equiv -1 \pmod{19} \end{cases}$$

Для этого нужно решить уравнения в целых числах $x = 1 + 29y = -1 + 19z$.

Решение линейного диофантова уравнения $29y - 19z = 1$ найдём с помощью алгоритма Евклида:

$$\begin{aligned}
&a_0 = 29, && y_0 = 1, && z_0 = 0, \\
&a_1 = -19, && y_1 = 0, && z_1 = 1, \\
&a_2 = 10, && y_2 = 1, && z_2 = 1, && \text{т.к. } 29 = (-1)\cdot(-19) + 10, \\
&a_3 = 1, && y_3 = 2, && z_3 = 3, && \text{т.к. } -19 = (-2)\cdot 10 + 1.
\end{aligned}$$

Из полученного решения $y = 2$, $z = 3$ умножением на -2 найдём решение уравнения $1+29y = -1+19z$ и выразим через него искомое значение $x = 1+29\cdot(-2)\cdot 2 = -115$.

Если хотеть в качестве ответа положительное число, то можно добавить $29 \cdot 19$ и получить 436. □

Есть и другой, более явный, способ. Он также основан на алгоритме Евклида. Вычислим $[\tilde{p}_1]_2 = [p_1]_2^{-1}$ и $[\tilde{p}_2]_1 = [p_2]_1^{-1}$ (то есть обратные к элементу по модулю второго элемента, напомним, что мы рассматриваем взаимно простую пару). Теперь x представляется как сумма $r_1\tilde{p}_2p_2 + r_2\tilde{p}_1p_1$, так как

$$\begin{aligned}
r_2\tilde{p}_1p_1 + r_1\tilde{p}_2p_2 &\equiv r_1(\tilde{p}_2p_2) \equiv r_1 \pmod{p_1}, \\
r_2\tilde{p}_1p_1 + r_1\tilde{p}_2p_2 &\equiv r_2(\tilde{p}_1p_1) \equiv r_2 \pmod{p_2}.
\end{aligned}$$

Пример 10.62. Решим систему сравнений

$$\begin{cases} x \equiv 17 \pmod{20} \\ x \equiv 11 \pmod{17} \end{cases}$$

Вычислим вспомогательные множители $\tilde{p}_1 \equiv 20^{-1} \pmod{17}$ и $\tilde{p}_2 \equiv 17^{-1} \pmod{20}$. Детали вычислений пропускаем (всегда можно применить расширенный алгоритм Евклида), приведём результаты: $\tilde{p}_1 = 6$; $\tilde{p}_2 = -7$.

Теперь получаем ответ:

$$x = 11 \cdot 6 \cdot 20 + 17 \cdot (-7) \cdot 17 = -703.$$

Разумеется, ответом является класс вычетов по модулю $20 \cdot 17 = 340$. В частности, другим вариантом того же ответа будет 2017 (проверьте!).

Обратите внимание, что в кольце целых чисел у такой системы бесконечно много решений. А в кольце вычетов $\mathbb{Z}/(340)$ — ровно одно. □

Теорема 10.60 имеет очевидные следствия для колец вычетов целых чисел, циклических групп и колец вычетов по модулю идеалов кольца многочленов. Для циклических групп мы уже эти следствия доказывали по отдельности. Построенные изоморфизмы применялись, в частности, для доказательства мультипликативности функции Эйлера. Теперь мы объединим теоремы 3.49 и 3.58 в одну общую теорему о кольцах вычетов.

Следствие 10.63. *Если p, q — взаимно простые числа, то $\mathbb{Z}_{pq} \cong \mathbb{Z}_p \oplus \mathbb{Z}_q$.*

Следствие 10.64. *Пусть K — поле. Если многочлены $f, g \in K[x]$ взаимно просты, то $K[x]/(fg) \cong K[x]/(f) \oplus K[x]/(g)$.*

Приведём пример использования китайской теоремы (ниже приводятся ещё примеры, см. разделы 10.6 про кольца вычетов колец многочленов и 11.9 про критерий неприводимости многочлена над конечным полем).

Пример 10.65. Найдём решения уравнения $x^2 - 1 = 0$ в кольце $\mathbb{Z}/(143)$. Так как $143 = 13 \cdot 11$, эти решения получатся из решений того же уравнения в полях $\mathbb{Z}/(11)$ и $\mathbb{Z}/(13)$.

Если характеристика поля не равна 2 (что то же самое, если $-1 \neq 1$), у уравнения $x^2 - 1 = 0$ есть ровно два решения в поле: ± 1.

Китайская теорема говорит, что $\mathbb{Z}/(143) = \mathbb{Z}/(11) \oplus \mathbb{Z}/(13)$. Единицей в сумме колец будет пара $(1, 1)$. Поэтому решения уравнения $x^2 - 1 = 0$ в кольце $\mathbb{Z}/(143)$ — это пары (x_1, x_2), в которых x_1, x_2 — решения уравнения $x^2 - 1 = 0$ (каждое в своём поле).

Всего таких пар $2 \cdot 2 = 4$. Две из них отвечают 1 и -1 в кольце $\mathbb{Z}/(143)$, это пары $(1, 1)$ и $(-1, -1)$. Две другие получаются из решений систем сравнений

$$\begin{cases} x \equiv 1 \pmod{11} \\ x \equiv -1 \pmod{13} \end{cases} \qquad \begin{cases} x \equiv -1 \pmod{11} \\ x \equiv 1 \pmod{13} \end{cases}$$

Это, конечно же, противоположные вычеты в кольце $\mathbb{Z}/(143)$. Поэтому достаточно решить одну из этих систем сравнений. Например, теми способами, которые описаны выше. Но в данном случае удачная догадка сильно экономит вычисления: у первой системы есть решение 12.

Итак, у уравнения $x^2 - 1 = 0$ в кольце $\mathbb{Z}/(143)$ есть ровно четыре решения ± 1, ± 12. □

10.6 Кольца вычетов колец многочленов

Кольцо многочленов с коэффициентами в поле F является евклидовым. Из общих свойств евклидовых колец получаем, что всякий собственный идеал I в $F[x]$ порождён одним многочленом: $I = (f)$.

Обратимые элементы в кольце многочленов — это многочлены степени 0, то есть ненулевые константы. Умножая на ненулевую константу, всегда можно добиться, чтобы старший коэффициент многочлена равнялся 1. Такие многочлены называются *нормированными*. Умножение многочлена на ненулевую константу не изменяет идеал, порождённый этим многочленом. Поэтому достаточно рассматривать только идеалы порождённые нормированными многочленами.

Кольцо вычетов $F[x]/(f)$ имеет сравнительно простой вид, похожий на вид колец вычетов целых чисел. Арифметические операции выполняются аналогично: чтобы найти сумму или произведение вычетов $[g]$, $[h]$, нужно выполнить аналогичную операцию в кольце многочленов.

Если стремиться к однозначности представления вычета, то можно использовать то обстоятельство, что класс вычетов образуют многочлены, имеющие одинаковый

остаток по модулю f. Возможные значения остатка — это многочлены $g(x) = a_0 + + a_1x + \cdots + a_{d-1}x^{d-1}$ степени меньше $d = \deg f$ (или нулевой). Значит, вычет однозначно задаётся многочленом степени меньше d (или нулевым). Для поддержания такого представления при выполнении арифметических операций нужно результат операции делить с остатком на f.

Из такого однозначного представления вычетов легко посчитать количество элементов в кольце $F[x]/(f)$ для конечного поля F, $|F| = q$. Всего есть d коэффициентов, задающих остаток, каждый может принимать q значений. Всего возможных остатков (и элементов в кольце вычетов) ровно q^d штук.

Хотя количество элементов в кольце вычетов кольца многочленов над конечным полем зависит только от количества элементов в поле коэффициентов и степени многочлена, алгебраически кольца вычетов по модулю многочленов одинаковой степени могут сильно различаться.

В частности, в силу теоремы о максимальном идеале $F[x]/(f)$ является полем тогда и только тогда, когда f является простым элементом в кольце многочленов. Для колец многочленов используется специальная терминология.

Определение 10.66. Многочлен $f \in F[x]$ называется *неприводимым над полем* F, если не существует разложения на собственные множители $f = gh$, $\deg g < \deg f$, $\deg h < \deg f$.

Пусть F — конечное поле характеристики p. Вспомним про сюръективный гомоморфизм значения $\mathrm{Ev}_\alpha \colon \mathbb{F}_p[x] \to F$, где α — порождающий мультипликативной группы поля. Из теоремы о гомоморфизме получаем, что $\mathbb{F}_p[x]/\operatorname{Ker}\mathrm{Ev}_\alpha \cong F$, а из свойств евклидовых колец $\operatorname{Ker}\mathrm{Ev}_\alpha = (f)$, где f — неприводимый многочлен. Таким образом, получаем одну из основных теорем о конечных полях.

Теорема 10.67. *Каждое конечное поле характеристики p изоморфно кольцу вычетов кольца многочленов $\mathbb{F}_p[x]$ по модулю идеала, порождённого неприводимым многочленом.*

Итак, чтобы найти конечное поле из p^n элементов, нужно найти неприводимый многочлен f степени n в кольце $\mathbb{F}_p[x]$ и взять кольцо вычетов $\mathbb{F}_p[x]/(f)$. Существование такого многочлена никак не следует из общей теории: в зависимости от поля коэффициентов множества неприводимых многочленов могут быть устроены поразному.

Многочлены первой степени всегда неприводимые: если $1 = a + b$, a, b — целые неотрицательные, то либо $a = 1$, либо $b = 1$.

Контрольный вопрос 10.68. Какое поле изоморфно кольцу вычетов $F[x]/(x + a)$, $a \in F$?

Определение 10.69. Поле F называется *алгебраически замкнутым*, если все неприводимые многочлены в $F[x]$ имеют степень 1.

Контрольный вопрос 10.70. Проверьте, что поле алгебраически замкнуто тогда и только тогда, когда любой многочлен положительной степени в $F[x]$ имеет корень.

Примером алгебраически замкнутого поля является поле комплексных чисел $\mathbb{C}$.

Теорема 10.71 (основная теорема алгебры)**.** *Всякий многочлен положительной степени над полем $\mathbb{C}$ имеет корень.*

Доказательство основной теоремы алгебры мы не приводим. См., например, учебник Э. Б. Винберга [5] или книгу В. В. Прасолова [20].

Поле $\mathbb{R}$ не является алгебраически замкнутым. Многочлен x^2+1 не имеет действительных корней, поэтому он неприводим.

Упражнение 10.72. Проверьте, что $\mathbb{R}[x]/(x^2+1)$ изоморфно полю комплексных чисел $\mathbb{C}$.

Вообще-то, отсутствие корней — лишь необходимое условие неприводимости многочлена (не делится на многочлен первой степени). Однако для многочленов степени 2 и 3 оно является и достаточным.

Лемма 10.73. *Если* $\deg f \leqslant 3$*, то* $f \in F[x]$ *неприводим тогда и только тогда, когда в поле F у многочлена f нет корней.*

Доказательство. Аналогично случаю степени 1: если $2 = a + b$, a, b — целые положительные, то либо $a = 1$, либо $b = 1$. То же самое верно и для 3: если $2 = a + b$, a, b — целые положительные, то либо $a = 1$, либо $b = 1$. □

Уже для многочленов 4й степени отсутствие корней не равносильно неприводимости.

Пример 10.74. Рассмотрим многочлен $x^4 + 1 \in \mathbb{R}[x]$. Корней у него нет. Однако он приводим:

$$x^4 + 1 = (x^2+1)^2 - (\sqrt{2}x)^2 = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1).$$ □

Более того, из основной теоремы алгебры сравнительно легко вывести, что все неприводимые многочлены в $\mathbb{R}[x]$ имеют степень не больше 2.

Задача 10.75. Докажите, что любой многочлен над полем действительных чисел степени больше 2 приводим.

Лемма 10.73 позволяет без труда построить поля из p^2 элементов, $p > 2$.

Пример 10.76. Пусть $p > 2$ — простое. Возьмём квадратичный невычет a по модулю p. По определению это означает, что многочлен $x^2 - a$ не имеет корней в $\mathbb{F}_p$. Значит, он неприводим. Поэтому кольцо вычетов $\mathbb{F}_p[x]/(x^2-a)$ является полем и в этом поле p^2 элементов. □

Пример 10.77. Рассмотрим примеры неприводимых многочленов в кольце $\mathbb{F}_2[x]$. Возможных значений коэффициентов всего два, как и возможных значений корней. Поэтому, пользуясь леммой 10.73 легко выписать все неприводимые многочлены степени не выше 3:

$$x,\ x+1,\ x^2+x+1,\ x^3+x^2+1,\ x^3+x+1.$$

Упражнение 10.78. Проверьте полноту и корректность этого списка.

Приводимый многочлен степени 4 либо имеет корень, либо является произведением неприводимых многочленов степени 2. В данном случае неприводимый многочлен степени 2 всего один, его квадрат равен $x^4 + x^2 + 1$.

Контрольный вопрос 10.79. Проверьте, что $(x^2+x+1)^2 = x^4+x^2+1$ в кольце $\mathbb{F}_2[x]$.

Отсюда нетрудно получить список неприводимых многочленов степени 4:

$$x^4 + x + 1,\ x^4 + x^3 + 1,\ x^4 + x^3 + x^2 + x + 1.$$

Упражнение 10.80. Проверьте полноту и корректность этого списка.

Эти примеры позволяют построить поля с 4, 8, 16 элементами. Не очень понятно пока, различны ли поля, построенные с помощью разных неприводимых многочленов, Скажем, изоморфны ли поля $\mathbb{F}_2[x]/(x^3+x^2+1)$ и $\mathbb{F}_2[x]/(x^3+x+1)$? В каждом из этих полей 8 элементов, но правило умножения задаётся по-разному.

Далее мы увидим, что эти поля изоморфны, как и любые конечные поля с одинаковым количеством элементов. □

Позже мы увидим, что в кольце $\mathbb{F}_p[x]$ есть неприводимые многочлены любой степени. Сейчас покажем, что кольцо $\mathbb{Q}[x]$ многочленов с рациональными коэффициентами обладает таким же свойством.

Аргументы будут теоретико-числовыми, поэтому удобно перейти к многочленам с целыми коэффициентами, хотя они и не образуют евклидова кольца.

Если $q(x)$ — ненулевой многочлен с рациональными коэффициентами степени n, то, приводя коэффициенты к общему знаменателю, можно записать:

$$q(x) = \alpha(a_0 + a_1 x + \cdots + a_n x^n) = \alpha q_0,$$

где все коэффициенты a_i — целые числа, не имеющие нетривиального общего делителя, $a_n > 0$, $\alpha \in \mathbb{Q}$. Легко видеть, что многочлен q_0 и число α определены однозначно. Будем называть q_0 *примитивным многочленом*, соответствующим многочлену q.

Лемма 10.81 (Гаусс)**.** $(uv)_0 = u_0 v_0$.

Доказательство. В сущности, нужно доказать, что если у каждого из многочленов $u_0, v_0 \in \mathbb{Z}[x]$ коэффициенты взаимно просты в совокупности, то у их произведения $u_0 v_0$ коэффициенты так же взаимно просты в совокупности. Для доказательства построим гомоморфизм кольца $\mathbb{Z}[x]$ на кольцо $\mathbb{F}_p[x]$, который называется *редукцией многочлена* по модулю p. Редукция f_p многочлена f получается приведением по модулю p каждого из коэффициентов f. Посмотрев на формулы для суммы и произведения многочленов, видим, что редукция действительно является гомоморфизмом.

Предположим, что у коэффициентов $u_0 v_0$ есть общий простой делитель p. Тогда $(u_0 v_0)_p = 0$ и для редукций по модулю p имеем $(u_0)_p (v_0)_p = 0$. Но в кольце $\mathbb{F}_p[x]$ нет делителей нуля, поэтому одна из редукций $(u_0)_p$, $(v_0)_p$ равна 0. Это противоречит примитивности u_0, v_0. □

Таким образом, вопрос о приводимости многочлена над полем рациональных чисел сводится к вопросу о разложении на множители меньшей степени многочлена с целыми коэффициентами. В этом направлении имеется следующее достаточное условие неприводимости:

Теорема 10.82 (критерий Эйзенштейна)**.** *Если для многочлена* $q = a_0 + a_1x + \cdots + a_nx^n$ *с целыми коэффициентами существует такое простое число* p*, что* $p \nmid a_n$*,* $p \mid a_i$ *при* $i = 0, \ldots, n-1$*,* $p^2 \nmid a_0$*, то этот многочлен неприводим.*

Доказательство. Предположим, что q приводимый многочлен: $q = uv$. Тогда $q_p = u_pv_p$. По условию теоремы $q_p = ax^n$, $a \neq 0$. Значит, $u_p = bx^k$, $v_p = cx^m$, где $k < n$ и $m < n$. Поэтому все коэффициенты многочленов u и v, кроме старших, делятся на p, а тогда свободный член многочлена q, (то есть a_0), равный u_0v_0, делится на p^2, что противоречит условию. □

Пример 10.83. Многочлен $2x^4 - 6x^3 + 15x^2 + 21$ неприводим над полем $\mathbb{Q}$. Достаточно взять $p = 3$ и применить критерий Эйзенштейна. □

Пример 10.84. Для всякого $n > 0$ многочлен $x^n - 2$ неприводим над $\mathbb{Q}$. Достаточно взять $p = 2$ и применить критерий Эйзенштейна. Отсюда вытекает, что над полем рациональных чисел существуют неприводимые многочлены любой степени. □

Критерий Эйзенштейна даёт только достаточное условие неприводимости. Он неприменим, скажем, к многочленам $x^n + 1$, среди которых есть неприводимые.

Пример 10.85. Многочлен $x^4 + 1$ неприводим над полем $\mathbb{Q}$. У него нет корней, поэтому единственно возможное разложение: произведение многочленов степени 2 (с рациональными коэффициентами). Но такое разложение выполняется и в поле действительных чисел. Получаем в $\mathbb{R}[x]$ два разложения

$$x^4 + 1 = (x^2 - \sqrt{2}x + 1)(x^2 + \sqrt{2}x + 1) = f(x)g(x).$$

Многочлены $x^2 - \sqrt{2}x + 1$ и $x^2 + \sqrt{2}x + 1$ не имеют действительных корней (дискриминант равен $2 - 4 = -2$), поэтому неприводимы (они же второй степени). В силу единственности разложения на простые в евклидовом кольце (основная теорема арифметики) это означает, что $f(x) = \alpha(x^2 - \sqrt{2}x + 1)$ или $g(x) = \alpha(x^2 - \sqrt{2}x + 1)$ в $\mathbb{R}[x]$. Но $f(x)$, $g(x)$ имеют рациональные коэффициенты, а $\sqrt{2}$ — иррациональное число. Поэтому такие равенства невозможны. □

Контрольный вопрос 10.86. Докажите последнее утверждение.

Пример 10.87. Интересный вариант этого примера получается в характеристике 3. Приводим ли многочлен $x^4 + 1$ в кольце $\mathbb{F}_3[x]$?

Корней в $\mathbb{F}_3$ у многочлена $x^4 + 1$ нет, это быстро проверяется прямыми вычислениями. Но нужно ещё убедиться, что у этого многочлена нет разложения в произведение двух неприводимых второй степени.

Заметим, что $2 = -1$ — квадратичный невычет в поле $\mathbb{F}_3$. Поэтому многочлен $x^2 + 1$ не имеет корней в $\mathbb{F}_3$ и его степень 2. Значит, он неприводим и кольцо вычетов

$F = \mathbb{F}_3[x]/(x^2+1)$ по модулю идеала, порождённого этим многочленом является полем (в котором 9 элементов).

В поле F уже есть квадратный корень из -1: это вычет α, содержащий x:

$$[x]^2 = [x^2] = [x^2 - (x^2+1)] = [-1] = -[1].$$

Аналогично предыдущему примеру с разложением над действительными числами, получаем разложение

$$x^4 + 1 = (x^2 - \alpha x + 1)(x^2 + \alpha x + 1).$$

Можно ли в данном случае утверждать, ссылаясь на единственность разложения на множители в евклидовом кольце, что разложения над $\mathbb{F}_3$ у этого многочлена нет? Оказывается, нет. Такое разложение существует:

$$x^4 + 1 = (x^2 + x - 1)(x^2 - x - 1) = (x^2-1)^2 - x^2 = x^4 - 2x^2 + 1 - x^2 = x^4 + 1$$

(вычисления в характеристике 3).

В чём разница между этими двумя примерами? В первом многочлены $x^2 - \sqrt{2}x + 1$ и $x^2 + \sqrt{2}x + 1$ были неприводимы в своём кольце многочленов. Во втором примере это не так. □

Упражнение 10.88. Найдите корень у многочлена $x^2 - \alpha x + 1$, $\alpha^2 = -1$ в поле $F = \mathbb{F}_3[x]/(x^2+1)$. (Подсказка: не бойтесь применить формулу для корней квадратного уравнения.)

Теперь вернёмся к структуре колец $F[x]/(f)$ для произвольного многочлена f. Основная теорема арифметики в применении к кольцу многочленов с коэффициентами в поле говорит, что любой многочлен разлагается в произведение степеней неприводимых

$$f(x) = g_1(x)^{a_1} \cdot \ldots \cdot g_s(x)^{a_s}, \tag{10.6}$$

g_i — неприводимые многочлены. Это разложение однозначно с точностью до перестановки множителей и умножения на обратимые элементы. Например, можно выделить разложение в произведение константы (многочлена нулевой степени) и каких-то неприводимых нормированных многочленов.

Определение 10.89. Если все $a_i = 1$ в разложении (10.6), то говорят, что многочлен *свободен от квадратов* (не делится на квадрат многочлена) или *не имеет кратных корней*.

Второе название более употребительно, но оно требует объяснений. Говорят, что многочлен f имеет корень a кратности k, если f делится на $(x-a)^k$ (то есть a — корень f), но не делится на $(x-a)^{k+1}$.

В разложении (10.6) могут встретиться степени выше первой неприводимых многочленов степени выше первой. При этом в поле коэффициентов у многочлена вообще нет корней. Однако, как мы увидим позже, у такого многочлена есть кратные корни в бóльшем поле.

Из (10.6) и китайской теоремы об остатках следует изоморфизм колец

$$F[x]/(f) \cong F[x]/(g_1^{a_1}) \oplus \cdots \oplus F[x]/(g_s^{a_s}).$$

Этот изоморфизм выражает структуру общего кольца вычетов по модулю идеала, порождённого произвольным многочленом, через кольца вычетов по модулю степеней неприводимых многочленов. Применение этого изоморфизма упрощает вычисления различных свойств колец вычетов. Например, порядок группы обратимых элементов кольца вычетов находится без особого труда.

Приведём типичные примеры.

Пример 10.90. Сколько обратимых элементов в кольце вычетов $\mathbb{F}_7[x]/(x^2+x-1)$?

Нужно разложить на множители многочлен x^2+x-1 в кольце $\mathbb{F}_7[x]$. Поскольку степень равна 2, вопрос сводится к поиску корней. Решать квадратные уравнения одинаково легко во всех полях характеристики, не равной 2.

Контрольный вопрос 10.91. Проверьте, что в любом поле F, $\operatorname{char} F \neq 2$, корни квадратного уравнения находятся по тем же формулам, что и в случае поля действительных чисел.

Таким образом, вопрос о существовании корней квадратного уравнения сводится к существованию квадратного корня из дискриминанта.

В данном случае $D = 1 - 4 \cdot (-1) = 5$ и это квадратичный невычет по модулю 7. Поэтому корней у многочлена $x^2 + x - 1$ нет. Значит, он неприводим, а кольцо $\mathbb{F}_7[x]/(x^2 + x - 1)$ является полем (из $7^2 = 49$ элементов). В поле обратимы все элементы, кроме 0.

Ответ: 48. □

Пример 10.92. Сколько обратимых элементов в кольце вычетов $\mathbb{F}_7[x]/(x^2-x-2)$?

Разложим многочлен $x^2 - x - 2$ на неприводимые множители:

$$x^2 - x - 2 = (x-2)(x+1)$$

(напомним, что многочлены степени 1 всегда неприводимые). Из китайской теоремы получаем изоморфизм колец

$$\mathbb{F}_7[x]/(x^2-x-2) \cong \mathbb{F}_7[x]/(x-2) \oplus \mathbb{F}_7[x]/(x+1) \cong \mathbb{F}_7 \oplus \mathbb{F}_7.$$

Теперь осталось найти количество обратимых элементов в прямой сумме полей $\mathbb{F}_7 \oplus \mathbb{F}_7$. Пара (x, y) обратима тогда и только тогда, когда $x \neq 0$ и $y \neq 0$. Если одно из этих условий нарушается, то такая пара либо нулевая, либо делитель нуля, то есть необратима. Если же оба условия выполняются, то

$$(x, y) \cdot (x^{-1}, y^{-1}) = (1, 1),$$

что и требовалось, так как $(1, 1)$ — единица в $\mathbb{F}_7 \oplus \mathbb{F}_7$.

Получаем ответ: количество обратимых элементов равно $(7-1) \cdot (7-1) = 36$. □

Пример 10.93. Сколько обратимых элементов в кольце вычетов $\mathbb{F}_7[x]/(x^2+x+2)$?

В данном случае дискриминант квадратного уравнения $x^2 + x + 2 = 0$ равен $1 - 4 \cdot 2 = -7 = 0$ по модулю 7. Это означает, что у многочлена кратные корни. Получаем разложение

$$x^2 + x + 2 = (x - 3)^2$$

в кольце $\mathbb{F}_7[x]$.

Поскольку $f(x) \mapsto f(x + a)$ является изоморфизмом колец, кольца вычетов $\mathbb{F}_7[x]/((x - 3)^2)$ и $\mathbb{F}_7[x]/(x^2)$ изоморфны. Будем искать обратимые элементы во втором из них.

Обозначим α класс вычетов, содержащий x. Тогда $\alpha^2 = 0$, а любой элемент кольца имеет вид $a + b\alpha$, $a, b \in \mathbb{F}_7$.

Если $a = 0$, $b \neq 0$, то элемент $b\alpha$ нильпотентный ($(b\alpha)^2 = 0$), поэтому необратим. А если $a \neq 0$, то элемент $a + b\alpha$ обратим, как показывает прямое вычисление

$$(a + b\alpha) \cdot \frac{a - b\alpha}{a^2} = \frac{a^2 - b^2\alpha^2}{a^2} = 1.$$

Таким образом, обратимыми являются $6 \cdot 7 = 42$ элемента этого кольца. □

Пример 10.94. Есть ли нильпотентные элементы в кольце $\mathbb{F}_{37}[x]/(x^2 + x - 9)$?

Вопрос сводится к разложению $x^2 + x - 9$ на множители. Дискриминант этого многочлена равен $1 - 4 \cdot (-9) = 0$ в $\mathbb{F}_{37}$. Поэтому $x^2 + x - 9 = f(x)^2$, $\deg f = 1$.

Контрольный вопрос 10.95. Найдите явное выражение для этого многочлена f.

Любое кольцо вычетов $F[x]/(f^2)$, $\deg f > 0$, имеет нильпотентные элементы, например, $[f]^2 = [f^2] = 0$. □

Пример 10.96. Сколько решений имеет уравнение $x^6 = 1$ в кольце $\mathbb{F}_7[x]/(x^2+3x-5)$?

Многочлен x^2+3x-5 в кольце $\mathbb{F}_7[x]$ разлагается на разные линейные множители, так как его дискриминант равен $9 + 20 \equiv 1 \pmod 7$.

Поэтому

$$\mathbb{F}_7[x]/(x^2 + 3x - 5) \cong \mathbb{F}_7[x]/(x - a) \oplus \mathbb{F}_7[x]/(x - b), \qquad a \neq b,$$

и вопрос сводится к нахождению числа решений уравнения $x^6 = 1$ в прямой сумме $\mathbb{F}_7 \oplus \mathbb{F}_7$. Пара (a, b) является решением этого уравнения тогда и только тогда, когда $a^6 = 1$ и $b^6 = 1$.

В поле $\mathbb{F}_7$ все ненулевые элементы являются решениями уравнения $x^6 = 1$ (так как 6 — порядок мультипликативной группы этого поля). Поэтому всего есть $6 \cdot 6 = 36$ решений уравнения $x^6 = 1$ в кольце $\mathbb{F}_7[x]/(x^2 + 3x - 5)$. □

11 Конечные поля

В предыдущих главах уже были выяснены многие свойства конечных полей. В этой главе мы закончим классификацию конечных полей. Будет доказано, что для любого p^n, p — простое, существует поле из p^n элементов. Как следует из теоремы 10.67, для этого достаточно доказать существование неприводимого многочлена степени n над полем $\mathbb{F}_p$ вычетов по модулю числа p. Однако мы приведём другое доказательство, которое кажется более полезным для понимания структуры полей. Впрочем, существование неприводимого многочлена степени n над полем $\mathbb{F}_p$ вычетов по модулю числа p также будет доказано.

11.1 Поле разложения многочлена

Рассмотрим многочлен $f \in F[x]$ от одной переменной над некоторым полем. Как мы знаем, количество корней этого многочлена не превосходит его степени. Если количество корней в точности равно степени многочлена, то он разлагается на линейные множители:

$$f(x) = (x - a_1)(x - a_2)\dots(x - a_d), \quad d = \deg f, \quad a_i \neq a_j \text{ при } i \neq j.$$

Бывает и так, что многочлен разлагается на линейные множители, но количество корней у него меньше степени. Простейшим примером является многочлен x^2.

Но в общем случае в разложении многочлена на неприводимые множители встречаются многочлены степени выше 1. Оказывается, этот эффект связан с недостаточным количеством элементов в поле. Его можно устранить, выбрав подходящее *расширение поля*.

Определение 11.1. Поле K называется *расширением* поля F, если $F \subseteq K$.

Вложение полей (инъективный гомоморфизм) определяет вложение (инъективный гомоморфизм) колец многочленов $F[x] \to K[x]$ очевидным образом (коэффициенты из поля F вкладываем в коэффициенты из поля K).

Пример 11.2. Поле действительных чисел $\mathbb{R}$ является расширением поля рациональных чисел $\mathbb{Q}$. Образ многочлена $x^2 - x - 1 \in \mathbb{Q}[x]$ при вложении колец многочленов $\mathbb{Q}[x] \to \mathbb{R}[x]$ равен $x^2 - x - 1$. Теперь мы рассматриваем ± 1 как действительные, а не рациональные числа.

Это может показаться пустым формализмом. Но алгебраические свойства многочлена $x^2 - x - 1$ различны в этих кольцах. В $\mathbb{Q}[x]$ он неприводим, так как не имеет корней и степень равна 2 (см. анализ в примере 9.27). А в кольце $\mathbb{R}[x]$ он раскладывается на линейные множители:

$$x^2 - x - 1 = (x - \phi)(x - \phi^{-1}), \quad \phi = \frac{1 + \sqrt{5}}{2}\,.$$

Это приводит к существенному различию в алгебраических свойствах колец вычетов $\mathbb{Q}[x]/(x^2 - x - 1)$ и $\mathbb{R}[x]/(x^2 - x - 1)$. Первое является полем, а второе — нет, оно изоморфно $\mathbb{R} \oplus \mathbb{R}$ как следует из китайской теоремы. В первом нет делителей нуля, а во втором они есть. □

Пример 11.3. Поле комплексных чисел $\mathbb{C}$ является расширением поля действительных чисел $\mathbb{R}$. Вложение многочленов и в этом случае может изменить их алгебраические свойства. Многочлен x^2+1 неприводим в $\mathbb{R}[x]$ (так как не имеет действительных корней и его степень 2), но разлагается на линейные множители в $\mathbb{C}[x]$:

$$x^2 + 1 = (x + i)(x - i).$$ □

Пример 11.4. Рассмотрим многочлен $x^5 + 1 \in \mathbb{F}_2[x]$. Он приводим, конечно:

$$x^5 + 1 = (x + 1)(x^4 + x^3 + x^2 + x + 1).$$

Но второй сомножитель в этом разложении уже неприводим (см. пример 10.77).

Рассмотрим однако расширение $\mathbb{F}_2 \subset F = \mathbb{F}_2[x]/(x^4 + x^3 + x^2 + x + 1)$. В этом поле многочлен x^5+1 уже раскладывается на линейные множители. Действительно, мультипликативная группа поля F — циклическая группа из 15 элементов. Поэтому в ней есть подгруппа из 5 элементов. Каждый элемент этой подгруппы является корнем многочлена $x^5 + 1$ (вспомним, что в поле характеристики 2 выполняется равенство $1 = -1$). Получаем разложение $x^5 + 1$ на линейные множители. □

Все эти примеры можно обобщить.

Теорема 11.5. *Для любого многочлена $f \in F[x]$ из кольца многочленов над полем F существует такое расширение $K \supseteq F$ (*поле разложения многочлена*), что f разлагается на линейные множители в кольце $K[x]$.*

Доказательство. Построим цепочку расширений полей

$$F = F_0 \subset F_1 \subset \dots$$

по следующему правилу. Разложим f на неприводимые множители в поле F_i. Если все эти множители имеют степень 1, то $F_i = K$ и построение цепочки закончено на поле разложения многочлена. Если в разложении есть неприводимый множитель, назовём его g, степени больше 1, то следующее поле в цепочке определим как

$$F_{i+1} = F_i[x]/(g).$$

(Напомним, что кольцо вычетов по модулю неприводимого многочлена является полем, так как идеал, порождённый эти многочленом, максимальный.)

Итак, цепочка расширений заканчивается на поле разложения. Осталось доказать, что процесс построения цепочки расширений всегда останавливается. Для этого заметим, что многочлен g в поле F_{i+1} обязательно имеет корень — это класс вычетов, содержащий x. Действительно, пусть $g = \sum_k g_k x^k$, $g_k \in F_i$. Элементам поля F_i в кольце $F_i[x]$ отвечают многочлены нулевой степени и 0. Поэтому

$$g([x]) = \sum_k g_k [x]^k = \sum_k [g_k] \cdot [x^k] = \left[\sum_k g_k x^k\right] = [0],$$

так как мы рассматриваем вычеты по модулю идеала, порождённого g.

Таким образом, количество множителей степени 1 в разложении многочлена f на неприводимые множители увеличивается при переходе от поля F_i к полю F_{i+1}

хотя бы на один. Но общее количество множителей степени 1 не превосходит степени многочлена (логарифмическое свойство степени). Поэтому процесс построения расширений рано или поздно остановится. □

Замечание 11.6. Эта теорема объясняет, почему многочлены, не делящиеся на квадраты, допустимо называть многочленами без кратных корней. Если перейти к полю разложения многочлена, но эти понятия совпадут.

Пример 11.7. Найдём поле разложения многочлена $x^2+1 \in \mathbb{F}_7[x]$. Так как $(-1)^3 = -1$, то -1 является квадратичным невычетом по модулю 7. Значит, многочлен x^2+1 не имеет корней в $\mathbb{F}_7[x]$ и потому является неприводимым в $\mathbb{F}_7[x]$. В поле $\mathbb{F}_7[x]/(x^2+1)$ из $7^2 = 49$ элементов этот многочлен имеет два корня: $\pm[x]$ (напомним, что $[x]$ — класс вычетов, содержащий многочлен x).

Это и есть поле разложения многочлена $x^2+1 \in \mathbb{F}_7[x]$. □

11.2 Основная конструкция

Пусть $q = p^n$. Рассмотрим многочлен $x^q - x \in \mathbb{F}_p[x]$. В некотором расширении F поля $\mathbb{F}_p$ (поле разложения) он раскладывается на линейные множители.

Утверждение 11.8. *Корни многочлена $x^q - x$ в поле разложения образуют поле.*

Для доказательства нам понадобится тождество, которое выполняется в любом поле характеристики p.

Лемма 11.9 (бином двоешника)**.** $(x+y)^p = x^p + y^p$.

Доказательство. В любом коммутативном кольце верна обычная формула для бинома

$$(a+b)^p = a^p + \binom{p}{1}a^{p-1}b + \cdots + \binom{p}{p-1}ab^{p-1} + b^p.$$

Чтобы доказать тождество из утверждения, достаточно проверить, что все $\binom{p}{k}$, $k \notin \{0, p\}$, делятся на p. Запишем формулу для биномиального коэффициента:

$$\binom{p}{k} = \frac{p!}{k!(p-k)!} = \frac{p\cdot(p-1)\cdot\ldots\cdot 1}{k!(p-k)!}.$$

Так как p — простое число, числитель дроби делится на p, а знаменатель — нет. В самом деле, разлагая сомножители знаменателя в произведение простых, видим, что каждый простой делитель знаменателя не превосходит максимума из k и $p-k$, то есть меньше p. □

Доказательство утверждения 11.8. Индукцией по n из бинома двоешника (утверждение 11.9) получается также тождество

$$(x+y)^{p^n} = (x^p + y^p)^{p^{n-1}} = \cdots = (x^{p^{n-1}} + y^{p^{n-1}})^p = x^{p^n} + y^{p^n}.$$

Поэтому для любой пары корней α, β многочлена $x^q - x$ выполняются равенства $(\alpha+\beta)^q = \alpha^q + \beta^q$, $(\alpha\beta)^q = \alpha^q\beta^q$. Отсюда получаем замкнутость множества корней многочлена $x^q - x$ относительно сложения и умножения. 0 и 1 принадлежат множеству корней по очевидным причинам. Но тогда множество корней замкнуто и относительно взятия обратных как по сложению, так и по умножению, поскольку множество корней ненулевого многочлена конечно. □

Итак, мы построили поле. Но сколько в нём элементов? Не больше q, так как степень многочлена $x^q - x$ равна q. Оказывается, что этот многочлен не делится на квадраты и потому количество различных его корней в точности равно q. Это и означает, что существует поле из q элементов.

Для завершения доказательства потребуется ещё одно техническое средство: производная многочлена.

11.3 Производная многочлена

Понятие производной принадлежит не только анализу, но и алгебре. Оно оказывается полезным для характеризации многочленов, не делящихся на квадраты.

Определение 11.10. *Производная* многочлена определяется как образ отображения $D\colon F[x] \to F[x]$, задаваемого следующими свойствами:

- $D(c) = 0$ для любой константы c (многочлена степени 0);
- $D(x^n) = nx^{n-1}$ (здесь n означает элемент поля коэффициентов F, который равен сумме n единиц поля);
- производная линейна: для любых $\alpha, \beta \in F$, $f, g \in F[x]$ выполняется равенство
$$D(\alpha f + \beta g) = \alpha D(f) + \beta D(g).$$

Лемма 11.11. *Указанные в определении свойства однозначно задают отображение* $D\colon F[x] \to F[x]$.

Доказательство. Мономы образуют базис в пространстве многочленов как векторном пространстве над полем F: любой многочлен однозначно представляется как линейная комбинация мономов. Поэтому значение производной на любом многочлене определяется из свойств, указанных в определении:

$$D\left(\sum_{i=0}^{d} \alpha_i x^i\right) = \sum_{i=0}^{d} \alpha_i D(x^i) = \sum_{i=1}^{d} i\alpha_i x^{i-1}.$$

Для завершения доказательства нужно заметить, что определённое таким образом отображение обладает всеми указанными в определении свойствами. □

Лемма 11.12 (формула Лейбница)**.** *Для любых многочленов* f, g *выполнено* $D(fg) = D(f)g + fD(g)$.

Доказательство. Если один из сомножителей равен 1, то $(f \cdot 1)' = f' = f' \cdot 1 + f \cdot 1'$ (производная константы — нуль). Для мономов x^n, x^k имеем

$$(x^n \cdot x^k)' = (n+k) \cdot x^{n+k-1} = nx^{n-1}x^k + kx^n x^{k-1} = (x^n)'x^k + x^n(x^k)'.$$

В остальных случаях формула выполняется в силу линейности производной.

Пусть формула Лейбница выполняется для пар многочленов f_1, g и f_2, g. Тогда она выполняется и для любой пары $\alpha f_1 + \beta f_2$, g ($\alpha, \beta \in F$):

$$\begin{aligned} D((\alpha f_1 + \beta f_2)g) = \\ = \alpha D(f_1 g) + \beta D(f_2 g) = \alpha D(f_1)g + \alpha f_1 D(g) + \beta D(f_2)g + \beta f_2 D(g) = \\ = D(\alpha f_1 + \beta f_2)g + (\alpha f_1 + \beta f_2)D(g). \end{aligned}$$

Используя это соображение вначале докажем, что формула выполняется для любой пары f, x^n (так как мономы образуют базис пространства многочленов). После этого для любого фиксированного f докажем, что формула Лейбница выполняется для любой пары f, g. □

Утверждение 11.13. *Пусть f и $D(f)$ взаимно просты. Тогда f не имеет кратных корней (не делится на квадрат).*

Доказательство. Предположим противное: $f = g^2 h$, $\deg g > 0$. Тогда по формуле Лейбница $D(f) = 2D(g)gh + g^2 D(h)$. Значит, g является общим делителем f и $D(f)$. Пришли к противоречию. □

Пример 11.14. Докажем, что поле разложения многочлена $x^7 - 1 \in \mathbb{F}_5[x]$ содержит больше 1000 элементов.

Обозначим это поле F, его характеристика 5. Если оно конечно, то содержит 5^n элементов. У многочлена $x^7 - 1$ в этом поле 7 различных корней, так как он взаимно прост со своей производной $7x^6$.

Пусть $\alpha \neq 1$ — один из корней этого многочлена. Так как 7 — простое число, порядок α в мультипликативной группе поля F равен 7 и является делителем порядка группы, то есть $5^n - 1$. Найдём минимальное n, для которого $7 \mid 5^n - 1$. Это порядок 5 в мультипликативной группе вычетов по модулю 7. Так как $7 \nmid 5^1 - 1 = 4$; $7 \nmid 5^2 - 1 = 24$; $7 \nmid 5^3 - 1 = 124$, этот порядок равен 6.

Итак, в поле разложения многочлена $x^7 - 1 \in \mathbb{F}_5[x]$ не менее $5^6 > 1000$ элементов. □

Обратное к утверждению 11.13 также верно.

Теорема 11.15. *Тогда $f \in F[x]$ не имеет кратных корней тогда и только тогда, когда f и $D(f)$ взаимно просты.*

Доказательство. В одну сторону это утверждение 11.13.

В другую сторону достаточно доказать для многочлена, который разлагается на линейные множители: перейдём к полю разложения; поскольку коэффициенты производной принадлежат исходному полю, то и наибольший общий делитель f и Df те же самые, что и в исходном поле.

Итак, пусть

$$f(x) = (x - a_1)(x - a_2) \cdot \ldots \cdot (x - a_d),$$

где все a_i различны. Запишем производную, пользуясь правилом Лейбница:

$$Df(x) = (x - a_2) \cdot \ldots \cdot (x - a_d) + (x - a_1)(x - a_3) \cdot \ldots \cdot (x - a_d) + \ldots \\ \cdots + (x - a_1)(x - a_2) \cdot \ldots \cdot (x - a_{d-1}). \quad (11.1)$$

В каждом слагаемом пропущен ровно один сомножитель из разложения многочлена f.

В силу единственности разложения на простые в евклидовом кольце достаточно проверить, что Df не делится ни на один неприводимый делитель f, то есть на $(x - a_i)$. Рассмотрим, не ограничивая общности, $x - a_1$ (с остальными делителями $x - a_i$ рассуждение аналогично). Все слагаемые в сумме (11.1) кроме первого делятся на $x - a_1$. А первое — нет, так как

$$(a_1 - a_2) \cdot \ldots \cdot (a_1 - a_d) \neq 0$$

(все корни различны). □

Как уже неоднократно упоминалось, свойство иметь кратные корни зависит от поля.

Пример 11.16. Имеет ли кратные корни многочлен $(x + 1)(x - 9)$? Ответ зависит от поля коэффициентов. Скажем, как многочлен из $\mathbb{Q}[x]$ этот многочлен не имеет кратных корней: $-1 \neq 9$ в $\mathbb{Q}$.

Однако если рассмотреть этот многочлен как элемент кольца $F[x]$, где F — поле из 25 элементов, этот многочлен оказывается квадратом линейного многочлена, так как $-1 = 9$ в поле характеристики 5. □

Приведём пример использования критерия отсутствия кратных корней.

Пример 11.17. Есть ли нильпотентные элементы в кольце $\mathbb{F}_7[x]/(x^3 + 4x - 2)$?

Проверим, есть ли у многочлена $x^3 + 4x - 2$ кратные корни. Вычислим производную $D(f) = 3x^2 + 4$. Так как $3^{-1} = -2$ в поле $\mathbb{F}_7$, то достаточно искать наибольший общий делитель у многочлена f и $-2D(f) = x^2 - 8 = x^2 - 1 = (x + 1)(x - 1)$.

Нетрудно видеть, что ± 1 не являются корнями многочлена $x^3 + 4x - 2$. Поэтому он взаимно прост со своей производной и не имеет кратных корней.

Но тогда в кольце $\mathbb{F}_7[x]/(x^3 + 4x - 2)$ нет нильпотентных элементов: из того, что g^k делится на $x^3 + 4x - 2$ следует, что g делится на каждый неприводимый делитель $x^3 + 4x - 2$, а, значит, и на сам $x^3 + 4x - 2$. □

Завершение доказательства существования конечного поля. Теперь мы можем закончить рассуждение с основной конструкцией. Нам ведь осталось показать, что $x^q - x$ не имеет кратных корней. Для этого вычислим производную

$$D(x^q - x) = qx^{q-1} - 1 = p^n x^{q-1} - 1 = -1$$

(напомним, что характеристика поля равна p).

Итак, многочлен $x^q - x$ взаимно прост со своей производной -1 и потому не имеет кратных корней. Его корни образуют поле, в котором ровно q элементов. □

Мы уже нашли возможные значения размеров конечных полей: это степени простых чисел. Но справедливы более сильные утверждения. Все конечные поля одинакового размера изоморфны, и про каждое поле можно определить, какие подполя оно содержит.

Эти утверждения связаны также с разложением многочлена $x^{p^n} - x$ на неприводимые множители в кольце $\mathbb{F}_p[x]$ и симметриями (автоморфизмами) полей.

В следующих разделах этой главы они разбираются подробнее.

11.4 Минимальный многочлен

Уже несколько раз у нас встречался гомоморфизм значения. Рассмотрим его ещё раз. Пусть $\alpha \in F$ — элемент конечного поля характеристики p. Напомним, что гомоморфизм значения $\mathrm{Ev}_\alpha \colon \mathbb{F}_p[x] \to F$ сопоставляет многочлену $f \in \mathbb{F}_p[x]$ его значение $f(\alpha)$ в точке α. (Такой гомоморфизм определён для любого элемента любого расширения поля.)

Ядро этого гомоморфизма порождено некоторым многочленом. Он называется *минимальным* многочленом элемента α. Обозначать минимальный многочлен будем $m_\alpha(x)$.

Утверждение 11.18. *Минимальный многочлен неприводим.*

Доказательство. Пусть $m_\alpha(x) = f(x)g(x)$. Так как $m_\alpha(\alpha) = 0$ (это и означает, что m_α принадлежит ядру гомоморфизма значения), то $f(\alpha) = 0$ или $g(\alpha) = 0$. Значит, один из сомножителей принадлежит ядру гомоморфизма Ev_α, то есть делится на m_α. Поэтому степень этого сомножителя совпадает со степенью m_α, а степень второго равна нулю. Это и означает неприводимость многочлена m_α. □

По теореме о гомоморфизмах колец кольцо вычетов $\mathbb{F}_p[x]/\operatorname{Ker}\mathrm{Ev}_\alpha$ изоморфно образу кольца многочленов при гомоморфизме значения. Так как минимальный многочлен неприводим, порождённый им идеал максимальный, а кольцо вычетов по модулю этого идеала является полем. Таким образом, значения многочленов $f(\alpha)$, $f \in \mathbb{F}_p[x]$, образуют подполе поля F. Обозначим его $\mathbb{F}_p(\alpha)$.

Нетрудно видеть, что $\mathbb{F}_p(\alpha)$ — наименьшее подполе, содержащее элемент α. Действительно, из замкнутости относительно умножения такое подполе обязано содержать все степени α^k элемента α, а из замкнутости относительно сложения — все мономы

$$s\alpha^k = \underbrace{\alpha^k + \alpha^k + \cdots + \alpha^k}_{s \text{ штук}}, \quad s \in \mathbb{Z}_+,$$

и их линейные комбинации. Но это и есть все элементы поля $\mathbb{F}_p(\alpha)$.

Соотношение между минимальными многочленами и минимальными подполями, содержащими данный элемент, помогает в разных ситуациях.

Пример 11.19. Пусть известно, что минимальный многочлен элемента α некоторого конечного поля имеет степень 6. Какова может быть степень минимального многочлена элемента $\alpha + 1$?

Это легко понять из равенства $\mathbb{F}_p(\alpha) = \mathbb{F}_p(\alpha + 1)$ (p — характеристика поля), которое следует из того, что $\alpha \in \mathbb{F}_p(\alpha+1)$ (значит, $\mathbb{F}_p(\alpha) \subseteq \mathbb{F}_p(\alpha+1)$) и $\alpha+1 \in \mathbb{F}_p(\alpha)$ (значит, $\mathbb{F}_p(\alpha + 1) \subseteq \mathbb{F}_p(\alpha)$).

Количество элементов в $\mathbb{F}_p(\alpha)$ равно p^d, где d — степень минимального многочлена: ведь $\mathbb{F}_p(\alpha) \cong \mathbb{F}_p[x]/(m_\alpha)$, а количество элементов в любом кольце вычетов по модулю многочлена степени d равно p^d.

То же самое справедливо и для поля $\mathbb{F}_p(\alpha + 1)$. Раз эти поля совпадают, что степени минимальных многочленов также совпадают. □

Пример 11.20. В поле $\mathbb{F}_{49}$ из 49 элементов как векторном пространстве над полем $\mathbb{F}_7$ из 7 элементов элементы 1, α образуют базис. Докажем, что $\alpha^2 \neq \alpha - 2$.

Действительно, из линейной независимости 1 и α следует, что $\alpha \notin \mathbb{F}_7$. Многочлен $x^2 - x + 2 \in \mathbb{F}_7[x]$ имеет кратный корень:

$$x^2 - x + 2 = (x + 3)^2.$$

Поэтому из $\alpha^2 = \alpha - 2$ следовало бы, что $\alpha + 3 = 0$, то есть $\alpha \in \mathbb{F}_7$. Значит, $\alpha^2 \neq \alpha - 2$. □

11.5 Разложение многочлена $x^q - x$

Рассмотрим поле F из $q = p^n$ элементов, p простое. Многочлен $x^q - x = x(x^{q-1} - 1)$ в этом поле разлагается на линейные множители: он делится на x и на любой $x - \alpha$, $\alpha \neq 0$, так как $\alpha^{q-1} = 1$ (теорема Лагранжа для мультипликативной группы поля F).

Сейчас мы определим неприводимые делители этого многочлена в кольце многочленов над простым полем характеристики p.

Утверждение 11.21. *Многочлен $x^q - x$ делится на любой неприводимый многочлен $f \in \mathbb{F}_p[x]$ степени n.*

Доказательство. Пусть $f \in \mathbb{F}_p[x]$ — какой-нибудь неприводимый многочлен степени n. Тогда кольцо вычетов $\mathbb{F}_p[x]/(f)$ является полем из $q = p^n$ элементов. Рассмотрим минимальный многочлен m_α элемента $\alpha = [x]$ этого поля, задаваемого вычетом, содержащим многочлен x.

Непосредственным вычислением проверяется, что $f \in \operatorname{Ker} \mathrm{Ev}_\alpha$:

$$f(\alpha) = \sum_i f_i \alpha^i = \sum_i f_i [x^i] = \left[\sum_i f_i x^i \right] = [f] = [0] \bmod (f).$$

Поэтому f делится на m_α. Так как f неприводим, то он равен m_α с точностью до умножения на константу (многочлен нулевой степени).

Как уже сказано выше, $\alpha^q = \alpha$ (это элемент поля из q элементов). Значит, $x^q - x \in \operatorname{Ker} \mathrm{Ev}_\alpha = (f)$, что и требовалось доказать. □

Чтобы закончить описание неприводимых делителей $x^q - x$ в кольце $\mathbb{F}_p[x]$, нам потребуется следующее наблюдение о подполях конечного поля, аналогичное условию на количество элементов в конечном поле.

Утверждение 11.22. *Пусть $K \subset F$ — конечные поля характеристики p, $|K| = p^k$, $|F| = p^n$. Тогда k делит n.*

Доказательство. Заметим, что F является конечномерным векторным пространством над полем K. Значит, в нём $|K|^d$ элементов, где d — размерность этого пространства. Отсюда следует равенство $n = kd$. □

Утверждение 11.23. *Пусть g — неприводимый делитель $x^q - x$ в кольце $\mathbb{F}_p[x]$, $q = p^n$. Тогда $\deg g$ делит n.*

Доказательство. Рассмотрим поле F из q элементов. В этом поле многочлен $x^q - x$ раскладывается на линейные множители. Поэтому любой его делитель имеет в этом поле корень.

Пусть α — корень g. Аналогично доказательству утверждения 11.21 проверяем, что g является минимальным многочленом α. Значит, подполе $\mathbb{F}_p(\alpha)$ содержит p^d элементов, $d = \deg g$. Из утверждения 11.22 следует, что d делит n. □

Теперь мы готовы доказать теорему о разложении $x^q - x$ в кольце $\mathbb{F}_p[x]$ на неприводимые множители.

Теорема 11.24. *Неприводимые делители многочлена $x^q - x$ в кольце $\mathbb{F}_p[x]$, $q = p^n$, — это в точности неприводимые многочлены, степени которых являются делителями n.*

Доказательство. В одну сторону утверждение теоремы — это утверждение 11.23. Осталось показать, что любой неприводимый многочлен степени $d \mid n$ является делителем $x^q - x$.

Случай $d = n$ разобран в утверждении 11.21. Общий случай сводится к нему, если убедиться, что из $d \mid n$ следует $x^{p^d} - x \mid x^{p^n} - x$. Поэтому неприводимый многочлен $g \in \mathbb{F}_p[x]$ степени d является не только делителем $x^{p^d} - x$ (по утверждению 11.21), но и делителем $x^{p^n} - x$.

Делимость $x^{p^d} - x \mid x^{p^{dk}} - x$ равносильна делимости $x^{p^d-1} - 1 \mid x^{p^{dk}-1} - 1$. Легко видеть, что $p^d - 1 \mid p^{kd} - 1$ (формула для суммы геометрической прогрессии):

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

Из той же формулы следует и искомая делимость многочленов:

$$x^{p^{dk}-1} - 1 = (x^{p^d-1} - 1)(x^{(N-1)(p^d-1)} + \cdots + x^{p^d-1} + 1),$$

здесь через N обозначено $(p^{kd} - 1)/(p^d - 1)$. □

Пример 11.25. Делится ли многочлен $x^{1023} - 1$ на $x^4 + x + 1$ в поле характеристики 2?

Прежде всего заметим, что у обоих многочленов коэффициенты принадлежат полю $\mathbb{F}_2$. Поэтому если первый многочлен делится на второй, то эта делимость выполняется и в кольце $\mathbb{F}_2[x]$. Но в этом кольце применение теоремы 11.24 к многочлену $x^{1024} - x = x(x^{1023} - 1)$ даёт следующую характеризацию неприводимых делителей $x^{1023} - 1$: это неприводимые в $\mathbb{F}_2[x]$ многочлены степени 10, 5, 2 и $x + 1$ (множитель x исключён, как видно из написанного выше разложения).

Многочлен $x^4 + x + 1$ неприводим в $\mathbb{F}_2[x]$ и его степень равна 4. Поэтому он не делит $x^{1023} - 1$. □

Теорема 11.24 позволяет находить степени неприводимых делителей многочленов без явного предъявления этих делителей.

Пример 11.26. Найдём степени неприводимых делителей многочлена $x^7 - 5 \in \mathbb{F}_{37}[x]$.

Во-первых, заметим, что $(7, 37 - 1) = 1$. Поэтому возведение в 7-ю степень является изоморфизмом мультипликативной группы поля $\mathbb{F}_{37}$. Это означает, что в $\mathbb{F}_{37}$ у многочлена $x^7 - 5$ ровно один корень, обозначим его α.

Теперь найдём количество корней многочлена $x^7 - 5$ в квадратичном расширении $\mathbb{F}_{37^2}$. Порядок $\mathbb{F}^*_{37^2}$ равен $37^2 - 1 = 36 \cdot 38$ и он взаимно прост с 7. Так что и в этом поле возведение в 7-ю степень является изоморфизмом мультипликативной группы поля. Корень по-прежнему один (и он лежит в подполе $\mathbb{F}_{37}$).

Так как $37^3 - 1 = 36 \cdot (37^2 + 37 + 1) \equiv 1 \cdot (2^2 + 2 + 1) = 7 \equiv 0 \pmod 7$, порядок группы $\mathbb{F}^*_{37^3}$ делится на 7. Эта группа циклическая, поэтому в ней есть подгруппа 7-го порядка. Если умножить α на любой элемент u этой подгруппы, получим корень многочлена $x^7 - 5$: $(\alpha u)^7 = \alpha^7 u^7 = 5 \cdot 1 = 5$.

Значит, в поле $\mathbb{F}_{37^3}$ у многочлена $x^7 - 5$ есть ровно 7 корней. Степени минимальных многочленов всех корней, отличных от α, равны 3 (это делители 3 и они отличны от 1). Каждый такой многочлен раскладывается на линейные множители над полем $\mathbb{F}_{37^3}$ (так как является делителем $x^{37^3} - x$). Значит, 6 корней многочлена $x^7 - 5$, отличных от α, разбиваются на две группы, каждая из которых образует корни некоторого неприводимого многочлена степени 3 в кольце $\mathbb{F}_{37}[x]$.

Окончательно получаем, что $x^7 - 5$ в кольце $\mathbb{F}_{37}[x]$ раскладывается в произведение неприводимых многочленов степеней 1, 3, 3. □

Из теоремы 11.24 следуют два важных свойства конечных полей, которые приведены в двух следующих разделах.

11.6 Изоморфизм полей с одинаковым количеством элементов

Теорема 11.27. *Пусть F_1, F_2 — поля характеристики p, в каждом из которых $q = p^n$ элементов. Тогда эти поля изоморфны: $F_1 \cong F_2$.*

Доказательство. Как уже доказано выше, любое конечное поле изоморфно кольцу вычетов по модулю идеала кольца $\mathbb{F}_p[x]$, порождённого неприводимым многочленом. Пусть $F_2 \cong \mathbb{F}_p[x]/(f)$, $\deg f = n$ (так как в поле p^n элементов).

Многочлен f неприводим и его степень равна n. По утверждению 11.21 он является делителем многочлена $x^q - x \in \mathbb{F}_p[x]$. Последний раскладывается в поле

F_1 (в котором q элементов) на линейные множители. Значит, в F_1 у многочлена f есть корень α, $f(\alpha) = 0$. Аналогично предыдущим случаям проверяем, что f — минимальный многочлен α. Поэтому подполе $\mathbb{F}_p(\alpha) \cong \mathbb{F}_p[x]/(f) \cong F_2$ содержит ровно $q = p^n$ элементов, то есть совпадает с F_1. Значит, поля F_1 и F_2 изоморфны. □

Теперь обозначение поля из q элементов как $\mathbb{F}_q$ становится понятнее: с точностью до изоморфизма есть ровно одно такое поле, поэтому достаточно указывать лишь количество элементов в нём.

Приведём примеры вычислений изоморфизмов, о которых говорится в теореме 11.27.

Пример 11.28. Изоморфны ли кольца вычетов $\mathbb{F}_7[x]/(x^2 - 2x + 2)$ и $\mathbb{F}_7[x]/(x^2 + 1)$? Степени многочленов 2, так что для разложения их на неприводимые множители нужно решить квадратные уравнения в $\mathbb{F}_7$.

Для первого многочлена дискриминант равен $4 - 8 = -4 = 2^2 \cdot (-1)$; для второго — такой же ($0 - 4 = -4$). Как нетрудно проверить, -1 является квадратичным невычетом по модулю 7: $(-1)^{(7-1)/2} = (-1)^3 = -1$. Значит, оба многочлена корней не имеют и потому неприводимы. Оба кольца вычетов являются полями, содержащими $7^2 = 49$ элементов.

Построим теперь изоморфизм этих колец явно. Для этого нужно в одном из них найти решение квадратного уравнения, задаваемого многочленом, порождающим второе.

Давайте искать в кольце $\mathbb{F}_7[x]/(x^2 - 2x + 2)$ корень уравнения $t^2 + 1 = 0$. Преобразуем порождающий:

$$x^2 - 2x + 2 = (x - 1)^2 + 1.$$

Итак, в первом кольце $[(x-1)^2+1] = [x-1]^2+[1] = [0]$. Поэтому один из корней равен $x - 1$ (второй ему противоположен, но для построения изоморфизма нам годится любой из корней).

Задаём изоморфизм $\varphi\colon \mathbb{F}_7[x]/(x^2 + 1) \to \mathbb{F}_7[x]/(x^2 - 2x + 2)$, как указано в теореме 11.27:

$$\varphi([1]) = [1]; \quad \varphi([x]) = [x - 1].$$

Обратите внимание, что квадратные скобки в левых и правых частях этих равенств означают разное. В левых частях они указывают на класс вычетов в кольце $\mathbb{F}_7[x]/(x^2 - 2x + 2)$, в правых частях — на класс вычетов в кольце $\mathbb{F}_7[x]/(x^2 + 1)$. Скажем, первое равенство не изменится, если его переписать в виде

$$\varphi([-x^2]) = [x^2 - 2x + 3].$$

Для остальных элементов кольца изоморфизм доопределяется по линейности:

$$\varphi([a + bx]) = [a] + [b(x - 1)] = [(a - b) + bx].$$ □

Контрольный вопрос 11.29. Найдите обратный изоморфизм

$$\psi\colon \mathbb{F}_7[x]/(x^2 - 2x + 2) \to \mathbb{F}_7[x]/(x^2 + 1).$$

11.7 Подполя конечных полей

Как доказано в утверждении 11.22, если $F_1 \subset F_2$ — два конечных поля, то $|F_2| = |F_1|^d$ для некоторого целого положительного d (размерности F_2 как векторного пространства над полем F_1).

Это необходимое условие является и достаточным. Теорема 11.24 говорит, что для любого делителя k числа n любой неприводимый многочлен $f \in \mathbb{F}_p[x]$ степени k является делителем $x^{p^n} - x$. Так как последний раскладывается в $\mathbb{F}_q$ на линейные множители, у многочлена f есть корень α в $\mathbb{F}_q$. Поле $\mathbb{F}_p(\alpha) \cong \mathbb{F}_p[x]/(f)$ содержит p^k элементов и является подполем F_{p^n}.

Заметим, что подполе из p^k элементов в поле из p^{kd} элементов если и существует, то единственно. Элементы любого такого подполя являются корнями многочлена $x^{p^k} - x$: нуль очевидно является корнем этого многочлена, а для любого ненулевого элемента α в силу теоремы Лагранжа для мультипликативной группы подполя F выполняется равенство $\alpha^{p^k-1} = 1$. Поскольку у многочлена степени p^k не больше p^k корней, то подполе данного размера единственно (если вообще существует).

Таким образом, конечные поля характеристики p образуют однозначную с точностью до изоморфизма структуру: $\mathbb{F}_p \subset \mathbb{F}_{p^k}$; $\mathbb{F}_{p^2} \subset \mathbb{F}_{p^{2k}}$ и т.д.

Рассмотрим несколько задач о подполях конечных полей.

Пример 11.30. Существует ли в поле из 64 элементов подполе из 16 элементов?

Ответ: нет, не существует. Нарушается указанное выше необходимое условие: $16 = 2^4$, $64 = 2^6$, поэтому $64 = 16^{3/2}$. □

Пример 11.31. Докажем, что существует такой многочлен $f \in \mathbb{F}_2[x]$ степени 10, что в его поле разложения не менее 2^{30} элементов.

Возьмём многочлен, который является произведением неприводимых многочленов степеней 2, 3 и 5. В поле разложения такого многочлена есть подполя из 2^2, 2^3 и 2^5 элементов. Таким образом, если размерность этого поля над простым полем $\mathbb{F}_2$ равна n, то n обязано делится на 2, 3, 5. То есть, n делится на 30. □

Пример 11.32. Сколько решений имеет уравнение $x^2 + x - 1 = 0$ в поле из 243 элементов?

Характеристика этого поля равна 3, а многочлен $x^2 + x - 1$ неприводим в $\mathbb{F}_3$ (так как его степень равна 2, то достаточно проверить, что у этого многочлена нет корней в $\mathbb{F}_3$).

Значит, минимальный многочлен корня многочлена $x^2 + x - 1$ совпадает с ним самим и порождает поле из $3^2 = 9$ элементов. Так как $243 = 3^5$, в поле $\mathbb{F}_{243}$ нет подполя из 9 элементов.

Ответ: 0 решений. □

11.8 Автоморфизм Фробениуса

У конечных полей есть симметрии (автоморфизмы). Примером такого автоморфизма является *автоморфизм Фробениуса*

$$F\colon x \mapsto x^p, \quad p \text{ — характеристика поля.} \tag{11.2}$$

Утверждение 11.33. *Отображение F является автоморфизмом.*

Доказательство. То, что это отображение является гомоморфизмом, уже фактически проверялось выше. Сохранение умножения следует из коммутативности умножения

$$F(xy) = (xy)^p = x^p y^p = F(x)F(y).$$

Сохранение сложения следует из линейности возведения в степень p, формула (11.9):

$$F(x+y) = (x+y)^p = x^p + y^p = F(x) + F(y).$$

Ядро F нулевое, так как из $x^p = 0$ следует $x = 0$ (в поле делителей нуля нет).

Осталось заметить, что инъективное отображение конечного множества в себя является биективным (взаимно однозначным). □

У автоморфизма Фробениуса есть неподвижные точки и это в точности элементы простого подполя.

Утверждение 11.34. *Если $F(x) = x$, то $x \in \mathbb{F}_p$.*

Доказательство. Условие $F(a) = a$ равносильно тому, что a является корнем многочлена $x^p - x$. Элементы простого поля являются корнями этого многочлена. Других корней нет, так как у многочлена над полем количество корней не превосходит степени. □

Автоморфизм Фробениуса является линейным отображением поля $\mathbb{F}_{p^n}$ как векторного пространства над простым полем $\mathbb{F}_p$. Это, в частности, позволяет решать квадратные уравнения в полях характеристики 2 (в этих полях обычная формула для корней квадратного уравнения не работает, так как в её знаменатель входит $2 = 0$ в поле характеристики 2).

Пример 11.35. Решим уравнение $x^2 + \alpha x + \alpha^3 = 0$ в поле $\mathbb{F}_{16}$, если известно, что минимальный многочлен для α равен $x^4 + x + 1 = 0$.

Это уравнение равносильно системе линейных уравнений в $\mathbb{F}_2^4$. Выберем базис 1, α, α^2, α^3 в поле $\mathbb{F}_{16}$ и обозначим (x_0, x_1, x_2, x_3) координаты x в этом базисе, $x_i \in \mathbb{F}_2$. Тогда

$$\begin{aligned} x^2 &= (x_0 + x_1\alpha + x_2\alpha^2 + x_3\alpha^3)^2 = x_0 + x_1\alpha^2 + x_2\alpha^4 + x_3\alpha^6 = \\ &= x_0 + x_1\alpha^2 + x_2(\alpha+1) + x_3\alpha^2(\alpha+1) = \\ &= (x_0 + x_2) + x_2\alpha + (x_1 + x_3)\alpha^2 + x_3\alpha^3; \\ \alpha x &= \alpha(x_0 + x_1\alpha + x_2\alpha^2 + x_3\alpha^3) = \alpha x_0 + x_1\alpha^2 + x_2\alpha^3 + x_3(1+\alpha) = \\ &= x_3 + (x_0 + x_3)\alpha + x_1\alpha^2 + x_2\alpha^3. \end{aligned}$$

Получаем систему линейных уравнений

$$\begin{array}{rrrr} x_0 & & +x_2 + x_3 & = 0 \\ x_0 & & +x_2 + x_3 & = 0 \\ & +2x_1 & + x_3 & = 0 \\ & & +x_2 + x_3 & = 1 \end{array}$$

Эту систему всегда можно решить методом Гаусса. Но в данном случае легко выписать решения прямо из вида системы (ведь $2 = 0$ в поле характеристики 2):

$$x_0 = x_2 = 1, \quad x_3 = 0, \quad x_1 = t,\ t \in \mathbb{F}_2.$$

То есть у уравнения два корня: $1 + \alpha^2$ и $1 + \alpha + \alpha^2$. □

Рассмотрим ещё пару примеров, в которых используется линейность автоморфизма Фробениуса (то есть, фактически, «бином двоешника»).

Пример 11.36. Сколько решений у уравнения $x^{25}+x^5+x = 1$ в поле из 125 элементов?

Характеристика поля равна 5 (так как $125 = 5^3$). Возведение в степень 5 и в степень 25 — линейные отображения поля $\mathbb{F}_{125}$ как 3-мерного векторного пространства над простым полем $\mathbb{F}_5$. Поэтому и отображение $A\colon x \mapsto x^{25} + x^5 + x$ является линейным отображением.

Если $y = x^{25} + x^5 + x$, то $y^5 = x^{125} + x^{25} + x^5 = x + x^{25} + x^5 = y$. Значит, образ отображения A лежит в подполе $\mathbb{F}_5 \subset \mathbb{F}_{125}$. Поскольку для $x \in \mathbb{F}_5$ выполняется $x^{25} + x^5 + x = x + x + x = 3x$, то этот образ совпадает с подполем $\mathbb{F}_5$.

Итак, образ отображения A одномерный, поэтому ядро имеет размерность 2. Элемент 1 принадлежит образу, поэтому в него переходит $5^2 = 25$ элементов поля $\mathbb{F}_{125}$.

Ответ: 25. □

Пример 11.37. Сколько решений у уравнения $x^{25}+x^5+x = 1$ в поле из 625 элементов?

Характеристика поля по-прежнему равна 5, так что $A\colon x \mapsto x^{25} + x^5 + x$ является линейным отображением поля $\mathbb{F}_{625}$ как 4-мерного векторного пространства над простым полем $\mathbb{F}_5$.

Найдём ядро этого отображения. Пусть $x^{25} + x^5 + x = 0$. Тогда, применяя автоморфизм Фробениуса, получаем равенства

$$\begin{aligned}
x^{25}+ \quad & x^5+ \quad & x \quad & = 0,\\
x^{125}+ \quad & x^{25}+ \quad & x^5 \quad & = 0,\\
x+ \quad & x^{125}+ \quad & x^{25} \quad & = 0,\\
x^5+ \quad & x+ \quad & x^{125} \quad & = 0.
\end{aligned}$$

Вычитая из третьего равенства четвёртое, получаем $x^{25}-x^5 = 0$; вычитая из второго третье, получаем $x^5 - x = 0$. То есть $x^{25} = x^5 = x$. Подставляя в первое уравнение, получаем $3x = 0$, то есть $x = 0$.

Итак, ядро нулевое. Поэтому отображение A является биекцией.

Ответ: 1. □

Одной из важных особенностей конечных полей является связь между корнями неприводимого над $\mathbb{F}_p$ многочлена и автоморфизмом Фробениуса.

Теорема 11.38. *Пусть $\beta \in \mathbb{F}_{p^n}$ — корень неприводимого многочлена $f(x)$ степени n с коэффициентами из $\mathbb{F}_p$. Тогда β, β^p, …, $\beta^{p^{n-1}}$ все различны и исчерпывают список корней этого многочлена.*

Из этой теоремы следует, что все корни неприводимого многочлена получаются применением автоморфизма Фробениуса к одному из них.

Доказательство. Вначале докажем, что если β — корень $f(x)$, то β^p — тоже корень.

Поскольку $a^p = a$ для всех $a \in \mathbb{F}_p$, для любого многочлена $f(x)$ с коэффициентами из $\mathbb{F}_p$ выполняется равенство

$$f(x)^p = f(x^p). \tag{11.3}$$

Действительно, возведение в степень p сохраняет операции сложения и умножения (утверждение 11.33). Поэтому

$$(a_0 + a_1x + \cdots + a_kx^k)^p = a_0^p + a_1^px^p + a_2^px^{2p} + \cdots + a_k^px^{kp} = \\ = a_0 + a_1(x^p) + a_2(x^p)^2 + \cdots + a_k(x^p)^k.$$

Если $f(\beta) = 0$, то и $f(\beta)^p = 0$. Из (11.3) получаем, что и $f(\beta^p) = 0$.

Итак, мы доказали, что β, β^p, …, $\beta^{p^{n-1}}$ — корни многочлена $f(x)$. Осталось доказать, что они все различны, тогда из леммы о числе корней многочлена будет следовать, что мы нашли все корни многочлена $\varphi(x)$.

Пусть орбита действия автоморфизма Фробениуса, содержащая β, имеет размер k, то есть $F^{\circ k}(\beta) = \beta$. Тогда $\beta^{p^k} = \beta$ и потому многочлен $x^{p^k} - x$ делится на многочлен f (минимальный многочлен элемента β). Из теоремы 11.24 заключаем, что $k = n$. □

Приведём простой пример, в котором помогает теорема 11.38.

Пример 11.39. Про элемент a поля $\mathbb{F}_{512}$ известно, что $a^{64} \neq a$. Какова степень минимального многочлена элемента a?

Поскольку минимальный многочлен неприводим, его корни в поле $\mathbb{F}_{512}$ образуют орбиту Фробениуса. С другой стороны, степень минимального многочлена должна равняться размерности $\mathbb{F}_2(a)$ как векторного пространства над $\mathbb{F}_2$. Так как $512 = 2^9$, возможные значения такой размерности 1, 3, 9 (делители 9). Первые два варианта не удовлетворяют условию, так как в таком случае $a^{64} = a^{2^6} = (a^{2^3})^{2^3} = a$.

Поэтому степень минимального многочлена a равна 9. □

Верно и обратное утверждение к теореме 11.38: многочлен

$$m_\alpha(x) = \prod_k (x - F^{\circ k}\alpha)$$

(корни образуют орбиту автоморфизма Фробениуса элемента α поля $\mathbb{F}_{p^n}$) лежит в кольце $\mathbb{F}_p[x]$ и неприводим в этом кольце. Действительно, множество корней этого многочлена не изменяется под действием автоморфизма Фробениуса (множители в произведении переставляются, но их набор остаётся тем же самым). Это означает, что и сам многочлен не изменяется, то есть его коэффициенты сохраняются при действии автоморфизма Фробениуса. Но это и означает, что они лежат в простом подполе $\mathbb{F}_p$.

Неприводимость этого многочлена над полем $\mathbb{F}_p$ сразу следует из теоремы 11.38: корни приводимого многочлена будут разбиваться на несколько (больше одной) орбит автоморфизма Фробениуса.

Поскольку α — корень многочлена, порождённого орбитой α автоморфизма Фробениуса, то из неприводимости получаем также, что это минимальный многочлен для α.

Контрольный вопрос 11.40. Минимальный многочлен элемента α поля характеристики 3 имеет степень 6. Найдите степень минимального многочлена элемента α^9.

Приведём пример использования теоремы 11.38.

Пример 11.41. Пусть $a \in \mathbb{F}_{121}$ — корень многочлена x^2-2x+4 в поле из 121 элемента. Каковы возможные значения a^{12}?

Прежде всего выясним, приводим ли этот многочлен в кольце $\mathbb{F}_{11}[x]$. Дискриминант $2^2 - 16 = -12 \equiv -1 \pmod{11}$, а $(-1)^5 = -1$. Поэтому дискриминант является квадратичным невычетом, у многочлена $x^2 - 2x + 4$ нет корней в $\mathbb{F}_{11}$, то есть он неприводим в кольце $\mathbb{F}_{11}[x]$.

Поэтому его корни образуют орбиту автоморфизма Фробениуса в $\mathbb{F}_{121}$, то есть корнями являются a и a^{11}. Но тогда $a^{12} = a \cdot a^{11}$ является произведением корней и по теореме Виета совпадает со свободным членом многочлена.

Ответ: $a^{12} = 4$. □

Если бы многочлен оказался приводимым, рассуждать пришлось бы иначе.

Пример 11.42. Пусть $a \in \mathbb{F}_{121}$ — корень многочлена x^2+6x+4 в поле из 121 элемента. Каковы возможные значения a^{12}?

Теперь дискриминант $36 - 16 = 20 \equiv 9 = 3^2 \pmod{11}$ является квадратичным вычетом и у уравнения $x^2 + 6x + 4 = 0$ два решения в $\mathbb{F}_{11}$: 1 и 4.

Поэтому возможные значения a^{12} равны $1^{12} = 1$ и $4^{12} = 4^2 = 16 = 5$ (вычисления по модулю 11, использована малая теорема Ферма). □

То обстоятельство, что в любом конечном поле корни неприводимого над простым полем $\mathbb{F}_p$ многочлена образуют орбиту автоморфизма Фробениуса, позволяет находить степени неприводимых множителей для многочленов $x^n - 1$.

Контрольный вопрос 11.43. Проверьте, что для любого простого p и любого целого положительного n многочлен $x^n - 1$ не имеет кратных корней.

Рассмотрим поле $\mathbb{F}_q$, в котором $x^n - 1$ раскладывается на линейные множители. Всего его корни различны и образуют подгруппу мультипликативной группы поля (произведение двух корней n-й степени из единицы также является корнем n-й степени из единицы). Отсюда следует, что n должно делить $q - 1$. А поскольку мультипликативная группа конечного поля циклическая, то подгруппа корней n-й степени также циклическая. Пусть она порождена элементом α. Рассмотрим орбиты Фробениуса на степенях α, их размеры и будут давать степени неприводимых делителей $x^n - 1$.

Пример 11.44. Рассмотрим многочлен $x^{15}-1 \in \mathbb{F}_2[x]$ и найдём степени его неприводимых делителей. Орбиты корней этого многочлена под действием автоморфизма Фробениуса ($x \mapsto x^2$ в случае характеристики 2) имеют вид

$$(\alpha^0 = 1),\ (\alpha, \alpha^2, \alpha^4, \alpha^8),\ (\alpha^3, \alpha^6, \alpha^{12}, \alpha^9),\ (\alpha^5, \alpha^{10}),\ (\alpha^7, \alpha^{14}, \alpha^{13}, \alpha^{11})$$

(степени считаем по модулю 15, так как α — примитивный корень 15-й степени из единицы; другими словами, — порождающий циклической подгруппы порядка 15 в мультипликативной группе поля).

Видим, что есть три неприводимых делителя степени 4, один — степени 2, и один — степени 1.

Это согласуется с проделанным ранее поиском всех неприводимых многочленов над $\mathbb{F}_2$ степени не выше 4. Обратите внимание, что $15 = 2^4 - 1$, поэтому неприводимыми делителями этого многочлена будут все неприводимые многочлены степеней 2 и 4 и один — степени 1 (второй такой многочлен x является делителем $x^{16} - x$, но не многочлена $x^{15} - 1$). □

Пример 11.45. Аналогично происходит рассуждение и для n, которые не равны степени двойки без единицы. Рассмотрим, скажем, многочлен $x^{23} - 1 \in \mathbb{F}_2[x]$. Теперь через ω обозначим примитивный корень 23-й степени из единицы. Получаем разбиение степеней ω на орбиты автоморфизма Фробениуса

$$\begin{aligned}&(\omega^0),\\ &(\omega, \omega^2, \omega^4, \omega^8, \omega^{16}, \omega^9, \omega^{18}, \omega^{13}, \omega^3, \omega^6, \omega^{12}),\\ &(\omega^5, \omega^{10}, \omega^{20}, \omega^{17}, \omega^{11}, \omega^{22}, \omega^{21}, \omega^{19}, \omega^{15}, \omega^7, \omega^{14}).\end{aligned}$$

Таким образом, есть один неприводимый делитель степени 1 и два неприводимых делителя степени 11. □

Контрольный вопрос 11.46. Найдите неприводимый делитель степени 1 многочлена $x^{23} - 1$.

Пример 11.47. Пусть a — порождающий мультипликативной группы поля $\mathbb{F}_{32}$. Найдём наименьшую степень многочлена из $\mathbb{F}_2[x]$, корнями которого являются a^3, a^9, a^{15}.

Для этого нужно найти орбиты автоморфизма Фробениуса $x \mapsto x^2$ указанных элементов (указываем показатели a):

$$(3,\ 6,\ 12,\ 24,\ 17);\quad (9,\ 18,\ 5,\ 10,\ 20);\quad (15,\ 30,\ 29,\ 27,\ 23).$$

Любой многочлен из $\mathbb{F}_2[x]$, корнями которого являются a^3, a^9, a^{15}, имеет корнями все эти элементы. Поэтому минимальная степень равна $2 \cdot 5 = 15$. □

Неподвижные точки какой-то степени автоморфизма Фробениуса образуют подполе и нетрудно в точности определить, какое подполе получается.

Утверждение 11.48. *Пусть p — простое, а $\text{НОД}(n, k) = d$. Тогда неподвижные точки k-й степени автоморфизма Фробениуса поля из p^n элементов образуют подполе из p^d элементов.*

Доказательство. Элементы подполя из p^d элементов, и только они, являются решениями уравнения $x^{p^d} - x = 0$, то есть неподвижными точками d-й степени автоморфизма Фробениуса.

Если $F^{\circ d}x = x$, то x неподвижная точка любой итерации $F^{\circ d}$. Для итераций d-й степени автоморфизма Фробениуса получаем выражения

$$F^{\circ d}x = x^{p^d}, \quad (F^{\circ d} \circ F^{\circ d})x = (x^{p^d})^{p^d} = x^{p^{2d}}, \quad \ldots, \quad (F^{\circ d})^{\circ t}x = x^{p^{dt}}.$$

Так как $d \mid k$, все элементы подполя $\mathbb{F}_{p^d}$ являются неподвижными точками $F^{\circ k}$.

Докажем, что других неподвижных точек нет. Пусть $F^{\circ k}x = x$. Так же, как и выше, проверяется, что для любого целого положительного t из следует

$$x = (F^{\circ k})^{\circ t}x = x^{p^{kt}}.$$

Уравнение $kt = d + ns$ имеет решения в целых положительных числах t, s (какое-то кратное k даёт остаток d по модулю n). Для такого t получаем

$$x = x^{p^{kt}} = x^{p^{d+ns}} = (x^{p^{ns}})^{p^d} = x^{p^d}$$

(так как для всех элементов поля выполняется равенство $x^{p^n} = x$). Это и означает, что $x \in \mathbb{F}_{p^d}$. □

Приведём пример использования этого наблюдения.

Пример 11.49. Проверим, что $f(x) = x^4 + x - 1 \in \mathbb{F}_3[x]$ неприводим. Корней в $\mathbb{F}_3$ у этого многочлена нет, что легко проверяется:

$$0^4 + 0 - 1 = -1 \neq 0, \quad 1^4 + 1 - 1 = 1 \neq 0, \quad (-1)^4 + (-1) - 1 = 1 - 2 = -1 \neq 0.$$

Нужно ещё проверить, что этот многочлен не разлагается в произведение неприводимых в $\mathbb{F}_3[x]$ многочленов степени 2.

Предположим противное. Тогда у многочлена $f(x)$ есть корень α в $\mathbb{F}_9$ (так как в $\mathbb{F}_9$ все многочлены второй степени разлагаются на линейные множители). Получаем такие равенства

$$1 = \alpha^8 = (1 - \alpha)^2 = 1 - 2\alpha + \alpha^2$$

(первое равенство следует из того, что порядок мультипликативной группы поля $\mathbb{F}_9$ равен 8; второе — из равенства $f(\alpha) = \alpha^4 + \alpha - 1 = 0$). Итак, α является корнем уравнения $x^2 - 2x = 0$, то есть принадлежит подполю $\mathbb{F}_3$. Приходим к противоречию, так как корней у $f(x)$ в этом поле нет.

Из неприводимости $f(x)$ следует, что кольцо вычетов $\mathbb{F}_3/(x^4 + x - 1)$ изоморфно полю $\mathbb{F}_{81}$. В этом поле есть подполе $\mathbb{F}_9$. Значит, такое подполе есть и в кольце вычетов $\mathbb{F}_3/(x^4 + x - 1)$. Из каких вычетов оно состоит? Разумеется, вычеты $[a]$, $a \in \mathbb{F}_3$ входят в это подполе, так как оно содержит простое подполе $\mathbb{F}_3$. Поле из 9 элементов — 2-мерное пространство над $\mathbb{F}_3$. Поэтому нужно найти ещё один элемент $[g]$ этого подполя и тогда все его элементы будут выражаться в виде $[a \cdot 1 + b \cdot g]$, $a, b \in \mathbb{F}_3$.

Для любого $\alpha \in \mathbb{F}_{81}$ выполняется равенство

$$(\alpha + \alpha^9)^9 = \alpha^9 + \alpha^{81} = \alpha + \alpha^9,$$

то есть по утверждению 11.48 элемент $\alpha + \alpha^9$ принадлежит подполю $\mathbb{F}_9$. Поэтому достаточно подобрать такое α, чтобы $\alpha + \alpha^9$ не лежало бы в простом подполе $\mathbb{F}_9$, это и даст искомый второй элемент базиса $\mathbb{F}_9$ над $\mathbb{F}_3$.

Попробуем вычет $[x]$. Для него

$$[x] + [x]^9 = [x] \cdot ([1] + [x^8]) = [x] \cdot ([1] + [(1-x)^2]) = [x] \cdot [2 - 2x + x^2] = [2x - 2x^2 + x^3]$$

не принадлежит подполю $\mathbb{F}_3$ (в любом классе вычетов есть ровно один многочлен степени меньше 4 и, как видно из проделанного вычисления, этот многочлен не является константой).

Таким образом, подполе из 9 элементов в кольце вычетов $\mathbb{F}_3/(x^4+x-1)$ образуют вычеты

$$a + 2bx - 2bx^2 + bx^3$$

и только они. □

Оказывается, что других автоморфизмов, кроме степеней автоморфизма Фробениуса, у конечных полей нет.

Теорема 11.50. *Любой автоморфизм поля $\mathbb{F}_q$, $q = p^n$, p — простое, является композицией автоморфизмов Фробениуса.*

Доказательство. Пусть $\varphi\colon \mathbb{F}_q \to \mathbb{F}_q$ — автоморфизм. Он сохраняет 0 и 1: $\varphi(0) = 0$, $\varphi(1) = 1$, значит, сохраняет и любую сумму единиц, то есть все элементы простого подполя $\mathbb{F}_p$:

$$\varphi(\underbrace{1 + 1 + \cdots + 1}_{a \text{ раз}}) = \underbrace{\varphi(1) + \varphi(1) + \cdots + \varphi(1)}_{a \text{ раз}} = \underbrace{1 + 1 + \cdots + 1}_{a \text{ раз}}.$$

Поэтому автоморфизм переводит корень β многочлена f степени n, неприводимого над простым полем, в другой корень:

$$\varphi(0) = \varphi(f(\beta)) = f(\varphi(\beta)) = 0.$$

Таким образом, $\varphi(\beta) = F^{\circ k}(\beta)$ для некоторого k.

Композиция автоморфизмов — автоморфизм. Рассмотрим автоморфизм

$$\varphi' = (F^{-1})^{\circ k}\varphi.$$

Этот автоморфизм переводит β в себя. Но тогда он переводит в себя все степени β и все линейные комбинации этих степеней с коэффициентами из простого поля. Осталось вспомнить, что $\mathbb{F}_p(\beta) = \mathbb{F}_q$. Поэтому автоморфизм φ' — тождественный, а $\varphi = F^{\circ k}$, что и требовалось доказать. □

Пример 11.51. Рассмотрим пример поля с 16 элементами. Пусть α — порождающий элемент мультипликативной группы $\mathbb{F}_{16}^*$. В примере 11.44 мы нашли разбиение степеней α на корни неприводимых многочленов из $\mathbb{F}_2[x]$ степеней 1, 2 и 4. Есть три неприводимых многочлена степени 4 в кольце $\mathbb{F}_2[x]$:

$$1 + x + x^4, \quad 1 + x + x^2 + x^3 + x^4, \quad 1 + x^3 + x^4.$$

На степенях α есть три орбиты автоморфизма Фробениуса размера 4:

$$(\alpha, \alpha^2, \alpha^4, \alpha^8),\ (\alpha^3, \alpha^6, \alpha^{12}, \alpha^9),\ (\alpha^7, \alpha^{14}, \alpha^{13}, \alpha^{11}).$$

Каждой орбите соответствует какой-то неприводимый многочлен. Можно ли восстановить это соответствие?

Порядок α^3 в мультипликативной группе поля $\mathbb{F}_{16}^*$ равен 5. Значит, второй многочлен отвечает второй орбите: минимальный многочлен α^3 должен быть делителем $x^5 - 1$.

Уточнить соответствие между оставшейся парой неприводимых многочленов и парой орбит автоморфизма Фробениуса уже невозможно. Действительно, вся эта конструкция начинается с выбора порождающего элемента в группе $\mathbb{F}_{16}^*$. Но если α — порождающий, то и α^7 — порождающий.

Контрольный вопрос 11.52. Докажите последнее утверждение.

Это означает, что две орбиты Фробениуса можно переставить между собой изменением выбора порождающего группы $\mathbb{F}_{16}^*$. Тем не менее, эта перестановка, как следует из теоремы 11.50 не даёт автоморфизма поля. Дело в том, что возведение в степень 7 является, конечно, автоморфизмом циклической группы 15 порядка, так как 7 и 15 взаимно просты. Однако аддитивная структура поля при этом не сохраняется: если α является корнем многочлена x^4+x+1, то α^7 не является корнем этого многочлена (не лежит в орбите автоморфизма Фробениуса, порождённой α). Наоборот, он является корнем многочлена $x^4 + x^3 + 1$.

Итак, выбор минимального многочлена для порождающего элемента мультипликативной группы конечного поля неоднозначен. В случае поля из 16 элементов есть ровно две возможности. □

В завершение рассмотрим довольно трудный пример, в котором возникает не только автоморфизм Фробениуса, но и другие соображения.

Пример 11.53. Степени минимальных многочленов элементов a, b в конечном поле $\mathbb{F}_{p^n}$ равны 5. Каковы возможные значения степени минимального многочлена элемента $a^2 + b^2$?

Из изоморфизма $\mathbb{F}_p(\alpha) \cong \mathbb{F}_p[x]/(m_\alpha)$ следует, что $\mathbb{F}_p(a)$ — поле из p^5 элементов, и $\mathbb{F}_p(b)$ — поле из p^5 элементов. Но такое подполе единственное, поэтому $\mathbb{F}_p(a) = \mathbb{F}_p(b)$.

Понятно тогда, что a^2+b^2 также лежит в поле из p^5 элементов. Возможные степени минимального многочлена элемента в таком поле — делители 5, то есть 1 или 5. Возможен также случай, когда $a^2 + b^2 = 0$, в этом случае степень минимального многочлена не определена.

Дальнейший анализ зависит от характеристики поля.

Пусть $p = 2$. Тогда возведение в квадрат — линейное преобразование. возможны ли равенства $(a + b)^2 = 1$ и $(a + b)^2 = 0$, где a и b порождают поле из 2^5 элементов.

Равенство $(a + b)^2 = 0$ выполняется при $b = -a = a$ (характеристика поля равна 2).

Равенство $a + b = 1$ выполняется при $b = a + 1$ (из примера 11.19 мы знаем, что степени минимальных многочленов a и $a + 1$ одинаковы).

Осталось убедиться, что степень минимального многочлена $(a+b)^2$ может равняться 5, когда степени минимальных многочленов a и b равны 5. Поскольку $a+b$ лежит в одной орбите Фробениуса с $a+b$, степени минимальных многочленов у них одинаковы. Выберем $b=a^2$. Равенства $a^2+a=0$ и $a^2+a=1$ невозможны: в первом случае два корня уже есть в простом поле, которому a не принадлежит, во втором случае степень минимального многочлена a равнялась бы 2.

Поэтому степень минимального многочлена a^2+a равна 5.

Итак, в характеристике 2 возможны все три случая: степень равна 1, равна 5 или не определена.

Далее считаем, что характеристика поля $p>2$.

Случай степени 5 для a^2+b^2 имеет место для $a=b$. Поскольку $\mathbb{F}_p(2a^2)=\mathbb{F}_p(a^2)$, достаточно проверить, что степень минимального многочлена для a^2 равна 5. Если бы эта степень равнялась 1, то степень минимального многочлена a была бы 2: в таком случае a является корнем уравнения $x^2-\alpha=0$, $a^2=\alpha\in\mathbb{F}_p$.

Равенство $a^2+b^2=0$ возможно если -1 является квадратичным вычетом по модулю p, то есть $(-1)^{(p-1)/2}=-1 \pmod p$, что равносильно $p\equiv 3 \pmod 4$. В этом случае в качестве b можно выбрать αa, $\alpha\in\mathbb{F}_p$, $\alpha^2=-1$.

С другой стороны, если $a^2+b^2=0$, то $-1=a^2/b^2=(a/b)^2$. Возможные корни уравнения $x^2+1=0$ в поле $\mathbb{F}_{p^5}$ лежат в $\mathbb{F}_p$, так как иначе появлялось бы подполе из p^2 элементов, что невозможно. Поэтому при $p\equiv 1 \pmod 4$ равенство $a^2+b^2=0$ невозможно.

Осталось понять, может ли степень минимального многочлена a^2+b^2 равняться 1, то $\alpha=a^2+b^2$ принадлежит простому подполю $\mathbb{F}_p$. Ответ положительный: всегда можно подобрать такие a, b, что $\mathbb{F}_p(a)=\mathbb{F}_p(b)=\mathbb{F}_{p^5}$ и $a^2+b^2=1$.

Для любого $t\in\mathbb{F}_{p^5}\setminus\mathbb{F}_p$ выполняется $t^2+1\neq 0$ (так как подполя из p^2 элементов в поле из p^5 элементов нет). Непосредственным вычислением проверяется, что сумма квадратов $a=2t/(1+t^2)$, $b=(1-t^2)/(1+t^2)$ равна 1:

$$\left(\frac{2t}{1+t^2}\right)^2+\left(\frac{1-t^2}{1+t^2}\right)^2=\frac{4t^2+1-2t^2+t^4}{(1+t^2)^2}=1.$$

При этом a, b не принадлежат простому подполю, если t не принадлежит простому подполю, так как в противном случае t должен равняться 0 или быть корнем квадратного уравнения с коэффициентами в $\mathbb{F}_p$.

Поскольку подполей кроме простого в $\mathbb{F}_{p^5}$ нет, $\mathbb{F}_p(a)=\mathbb{F}_p(b)=\mathbb{F}_{p^5}$. □

11.9 Критерий неприводимости многочлена над конечным полем

Теорема 11.54. *Многочлен $f\in\mathbb{F}_q[x]$ неприводим тогда и только тогда, когда он взаимно прост со своей производной и число решений уравнения $h^q-h=0$ в кольце вычетов $\mathbb{F}_q[x]/(f)$ равно в точности q.*

Доказательство. Взаимная простота с производной необходима для неприводимости (иначе у многочлена есть необратимый делитель и он приводим). Поэтому далее мы рассматриваем только случай многочленов, взаимно простых с производной. Как

было показано выше, это означает, что в разложении многочлена на неприводимые в $\mathbb{F}_q[x]$ множители

$$f = f_1 \cdot f_2 \cdot \ldots \cdot f_m$$

все неприводимые множители различны (входят в разложение в степени 1).

По китайской теореме

$$\mathbb{F}_q[x]/(f) \cong \mathbb{F}_q[x]/(f_1) \oplus \cdots \oplus \mathbb{F}_q[x]/(f_m) \cong \mathbb{F}_{q^{d_1}} \oplus \cdots \oplus \mathbb{F}_{q^{d_m}},$$

где $d_i = \deg f_i$. Второй изоморфизм следует из того, что кольцо вычетов по модулю неприводимого многочлена является полем, причём количество классов вычетов в кольце $\mathbb{F}_q[x]/(f_i)$ равно как раз q^{d_i}.

Количество решений уравнения $h^q - h = 0$ в поле $\mathbb{F}_{q^d}$ равно в точности q: все элементы подполя $\mathbb{F}_q \subseteq \mathbb{F}_{q^d}$ являются решениями этого уравнения, а больше q корней у многочлена степени q быть не может.

Таким образом, количество решений уравнения $h^q - h = 0$ в прямой сумме полей $\mathbb{F}_{q^{d_1}} \oplus \cdots \oplus \mathbb{F}_{q^{d_m}}$ равно q^m. Эти решения имеют вид $(g_1, g_2, \ldots, g_m)$, где g_i — какое-то решение уравнения $h^q - h = 0$ в поле $\mathbb{F}_{q^{d_i}}$.

Осталось заметить, что неприводимость многочлена равносильна $m = 1$. □

Линейность отображения $h \mapsto h^q - h$ в поле $\mathbb{F}_q$ позволяет свести проверку условия в теореме 11.54 к линейной алгебре. Нужно записать матрицу этого отображения в каком-нибудь базисе кольца вычетов $\mathbb{F}_q[x]/(f)$ как векторного пространства над полем $\mathbb{F}_q$ и найти размерность ядра (скажем, приводя матрицу к диагональному виду методом Гаусса).

Такого рода вычисления очень эффективно выполняются компьютерами. Для ручного счёта они оказываются довольно громоздкими. Приведём тем не менее несколько примеров.

Пример 11.55. Проверим неприводимость многочлена $x^2 - a$ в кольце $\mathbb{F}_p[x]$, $p > 2$.

В данном случае пространство двумерно, его базис образуют вычеты $[1]$ и $[x]$. Запишем действие отображения $h \mapsto h^p - h$ на базисных векторах:

$$[1] \mapsto [1^p - 1] = 0,$$
$$[x] \mapsto [x^p - x] = [x] \cdot [(x^2)^{(p-1)/2} - 1] = [x] \cdot (a^{(p-1)/2} - 1).$$

Матрица этого отображения имеет вид

$$\begin{pmatrix} 0 & 0 \\ 0 & a^{(p-1)/2} - 1 \end{pmatrix}.$$

Размерность ядра не меньше 1 (есть нулевой столбец) и равна 2 в том и только том случае, когда $a^{(p-1)/2} - 1 \equiv 0 \pmod p$. Получаем уже известный нам критерий квадратичного невычета: $a^{(p-1)/2} - 1 \not\equiv 0 \pmod p$, что равносильно $a^{(p-1)/2} - 1 \equiv \equiv -1 \pmod p$ (вспомните, почему так). □

Контрольный вопрос 11.56. Докажете, что многочлен $x^2 - a$ неприводим в $\mathbb{F}_p[x]$ тогда и только тогда, когда a квадратичный невычет.

Пример 11.57. Проверим неприводимость многочлена $f(x) = x^p - x - 1$ в кольце $\mathbb{F}_p[x]$.

Так как в кольце вычетов $\mathbb{F}_p[x]/(f(x))$ выполняется равенство $[x^p] = [x] + [1]$, получаем действие отображения $h \mapsto h^p - h$ на базисных векторах $[x^i]$, $0 \leqslant i < p$:

$$[1] \mapsto [1^p - 1] = 0,$$
$$[x^i] \mapsto [x^{pi} - x^i] = [(x+1)^i - x^i], \quad i > 0.$$

Поэтому любой элемент ядра этого отображения $[g(x)]$ обязан удовлетворять уравнению $g(x+1) = g(x)$. Все константы, конечно, годятся. Докажем, что решений первой степени нет, это легко: $a(x+1) + b = ax + b$ равносильно системе $a = a$, $a + b = b$, откуда следует $a = 0$.

Далее рассуждаем индукцией по степени многочлена. База — многочлены степени 1 — уже доказана.

Предположим, что нет многочленов степени $d-1$, удовлетворяющих уравнению $g(x+1) = g(x)$. Рассмотрим многочлен $g = g_0 + g_1 x + \cdots + g_d x^d$ степени d. Если $g(x+1) = g(x)$, то

$$g_0 + g_1(x+1) + \cdots + g_d(x+1)^d = g_0 + g_1 x + \cdots + g_d x^d$$

и многочлен $h = (g - g_0)/x$ также является решением уравнения, но его степень $d - 1$. Полученное противоречие доказывает индуктивный переход. По принципу математической индукции заключаем, что решениями уравнения $g(x+1) = g(x)$ являются только константы.

Значит, многочлен $x^p - x - 1$ неприводим в кольце $\mathbb{F}_p[x]$. □

12 Корректирующие коды

В этом разделе мы рассмотрим приложения конечных полей к построению корректирующих кодов (другое название — коды, исправляющие ошибки).

12.1 Определения и основные свойства

Рассмотрим такую ситуацию. Есть набор сообщений S_i, $1 \leqslant i \leqslant N$, которые нужно передавать по каналу связи. Сообщения будем кодировать нулями и единицами, то есть каждому сообщению будем сопоставлять слово из нулей и единиц (бинарный набор), который обычно называется *кодовым словом*. Шум в канале связи приводит к ошибкам передачи: полученное в приёмнике слово может отличаться от переданного слова.

Неформально задача теории кодирования состоит в разработке таких способов кодирования сообщений, при которых хотя бы часть возможных ошибок обнаруживается (и, что ещё лучше, *исправляется*, то есть на приёмном конце удаётся восстановить переданное слово).

Далее мы ограничимся только частным случаем этой задачи.

Во-первых, будем считать, что все сообщения кодируются словами одинаковой длины. Зафиксируем это в определении.

Определение 12.1. Код длины n — это подмножество C множества $\{0,1\}^n$ двоичных слов длины n.

Во-вторых, будем считать, что ошибки при передаче могут только изменять значения некоторых битов. Вообще говоря, это не единственный вид ошибок. Возможны, например, выпадения и вставки — какой-то из битов может не дойти до приёмника или, наоборот, из-за помех может произойти ложное срабатывание приёмника и получится бит, которого никто не посылал. Мы такие ситуации рассматривать не будем.

В-третьих, мы собираемся исправлять некоторое заранее заданное количество ошибок r. Вообще говоря, можно формулировать задачу более сложным образом: скажем, задать вероятностную модель появления ошибок и искать код, при котором вероятность правильного восстановления сообщения максимальная. Исправление заданного количества ошибок связано с одной из самых простых и естественных моделей — ошибка происходит в каждом бите независимо с некоторой вероятностью, одинаковой для всех битов.

Наконец, мы хотим передавать сообщения как можно быстрее, то есть брать длину кодового слова как можно меньше.

Обратите внимание, что мы не учитываем трудность восстановления переданного слова. Эта задача в общем случае трудна и пригодные для практики коды должны допускать её эффективное решение. Мы, однако, не будем обсуждать эту задачу *декодирования*. Такой подход разумен, когда длина кодового слова невелика. Исторически так и было: кодовые слова генерировались электронными схемами,

которые во время возникновения теории кодирования (середина прошлого века) были дорогими и не слишком большими.

Пример 12.2 (Код повторений)**.** Пусть сообщений всего 2. Одно будем кодировать словом 0^n длины n из одних нулей, второе — словом 1^n из одних единиц.

Если при передаче первого сообщения произошло меньше $n/2$ ошибок, то в полученном приёмником слове будет больше нулей, чем единиц. Если при передаче второго сообщения произошло меньше $n/2$ ошибок, то в полученном приёмником слове будет больше единиц, чем нулей.

Таким образом, *декодировать* полученное сообщение можно по очень простому правилу: взять то из булевых значений 0, 1, которое чаще встречается в полученном слове, и заменить полученное слово на код этого булева значения.

Код повторений очень хорош в смысле исправления ошибок: можно исправить любое количество ошибок, лишь бы их было меньше половины.

Код повторений очень плох в смысле скорости передачи: для передачи одного полезного бита (обычно говорят, *информационного*) нужно передать n битов. Другими словами, скорость передачи $1/n$. □

Пример 12.3 (Проверка на чётность)**.** Пусть количество сообщений равно $N = 2^k$. Закодируем их сначала двоичными словами длины k каким-нибудь способом. Затем добавим ещё один бит к этим словам. Значение в этом бите равно 0, если количество единиц чётно, и 1, если нечётно.

При $k = 2$ получаем такой код из 4 слов длины 3:

$$(000),\ (011),\ (101),\ (110).$$

Скорость передачи у кода с проверкой на чётность близка к максимальной, для передачи k информационных битов нужно передать всего $k + 1$ кодовый бит. Скорость $1 - 1/(k + 1)$.

В коде с проверкой на чётность можно обнаружить одну ошибку. Действительно, по построению кода количество единиц в кодовом слове чётно. Если происходит ровно одна ошибка, количество единиц в полученном слове нечётно.

Однако исправить обнаруженную ошибку невозможно. Скажем, в рассмотренном выше примере приёмник получил слово 001. Оно могло получиться инвертированием одного бита в трёх разных кодовых словах: 101, или 011, или 000. В общем случае инвертирование любого бита в слове с нечётным количеством единиц даёт кодовое слово — слово с чётным количеством единиц. □

Хотя проверка не чётность выглядит не очень привлекательно (ни одной ошибки исправить не удаётся), исторически это был важный пример, из которого выросла дальнейшая теория и который применялся на практике. До сих пор по-английски корректирующие коды зачастую называются parity check codes, хотя сами коды намного сложнее этого простого примера проверки на чётность.

Итак, при сделанных предположениях получаем такую комбинаторную задачу: при заданном количестве сообщений N и заданном количестве ошибок r, которые нужно исправлять, нужно выбрать как можно меньшую длину слова n, для которой

существует код длины n (то есть множество двоичных слов длины n), допускающий исправление r ошибок.

Эта задача — одна из самых популярных задач комбинаторики. На эту тему написаны десятки тысяч (это не преувеличение) статей. Полного решения этой задачи до сих пор неизвестно.

Из рассмотренных примеров ясно, что нужно найти компромисс между двумя крайними вариантами, которые выше представлены кодом повторений и кодом проверки на чётность. Вариантов такого компромисса очень много, что и определяет отчасти трудность задачи.

Более удобно рассматривать другую задачу: при заданных n и r требуется найти максимальное число сообщений N в коде длины n, исправляющем r ошибок. Решив задачу про максимальное число сообщений, нетрудно получить и решение предыдущей задачи про код минимальной длины.

Мы приведём конструкции, которые дают хорошие приближения к оптимуму при некоторых значениях параметров. Точное решение будет дано лишь для случая $n = 2^m - 1$ и $r = 1$, а также для $n = 23$, $r = 3$ (см. раздел 12.6).

Будем использовать следующие обозначения.

Расстояние $\rho(\tilde{\alpha}, \tilde{\beta})$ между двоичными словами (бинарными наборами) $\tilde{\alpha}$, $\tilde{\beta}$ одинаковой длины равно по определению количеству позиций (координат), в которых эти слова различаются. Это расстояние обычно называется *метрикой Хэмминга* или расстоянием Хэмминга.

Утверждение 12.4. *Расстояние Хэмминга удовлетворяет свойствам метрики, то есть*

- $\rho(\tilde{\alpha}, \tilde{\beta}) = 0$ *тогда и только тогда, когда* $\tilde{\alpha} = \tilde{\beta}$;
- $\rho(\tilde{\alpha}, \tilde{\beta}) = \rho(\tilde{\beta}, \tilde{\alpha})$;
- *выполняется неравенство треугольника* $\rho(\tilde{\alpha}, \tilde{\beta}) + \rho(\tilde{\beta}, \tilde{\gamma}) \geqslant \rho(\tilde{\alpha}, \tilde{\gamma})$.

Доказательство. Равенство расстояния Хэмминга нулю означает, что во всех позициях в словах стоят одни и те же символы. Но это и значит, что слова равны.

Симметричность (второе равенство) очевидна из определения, которое само по себе симметрично.

Осталось доказать неравенство треугольника. Если в какой-то позиции у слов $\tilde{\alpha}$ и $\tilde{\beta}$ одинаковые символы и у слов $\tilde{\beta}$ и $\tilde{\gamma}$ одинаковые символы, то тогда символ слова $\tilde{\alpha}$ в этой позиции тот же, что и слова $\tilde{\gamma}$ (транзитивность равенства). Поэтому позиции, в которых различаются слова $\tilde{\alpha}$ и $\tilde{\gamma}$, лежат в объединении множеств позиций, в которых различаются $\tilde{\alpha}$ и $\tilde{\beta}$, и позиций, в которых различаются $\tilde{\beta}$ и $\tilde{\gamma}$. Но размер объединения множеств не превосходит суммы размеров множеств. Это и означает неравенство треугольника для расстояния Хэмминга. □

Контрольный вопрос 12.5. Пусть слова $\tilde{\alpha}$ и $\tilde{\beta}$ различаются в множестве позиций X, а слова $\tilde{\beta}$ и $\tilde{\gamma}$ — в множестве позиций Y. Проверьте, что тогда $\tilde{\alpha}$ и $\tilde{\gamma}$ различаются в

множестве позиций

$$X \oplus Y = (X \setminus Y) \cup (Y \setminus X)$$

(*симметрическая разность* множеств X и Y).

Контрольный вопрос 12.6. Чему равно расстояние Хэмминга между наборами

$$(001101000) \quad \text{и} \quad (001000110)?$$

Множество тех $\tilde{\beta}$, для которых $\rho(\tilde{\alpha}, \tilde{\beta}) \leqslant r$, назовём *шаром Хэмминга* $S_r(\tilde{\alpha})$ с центром $\tilde{\alpha}$ и радиусом r.

Утверждение 12.7. *Количество слов в шаре Хэмминга радиуса r равно*

$$V_n(r) = \sum_{i=0}^{r} \binom{n}{i},$$

где n — длина рассматриваемых двоичных слов.

Доказательство. Найдём количество слов на расстоянии ровно i от данного. Эти слова отличаются от данного слова ровно в i позициях, в остальных совпадают. Поскольку мы рассматриваем двоичный алфавит (на каждой позиции стоит 0 или 1), то количество таких слов равно количеству способов выбрать i позиций из n, что равно биномиальному коэффициенту. □

Определение 12.8. *Кодовым расстоянием* кода $C \subseteq \{0,1\}^n$ называется минимальное расстояние между различными кодовыми словами.

Пример 12.9. Для кода повторений минимальное расстояние равно n (кодовые слова различаются во всех позициях). □

Утверждение 12.10. *Код исправляет r ошибок тогда и только тогда, когда кодовое расстояние больше $2r$.*

Доказательство. При передаче кодового слова $\tilde{\alpha}$ любое полученное приёмником слово лежит в шаре радиуса r с центром $\tilde{\alpha}$. Возможность правильного восстановления переданного слова (в предположении, что произошло не больше r ошибок) равносильна тому, что ни одно слово из шара радиуса r с центром в кодовом слове не попадает в шар радиуса r с центром в другом кодовом слове.

Но это последнее условие — шары радиуса r с центрами в кодовых словах не пересекаются — равносильно тому, что кодовое расстояние больше $2r$. Это просто неравенство треугольника: если шары радиуса r пересекаются, то расстояние между их центрами не больше суммы расстояний от центров до точки пересечения, то есть не больше $2r$. А если расстояние между центрами шаров $\tilde{\alpha}$ и $\tilde{\beta}$ больше $2r$, то для любой точки $\tilde{\gamma}$ из неравенства $\rho(\tilde{\alpha}, \tilde{\gamma}) \leqslant r$ следует $\rho(\tilde{\beta}, \tilde{\gamma}) > r$. □

Из этого простого геометрического наблюдения следуют верхние и нижние оценки на размеры оптимальных кодов.

Теорема 12.11 (Верхняя граница Хэмминга)**.** *В любом коде длины n, исправляющем r ошибок, не больше*

$$\frac{2^n}{\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{r}}$$

слов.

Доказательство. Если в коде N слов и он исправляет r ошибок, то шары радиуса r с центрами в кодовых словах не пересекаются. Общее количество слов в таких шарах $NV_n(r)$ и оно не превосходит общего количества 2^n двоичных слов длины n. Отсюда и получаем искомую оценку. □

Граница Хэмминга достижима далеко не всегда. Но всегда можно построить код с несколько худшими значениями параметров, как показывает следующее рассуждение.

Теорема 12.12 (Нижняя граница Варшамова–Гилберта)**.** *При $2r < n$ существует код длины n, исправляющий r ошибок, в котором не меньше*

$$\frac{2^n}{\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{2r}}$$

слов.

Доказательство. Возьмём произвольное слово $\tilde{\alpha}^1$ и построим вокруг него шар радиуса $2r$. Следующим словом возьмём произвольное слово $\tilde{\alpha}^2$ вне этого шара. Третье слово выберем вне этих двух шаров. Далее аналогично, пока объединение шаров радиуса $2r$ с центрами в выбранных словах не покроет весь булев куб $\{0,1\}^n$.

Заметьте, что каждый новый шар содержит не более

$$\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{2r}$$

слов.

Значит, число таких шаров в конце процесса будет не меньше

$$\frac{2^n}{\binom{n}{0} + \binom{n}{1} + \ldots + \binom{n}{2r}}\,.$$

Расстояния между различными выбранными словами больше $2r$ по построению. Поэтому полученный код исправляет r ошибок. □

Конструкция в этом рассуждении требует перебора множества всех двоичных слов длины n. Это много. Хочется дать «явную» конструкцию хорошего кода, то есть сформулировать простое правило, которое выделяет кодовые слова. Смысл слова «явный» не вполне определён, его можно уточнять разными способами.

Однако есть общее неформальное наблюдение: если нужно предъявить «явный» пример комбинаторного объекта, про который известно доказательство существования, то полезными оказываются алгебраические соображения. В следующих разделах мы приведём первые примеры на этом пути.

12.2 Линейные коды

Булев куб, то есть множество двоичных слов длины n, можно снабдить дополнительной структурой. Например, можно рассмотреть координатное векторное пространство $\mathbb{F}_2^n$ над полем из 2 элементов. Координаты как раз принимают два значения: 0 и 1. Но теперь у нас появляется структура группы. В частности, теперь определено сложение двоичных слов (покомпонентное сложение координат по модулю 2).

Чтобы подчеркнуть наличие этой дополнительной структуры, будем теперь слова называть векторами.

Определение 12.13. *Линейный код* C — это подпространство $\mathbb{F}_2^n$.

Количество слов в линейном коде 2^k, где k — размерность подпространства C. Обычно параметры линейного кода задают как тройку $[n, k, d]$, где n — длина кода, k — размерность, а d — кодовое расстояние.

Для линейных кодов кодовое расстояние можно описать несколько проще, чем в общем случае.

Введём ещё одно определение: *норма* (Хэмминга) $\|\tilde{\gamma}\|$ — это число единичных координат в $\tilde{\gamma}$.

Расстояние Хэмминга легко выражается через норму суммы.

Утверждение 12.14. $\rho(\tilde{\alpha}, \tilde{\beta}) = \|\tilde{\alpha} + \tilde{\beta}\|$.

Доказательство. Сумма различных элементов поля $\mathbb{F}_2$ равна 1, а сумма одинаковых равна 0. □

Пример 12.15. $\tilde{\alpha} = (101011)$, $\tilde{\beta} = (000110)$, $\rho(\tilde{\alpha}, \tilde{\beta}) = 4$, $\tilde{\alpha} \oplus \tilde{\beta} = (101101)$, $\|\tilde{\alpha} \oplus \tilde{\beta}\| = 4$. □

Следствие 12.16. *Кодовое расстояние линейного кода равно наименьшей норме ненулевого вектора в этом коде.*

Доказательство. Норма — это расстояние от нулевого вектора, который входит в каждый линейный код. А расстояние между любыми двумя векторами в линейном коде равно норме их суммы, которая также принадлежит линейному коду (это же подгруппа по сложению). □

Пример 12.17 (Ещё раз о проверке на чётность)**.** Код проверки на чётность линейный. Он состоит из тех векторов, сумма координат которых равна 0 (в поле $\mathbb{F}_2$, разумеется). Другими словами, количество единиц в таких словах чётно.

Наименьшее количество единиц в ненулевом слове из чётного числа единиц равно 2, как легко понять. □

Если кодовое расстояние равно 3, то такой код исправляет одну ошибку. Мы сейчас построим пример линейных кодов с таким кодовым расстоянием. Это знаменитый код Хэмминга (на самом деле, семейство кодов).

Длину кода выберем $n = 2^m - 1$.

Рассмотрим такую таблицу:

$100\ldots$	$\ldots$	$\ldots 000$	$1100\ldots 000$
$010\ldots$	$\ldots$	$\ldots 000$	$1010\ldots 000$
$001\ldots$	$\ldots$	$\ldots 000$	$1001\ldots 000$
	$\ldots$		$\ldots$
$000\ldots$	$\ldots$	$\ldots 100$	$1111\ldots 101$
$000\ldots$	$\ldots$	$\ldots 010$	$1111\ldots 110$
$000\ldots$	$\ldots$	$\ldots 001$	$1111\ldots 111$

В правой части таблицы выписаны все $2^m - m - 1$ двоичных слов длины m и нормы больше 1, то есть содержащих более одной единичной координаты.

Контрольный вопрос 12.18. Проверьте, что количество таких слов как раз $2^m - m - 1$.

В левой части таблицы записана единичная матрица размера $(2^m - (m+1)) \times (2^m - (m+1))$. Таким образом, длины строки этой таблицы равна $(2^m - m - 1) + m = 2^m - 1 = n$.

Слова из кода Хэмминга — это все возможные суммы строк из этой таблицы, рассматриваемых как векторы в координатном пространстве $\mathbb{F}_2^n$ (то есть суммирование происходит по модулю 2).

Всего слов в коде Хэмминга $2^{2^m-(m+1)}$, так как суммы разных наборов строчек различаются (уже в левой части таблицы). Заметим, что

$$2^{2^m-(m+1)} = \frac{2^{2^m-1}}{2^m} = \frac{2^n}{n+1} = \frac{2^n}{V_n(1)}\ .$$

Теорема 12.19. *Кодовое расстояние кода Хэмминга равно* 3.

Доказательство. Кодовое расстояние не меньше 3, так как норма любого ненулевого набора, получаемого суммированием строчек таблицы, не меньше трёх: если суммируем не менее трёх строчек, то в левой части будет не менее трёх единиц; если суммируем две строчки, то в левой части будет две единицы, а в правой (наборы разные) — хотя бы одна; в любой строчке таблицы также содержится не менее трёх единиц.

С другой стороны, в это таблице есть строчки с тремя единицами (в правой части такой строки стоит двоичное слово с двумя единицами и ещё одна единица всегда есть в левой части). □

Итак, код Хэмминга исправляет 1 ошибку и при этом достигает оценки Хэмминга. Шары радиуса 1 с центрами в словах из кода Хэмминга не только не пересекаются, но и покрывают без пропусков весь двоичный куб (задают *разбиение* куба).

Пример 12.20. Составим таблицу для кода Хэмминга длины 7:

1	0	0	0	1	0	1
0	1	0	0	1	1	0
0	0	1	0	0	1	1
0	0	0	1	1	1	1

Складываем строчки произвольным образом и получаем 16 возможных комбинаций. Ими можно закодировать 16 сообщений, например, все 10 цифр и знаки операций. При передаче с помощью кода Хэмминга можно исправить одну ошибку, возникающую при передаче. □

12.3 Циклические коды

Одной из самых важных конструкций кодов являются циклические коды.

Определение 12.21. Линейный код C называется *циклическим*, если он инвариантен относительно циклических сдвигов: из того, что набор $(\alpha_0, \dots, \alpha_{n-1})$ принадлежит C, следует, что и всякий набор $(\alpha_s, \alpha_{s+1}, \dots, \alpha_{n-1}, \alpha_0, \dots, \alpha_{s-1})$ принадлежит C.

Определение циклического кода комбинаторное. Его можно применить и к нелинейным кодам. Однако для линейных кодов оно имеет очень простой алгебраический смысл.

Рассмотрим кольцо вычетов $R_n = \mathbb{F}_2[x]/(x^n - 1)$. Оно является векторным пространством размерности n над полем $\mathbb{F}_2$. Остатки при делении на $x^n - 1$ — это многочлены степени меньше n и 0. Каждый такой остаток однозначно задаётся первыми n коэффициентами: двоичному набору $(a_0, \dots, a_{n-1})$ отвечает многочлен

$$a_0 + a_1 x + \dots + a_{n-1} x^{n-1},$$

то есть вычеты $[1], [x], \dots, [x^{n-1}]$ образуют базис в R_n. Мы считаем этот базис выделенным и отождествляем элементы R_n с двоичными словами по указанному выше правилу (координаты в выделенном базисе).

Теорема 12.22. *Циклические коды — это в точности идеалы в кольце*

$$R_n = \mathbb{F}_p[x]/(x^n - 1).$$

Доказательство. Идеал $J \subseteq R_n$ — это подпространство (подгруппа по сложению).

Кроме того, должно выполняться свойство втягивания. В частности, пространство должно быть замкнуто относительно умножения на вычет $[x]$.

Запишем действие умножения на x в координатах в выделенном базисе:

$$x \cdot 1 = x^1,\ x \cdot x^1 = x^2,\ \dots, x \cdot x^{n-1} = x^n = 1$$

(так как $x^n - 1$ принадлежит идеалу, порождающему циклическое кольцо R_n). Значит,

$$x(a_0 + a_1 x + \dots + a_{n-1} x^{n-1}) = a_0 x + a_1 x^2 + \dots + a_{n-1},$$
$$(a_0, a_1, \dots, a_{n-1}) \overset{x}{\mapsto} (a_{n-1}, a_0, \dots, a_{n-2}).$$

Поэтому J замкнут относительно циклических сдвигов в координатах

Контрольный вопрос 12.23. Чему отвечает циклический сдвиг на s позиций?

В обратную сторону: если J — циклическое подпространство, то вычеты, его составляющие, замкнуты относительно умножения на x, x^2, … Поэтому из $[f] \in J$ следует $[f] \cdot [x^k] \in J$ и потому $[f] \cdot [g] \in J$ для любого многочлена $g \in \mathbb{F}_2[x]$.

Контрольный вопрос 12.24. Проверьте последнее утверждение.

Поэтому I — идеал. □

Итак, циклические коды — это то же самое, что идеалы в циклическом кольце. Как устроены такие идеалы и сколько их?

Во-первых, заметим, что кольцо R_n является гомоморфным образом кольца многочленов $\mathbb{F}_2[x]$ и потому является кольцом главных идеалов (см. выше лемму 10.30).

Во-вторых, элементы, различающиеся на обратимый элемент кольца, порождают один и тот же идеал.

Утверждение 12.25. *Обратимые элементы в кольце R_n — это классы вычетов $[f]$, которые состоят из многочленов, взаимно простых с $x^n - 1$ в кольце $\mathbb{F}_2[x]$.*

Доказательство. Условие обратимости $[f] \cdot [g] = [1]$ означает, что в кольце многочленов $\mathbb{F}_2[x]$ выполняется равенство

$$f(x)g(x) = 1 + h(x)(x^n - 1)$$

Для многочлена f выполнить такое равенство выбором подходящих g и h возможно тогда и только тогда, когда f взаимно прост с $x^n - 1$. □

Умножая на подходящий обратимый элемент кольца R_n, можно преобразовать порождающий идеала к более простому виду. А именно, можно выбрать порождающим делитель многочлена $x^n - 1$.

Лемма 12.26. *Каждый идеал в циклическом кольце R_n порождён классом вычетов, содержащим некоторый делитель многочлена $x^n - 1$.*

Доказательство. Пусть $J = ([f])$ и $g = \text{НОД}(f, x^n - 1)$. Тогда $h = f/g$ взаимно прост с $x^n - 1$ и по предыдущему утверждению $J = ([f]) = ([g] \cdot [h]) = ([g])$. □

Оказывается, разным делителям многочлена $x^n - 1$ отвечают разные идеалы в кольце R_n. Чтобы это доказать, полезно дать двойственную характеризацию идеала. Пока мы задавали идеал с помощью порождающего множества: идеал $([f])$ порождён как векторное пространство классами вычетов $[f]$, $[fx]$, …, $[fx^{n-1}]$. Но векторное пространство можно задавать и как решение системы линейных уравнений (то есть линейными функционалами, которые порождают подпространство линейных функционалов, обращающихся в 0 на данном подпространстве).

По определению, для делителя f многочлена $x^n - 1$ дополнительный делитель g удовлетворяет равенству $fg = x^n - 1$. Искомые линейные уравнения легко выражаются с помощью дополнительного делителя.

Лемма 12.27. $[h] \in ([f])$ *тогда и только тогда, когда* $[gh] = 0$ *в кольце R_n (то есть многочлен gh делится на $x^n - 1$).*

Контрольный вопрос 12.28. Почему лемма задаёт систему линейных уравнений, хотя написано всего лишь одно линейное условие?

Доказательство леммы 12.27. Пусть $[h] = [q] \cdot [f] = [qf]$. Тогда

$$[gh] = [gqf] = [q(x^n - 1)] = [0]$$

в кольце R_n, то есть gh делится на $x^n - 1$.

В другую сторону: если $[gh] = 0$ в кольце R_n, то $gh = q(x^n - 1) = qfg$. Поэтому $g = qf$ (применяем закон сокращения в кольце многочленов $\mathbb{F}_2[x]$, в котором нет делителей нуля) и $[g] \in ([f])$. □

Эта лемма оказывается полезной во многих вопросах. В частности, она помогает убедиться, что разные делители $x^n - 1$ порождают разные идеалы в кольце R_n.

Теорема 12.29. *Пусть* $f_1 \mid x^n - 1$, $f_2 \mid x^n - 1$ *— делители многочлена* $x^n - 1$. *Тогда из* $([f_1]) = ([f_2])$ *следует* $f_1 = f_2$.

Доказательство. Обозначим дополнительные делители g_1 и g_2 соответственно. Из равенства идеалов следуют два включения:

$$f_1 \in ([f_2]) \quad \text{и} \quad f_2 \in ([f_1]).$$

Применяя критерий из леммы 12.27, получаем

$$[f_1 g_2] = 0 \quad \text{и} \quad [f_2 g_1] = 0,$$

то есть

$$f_1 g_2 = q'(x^n - 1) = q' f_1 g_1 \quad \text{и} \quad f_2 g_1 = q''(x^n - 1) = q' f_2 g_1.$$

Это равенства в кольце многочленов $\mathbb{F}_2[x]$. Сокращая первое на f_1, а второе на f_2, видим, что многочлен g_1 делится на g_2 и наоборот. Значит, эти многочлены различаются на обратимый элемент кольца многочленов — многочлен степени 0, то есть на ненулевую константу. В случае поля $\mathbb{F}_2$ такая константа единственна и равна 1. □

В силу этой теоремы и леммы 12.26 разнообразие циклических кодов зависит от разнообразия делителей многочлена $x^n - 1$. С помощью автоморфизма Фробениуса нетрудно найти степени неприводимых делителей $x^n - 1$ и общее количество циклических кодов длины n.

Пример 12.30. Сколько есть циклических кодов длины 15?

Мы уже находили степени неприводимых делителей многочлена $x^{15} - 1 \in \mathbb{F}_2[x]$ в примере 11.44. Многочлен $x^{15} - 1$ разлагается в произведение одного неприводимого многочлена степени 1, одного неприводимого многочлена степени 2 и трёх неприводимых многочленов степени 4. Всего 5 различных неприводимых делителей, поэтому общее количество делителей равно $2^5 = 32$. □

Пример 12.31. Сколько есть циклических кодов длины 14?

Нужно опять найти степени неприводимых делителей многочлена $x^{14} - 1 \in \mathbb{F}_2[p]$ как в примере 11.44. Прежде всего заметим, что $x^{14} - 1 = (x^7 - 1)^2$, поэтому нужно искать неприводимые делители многочлена $x^7 - 1$. Орбиты корней этого многочлена под действием автоморфизма Фробениуса ($x \mapsto x^2$ в случае характеристики 2) имеют вид

$$(\alpha^0 = 1),\ (\alpha, \alpha^2, \alpha^4),\ (\alpha^3, \alpha^6, \alpha^5)$$

(показатели считаем по модулю 7).

Поэтому $x^{14}-1 = f_1^2 f_2^2 f_3^2$, где $\deg f_1 = 1$, $\deg f_2 = \deg f_3 = 3$. Всего получается $3^3 = 27$ различных делителей. □

Контрольный вопрос 12.32. Вернувшись к примеру 11.45, найдите количество циклических кодов длины 23.

Пример 12.33. Пусть $n = 7$. Запишем разложение на неприводимые множители:

$$x^7 - 1 = (1+x)(1+x^2+x^3)(1+x+x^3).$$

В качестве порождающего идеала многочлена возьмём последний множитель $g = 1 + x + x^3$ (точнее, содержащий его класс вычетов). Умножая его на степени x получим базис в подпространстве, которое является кодом:

(1101000)	$[g]$
(0110100)	$[g \cdot x]$
(0011010)	$[g \cdot x^2]$
(0001101)	$[g \cdot x^3]$

Можно проверить, что кодовое расстояние для этого кода равно 3.

В самом деле, умножение на x в кольце вычетов по модулю $x^7 - 1$ приводит к циклическому сдвигу коэффициентов. Если есть набор коэффициентов с двумя единицами, то расстояние между единицами в наборе не больше 2 (либо в одну сторону, либо в другую). Но тогда есть такой циклический сдвиг этого набора, у которого единицы помещаются в младшей (левой половине). Значит, для некоторого многочлена f степени не выше 3 с двумя ненулевыми коэффициентами выполняется

$$f(x) = g(x)h(x) + q(x)(x^7 - 1),$$

то есть f делится на $g = 1 + x + x^3$, что невозможно.

Таким образом, минимальное число единиц (равное кодовому расстоянию) для этого кода равно 3. □

Каковы параметры циклических кодов? Размерность найти легко.

Лемма 12.34. *Если циклический код порождён делителем $f \mid x^n - 1$, то его размерность как подпространства равна $n - \deg f$, а количество элементов в этом коде равно $2^{n-\deg f}$.*

Доказательство. Как мы уже установили, циклический код — это идеал $([f])$ в кольце R_n. Пусть g — дополнительный делитель к f. Его степень равна как раз $n - \deg f$.

Рассмотрим два многочлена r_1, r_2 степени меньше $\deg g$ (нулевой многочлен также возможен). Им соответствуют элементы $[r_1 f]$, $[r_2 f]$ идеала $([f])$. Проверим, что эти элементы различны. Действительно, $[f(r_1 - r_2)]$ не делится на $x^n - 1$, так как $\deg(r_1 - r_2) < \deg d$ и потому $\deg(f(r_1 - r_2)) < n$ (логарифмическое свойство степени).

Контрольный вопрос 12.35. Почему найденные элементы образуют подпространство размерности $\deg g$?

Убедимся, что мы нашли все элементы идеала.

Пусть $[hf] \in ([f])$. Разделим h на дополнительный к f делитель g с остатком: $h = h_1 g + r$, $\deg r < \deg g$ или $r = 0$. Тогда $hf = h_1 fg + rf = h_1(x^n - 1) + rf$, то есть $[hf] = [rf]$. □

Вопрос о кодовом расстоянии циклического кода в общем случае безнадёжно труден. Но есть несколько важных конструкций циклических кодов с хорошими параметрами. Далее мы приводим несколько примеров.

12.4 Код Хэмминга как циклический код

Разложение многочлена $x^n - 1 = x^{2^m-1} - 1$ на неприводимые множители мы уже находили. Это все неприводимые многочлены в кольце многочленов $\mathbb{F}_2[x]$, степени которых делят m (кроме многочлена x).

Если взять циклический код (идеал в циклическом кольце R_n), порождённый подходящим неприводимым многочленом степени m, то получим как раз код Хэмминга.

Пусть α — порождающий мультипликативной группы поля $\mathbb{F}_{2^m}$, а f — его минимальный многочлен. Он, как мы уже выясняли, неприводим и его степень равна m (так как поле $\mathbb{F}_2(\alpha)$ совпадает с $\mathbb{F}_{2^m}$).

Пусть $J = ([f])$ в кольце R_n. Размерность этого кода получается из леммы 12.34. Она равна $n - \deg f = 2^m - 1 - m$, то есть совпадает с размерностью кода Хэмминга H_n.

Теорема 12.36. *Кодовое расстояние кода J равно* 3.

Доказательство. Пусть $[h] \in J$ и количество единиц среди координат $[h]$ в базисе мономов $[1]$, $[x]$, $\ldots$, $[x^{n-1}]$ меньше 3. Поскольку код циклический, то он замкнут относительно циклических сдвигов координат. Так что без ограничения общности считаем, что у константы $[1]$ координата равна 1.

Если это единственная единичная координата, то получаем $[1] = [h_1 f]$, то есть $1 = h_1 f + h_2(x^n - 1)$, что невозможно, так как f делит $x^n - 1$.

Если ещё и координата при x^k, $0 < k < n$, равна единице (а все остальные равны нулю), то получаем $[1 + x^k] = [h_1 f]$, то есть $1 + x^k = h_1 f + h_2(x^n - 1)$. Слагаемые в правой части равенства обращаются в 0 в точке α (так как f выбран минимальным многочленом α). Но тогда и левая часть равенства обращается в 0, то есть $\alpha^k + 1 = 0$, что равносильно $\alpha^k = 1$ в характеристике 2. Но это невозможно: ведь порядок α в мультипликативной группе поля $\mathbb{F}^*_{2^m}$ равен $2^m - 1 = n$.

Итак, кодовое расстояние построенного кода не меньше 3. Больше оно быть не может из-за границы Хэмминга: количество элементов в коде с кодовым расстоянием 3 не превосходит $2^n/(1 + n) = 2^{2^m-m-1}$, а размерность кода как раз равна $n - m = 2^m - m - 1$. □

У построенного кода те же параметры, что и кода Хэмминга. Но пока не вполне понятно, почему это по существу один и тот же код. «По существу» здесь означает, что один код может быть получен из другого перестановкой координат (обратите внимание, что после перестановки координат код перестанет быть циклическим, хотя «по существу» ничего не изменилось).

Любой элемент поля $\mathbb{F}_{2^m}$ выражается как линейная комбинация степеней α^i, $0 \leqslant i < m$, с коэффициентами из $\mathbb{F}_2$.

Контрольный вопрос 12.37. Почему?

Запишем такие выражения для всех степеней α, начиная с m-й и до $n-1$ (порядок степеней в правых частях равенств убывающий):

$$\begin{aligned}
\alpha^m &= a_{m-1}^{(m)}\alpha^{m-1} + a_{m-2}^{(m)}\alpha^{m-2} + \cdots + a_1^{(m)}\alpha + a_0^{(m)};\\
\alpha^{m+1} &= a_{m-1}^{(m+1)}\alpha^{m-1} + a_{m-2}^{(m+1)}\alpha^{m-2} + \cdots + a_1^{(m+1)}\alpha + a_0^{(m+1)};\\
&\dots\\
\alpha^{n-1} &= a_{m-1}^{(n-1)}\alpha^{m-1} + a_{m-2}^{(n-1)}\alpha^{m-2} + \cdots + a_1^{(n-1)}\alpha + a_0^{(m+1)}.
\end{aligned}$$

Всего получили $n - m = 2^m - m - 1$ разложений по начальным степеням α.

Многочлены

$$f_s(x) = x^s + \sum_{i=0}^{m-1} a_i^{(s)}, \quad m \leqslant s < n,$$

обращаются в 0 в точке α. Так как f — минимальный многочлен для α, то все они делятся на α, поэтому $[f_s] \in J$.

Составим, как и при первом построении кода Хэмминга, таблицу $(2^m - m - 1) \times (2^m - 1)$, записав в её левую часть единичную матрицу, а в правую — коэффициенты разложений $a_i^{(j)}$, $0 \leqslant i < m$, $m \leqslant j < n$. Легко убедиться, аналогично доказательству теоремы 12.36, что среди коэффициентов разложений для любой степени хотя бы два единичных (иначе придём к противоречию с выбором α как элемента порядка $n - 1$ в мультипликативной группе $\mathbb{F}_{2^m}^*$). По той же причине среди строк правой части таблицы нет одинаковых (иначе $\alpha^s - \alpha^p = 0$ при $m \leqslant s, p < n$ и $s \neq p$). Итак, мы фактически построили ту же таблицу, что и в первый раз. Теперь мы лишь уточнили порядок, в котором идут столбцы таблицы (и в левой части, и в правой).

Представление кода Хэмминга как циклического позволяет задавать его очень компактно.

Пример 12.38. Пусть $n = 15$, $m = 4$. В качестве порождающего мультипликативной группы $\mathbb{F}_{16}^*$ возьмём корень многочлена $x^4 + x + 1$ (обсуждение этого выбора см. в примере 11.51).

Запишем коэффициенты этого многочлена в возрастающем порядке в виде двоичного вектора длины 15:

$$(1,\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0).$$

Суммы циклических сдвигов этого вектора дают в точности слова из кода Хэмминга. Поэтому таких сумм $2^{15-4} = 2^{11}$ и все ненулевые из них имеют вес Хэмминга по крайней мере 3. □

12.5 Коды БЧХ

Это циклические коды длины $n = 2^m - 1$. Они обобщают код Хэмминга. Конструкция этих кодов придумана Боузом, Чоудхури и Хоквингемом. Поэтому эти коды и называются БЧХ-кодами.

Длина этих кодов, как и у кода Хэмминга, $n = 2^m - 1$. Допустим, мы хотим исправлять r ошибок. Чтобы построить БЧХ код с этим свойством, необходим многочлен $f(x) \in \mathbb{F}_2[x]$, корнями которого являются α, α^2, …, α^{2r}. Тогда код $B_{n,r}$ будет идеалом в циклическом кольце $R_n = \mathbb{F}_2[x]/(x^n - 1)$, который порождён классом вычетов $[f]$.

Утверждение 12.39. *Размерность k кода $B_{n,r}$ не меньше $n - mr$.*

Доказательство. Многочлен f можно получить, перемножив минимальные многочлены степеней α, α^2, …, α^{2r}. Однако дважды включать один и тот же многочлен в это произведение необязательно.

Как мы знаем, минимальный многочлен неприводим. А поскольку корни неприводимого многочлена образуют орбиту автоморфизма Фробениуса, то у элементов α^i и α^{2i} минимальный многочлен один и тот же (характеристика поля равна 2).

Поэтому в произведение достаточно взять не более r минимальных многочленов. Степень каждого из них не превосходит m. Отсюда получается оценка $\deg f \leqslant mr$. Оценка на размерность БЧХ кода следует теперь из общей формулы для размерности циклического кода, лемма 12.34. □

Как оценить кодовое расстояние БЧХ кода? Простое рассуждение для кода Хэмминга нужно усовершенствовать.

Лемма 12.40. *Пусть $n = 2^m - 1$, $g \in \mathbb{F}_2[x]$, а $\langle\alpha\rangle = \mathbb{F}_{n+1}^*$. Если $g(\alpha^s) = 0$ при $1 \leqslant s \leqslant 2r$ и степень g меньше n, то у g не менее $2r+1$ ненулевого коэффициента.*

Доказательство. Коэффициенты g — это нули и единицы. Фактически, g является суммой ℓ различных мономов. Обозначим показатели степени в этих мономах через a_1, a_2, …, a_ℓ. Все они меньше n по условию. В выбранных обозначениях

$$g(x) = x^{a_1} + x^{a_2} + \cdots + x^{a_\ell}.$$

Условия $g(\alpha^s) = 0$ в развёрнутом виде записываются как

$$(\alpha^s)^{a_1} + (\alpha^s)^{a_2} + \cdots + (\alpha^s)^{a_\ell} = 0,$$

что равносильно

$$(\alpha^{a_1})^s + (\alpha^{a_2})^s + \cdots + (\alpha^{a_\ell})^s = 0. \tag{12.1}$$

Они выполняются для $1 \leqslant s \leqslant 2r$.

Предположим, что $\ell \leqslant 2r$. Ограничимся условиями (12.1) для $1 \leqslant s \leqslant \ell$. Перепишем их в другом виде. Определим в векторном пространстве $\mathbb{F}_{2^m}^{\ell}$ наборы векторов

$$\begin{aligned} v_1 &= \bigl(\alpha^{a_1}, (\alpha^{a_1})^2, \ldots, (\alpha^{a_1})^{\ell}\bigr), \quad u_1 = \alpha^{-a_1} v_1; \\ v_2 &= \bigl(\alpha^{a_2}, (\alpha^{a_2})^2, \ldots, (\alpha^{a_2})^{\ell}\bigr), \quad u_2 = \alpha^{-a_2} v_2; \\ &\ldots \qquad\qquad\qquad\qquad\qquad\quad \ldots \\ v_{\ell} &= \bigl(\alpha^{a_{\ell}}, (\alpha^{a_{\ell}})^2, \ldots, (\alpha^{a_{\ell}})^{\ell}\bigr), \quad u_{\ell} = \alpha^{-a_{\ell}} v_{\ell}. \end{aligned}$$

Тогда из (12.1) следует линейная зависимость этих векторов

$$\sum_{i=1}^{\ell} v_i = \sum_{i=1}^{\ell} \alpha^{a_i} u_i = 0.$$

Действительно, равенство (12.1) в точности означает, что s-я координата в сумме $v_1 + \cdots + v_{\ell}$ равна 0.

Линейная зависимость векторов u_i означает линейную зависимость строк квадратной матрицы

$$\begin{pmatrix} 1 & \alpha^{a_1} & (\alpha^{a_1})^2 & \ldots & (\alpha^{a_1})^{\ell-1} \\ 1 & \alpha^{a_2} & (\alpha^{a_2})^2 & \ldots & (\alpha^{a_2})^{\ell-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 1 & \alpha^{a_{\ell}} & (\alpha^{a_{\ell}})^2 & \ldots & (\alpha^{a_{\ell}})^{\ell-1} \end{pmatrix}$$

По теореме о ранге из линейной алгебры линейная зависимость строк квадратной матрицы влечёт линейную зависимость столбцов. Это означает, что для некоторых $h_i \in \mathbb{F}_{2^m}$, $0 \leqslant i < \ell$, не все из которых равны нулю, выполняются равенства

$$\sum_{i=0}^{\ell-1} h_i (\alpha^{a_s})^i = 0, \qquad 1 \leqslant s \leqslant \ell.$$

Другими словами, некоторый ненулевой многочлен $h \in \mathbb{F}_{2^m}[x]$ степени меньше ℓ обращается в 0 во всех точках α^{a_i}, $1 \leqslant s \leqslant \ell$.

Так как все a_i меньше n, среди степеней α^{a_i} нет одинаковых элементов поля. То есть мы нашли ненулевой многочлен степени меньше ℓ, у которого по крайней мере ℓ корней. Приходим к противоречию с леммой о числе корней многочлена с коэффициентами в поле.

Это означает, что сделанное предположение ложно и $\ell > 2r$. □

Из доказанных утверждений непосредственно следует теорема.

Теорема 12.41. *Для параметров кода $B_{n,r}$ выполняются соотношения $n = 2^m - 1$, $k \geqslant n - mr$, $d > 2r$.*

Такой код исправляет r ошибок.

Пример 12.42. Пусть $n = 15$, $m = 4$, $r = 2$. В качестве порождающего мультипликативной группы $\mathbb{F}_{16}^*$ как и в случае кода Хэмминга (пример 12.38) берём корень многочлена $f_1 = x^4 + x + 1$.

Корнями этого многочлена являются также α^2 и α^4. Для построения БЧХ кода с заданными параметрами нужен ещё многочлен, корнем которого является α^3. Как следует из обсуждения в примере 11.51, это $f_2 = x^4 + x^3 + x^2 + x + 1$.

Порождающий БЧХ кода — это произведение этих многочленов $f_1 f_2$:

$$\begin{array}{cccccccccc}
1 & +x & +x^2 & +x^3 & +x^4 & & & & & \\
 & +x & +x^2 & +x^3 & +x^4 & +x^5 & & & & \\
 & & & & +x^4 & +x^5 & +x^6 & +x^7 & +x^8 & \\
\hline
1 & & & & +x^4 & & +x^6 & +x^7 & +x^8 &
\end{array}$$

Запишем коэффициенты этого многочлена в возрастающем порядке в виде двоичного вектора длины 15:

$$(1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 1,\ 1,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0).$$

Суммы циклических сдвигов этого вектора дают в точности слова из кода БЧХ. Поэтому таких сумм $2^{15-8} = 2^7 = 128$ и расстояния Хэмминга между ними не меньше 5.

Сравним мощность этого кода с границей Хэмминга. При данных параметрах она имеет вид

$$U_{15,2} = \frac{2^{15}}{1 + 15 + \binom{15}{2}} = \frac{2^{15}}{121}.$$

Отношение $|B_{15,2}|/U_{15,2} = 121/2^8$ чуть меньше $1/2$. Таким образом, данный код БЧХ теряет по сравнению с идеальным кодом (несуществующим) чуть больше 1 бита. Не так уж и мало с точки зрения скорости передачи: $1/15 \approx 6\%$. □

Пример 12.43. Пусть $n = 15$, $m = 4$, $r = 3$, то есть код, исправляющий три ошибки. Многочлен из предыдущего примера имеет корнями α, α^2, α^3, α^4 и α^6.

Контрольный вопрос 12.44. Проверьте это утверждение.

Нужен ещё многочлен, корнем которого является α^5. Этот элемент принадлежит подполю $\mathbb{F}_4$, поэтому его минимальный многочлен $1 + x + x^2$. Порождающий БЧХ кода получается как произведение:

$$\begin{array}{cccccccccccc}
1 & & & +x^4 & & +x^6 & +x^7 & +x^8 & & & & \\
 & x & & & +x^5 & & +x^7 & +x^8 & +x^9 & & & \\
 & & x^2 & & & +x^6 & & +x^8 & +x^9 & +x^{10} & & \\
\hline
1 & +x & +x^2 & +x^4 & +x^5 & & & +x^8 & & +x^{10} & &
\end{array}$$

Запишем коэффициенты этого многочлена в возрастающем порядке в виде двоичного вектора длины 15:

$$(1,\ 1,\ 1,\ 0,\ 1,\ 1,\ 0,\ 0,\ 1,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0).$$

Суммы циклических сдвигов этого вектора дают в точности слова из кода БЧХ, исправляющего три ошибки. Таких сумм $2^{15-10} = 2^5 = 32$ и расстояния Хэмминга между ними не меньше 7.

Сравним мощность этого кода с границей Хэмминга. При данных параметрах она имеет вид

$$U_{15,3} = \frac{2^{15}}{1 + 15 + \binom{15}{2} + \binom{15}{3}} = \frac{2^{15}}{576}.$$

Отношение $|B_{15,3}|/U_{15,3} = 576/2^{10}$ теперь чуть больше $1/2$. Таким образом, данный код БЧХ теряет по сравнению с идеальным кодом (несуществующим) чуть меньше 1 бита. □

Если $n \to \infty$, а $r = O(1)$, параметры БЧХ кода близки к границе Хэмминга. Заметим, что $V_n(r) \sim n^r/r!$ при такой асимптотике. Поэтому

$$\log_2 U_{n,r} - \log_2 |B_{n,r}| \leqslant n - r\log_2 n + \log r! - n + mr = O(1),$$

так как $r! = O(1)$ при сделанных предположениях.

Таким образом, при растущей длине кода и фиксированном количестве ошибок разница между границей Хэмминга и кодом БЧХ всего константа битов.

Однако эта константа быстро растёт с ростом r. Поэтому БЧХ коды не годятся для построения *асимптотически оптимальных кодов*, у которых и скорость передачи k/n, и доля исправляемых ошибок d/n ограничены снизу положительной константой.

12.6 Двоичный код Голея

Совершенные коды достигают границы Хэмминга. Таковы коды Хэмминга. Оказывается, есть всего один двоичный совершенный код, который не является кодом Хэмминга. Он называется кодом Голея. Ему отвечают такие значения параметров: $n = 23$, $k = 12$, $d = 7$. То есть код Голея исправляет 3 ошибки и делает это наилучшим возможным образом.

Прежде всего проверим, что код с такими параметрами и впрямь достигает границы Хэмминга. Для этого нужно найти количество точек в 23-мерном шаре радиуса 3:

$$1 + \binom{23}{1} + \binom{23}{2} + \binom{23}{3} = 2^{11}. \tag{12.2}$$

Если вложить без пересечений в 23-мерный булев куб 2^{12} шаров радиуса 3, то из (12.2) следует, что они покроют все точки булева куба. Код с такими параметрами будет совершенным.

Код Голея также циклический, как и предыдущие конструкции. Его можно задать как суммы циклических сдвигов вектора

$$0\,0\,0\,0\,1\,0\,1\,0\,0\,1\,1\,0\,0\,1\,1\,0\,1\,0\,1\,1\,1\,1\,0\,. \tag{12.3}$$

Конечно, трудно понять, глядя на координаты этого вектора, почему в таком коде будет именно 2^{12} точек и почему расстояния между ними не меньше 7.

Чтобы понять, откуда берётся строка Голея (12.3), нужно увидеть в ней таблицу квадратичных вычетов по модулю 23. Действительно, припишем к этой строке

вычеты по модулю 23:

22	21	20	19	18	17	16	15	14	13	12	11	10	9	8	7	6	5	4	3	2	1	0
0	0	0	0	1	0	1	0	0	1	1	0	0	1	1	0	1	0	1	1	1	1	0.

Теперь уже можно прямым вычислением проверить, что единицы в строке Голея стоят в точности на местах, которые являются квадратичными вычетами. Эту проверку можно выполнить чуть быстрее, чем прямое использование критерия квадратичного вычета.

Пользуясь этим критерием, убедимся в том, что 2 — квадратичный вычет по модулю 23:

$$2^{(23-1)/2} = 2^{11} = 2 \cdot 32^2 \equiv 2 \cdot 81 = 162 = 1 + 23 \cdot 7 \equiv 1 \pmod{23}.$$

Подгруппа квадратичных вычетов мультипликативной группы $\mathbb{F}_{23}^*$ имеет индекс 2, то есть порядок 11, а это простое число. Значит, любой неединичный элемент этой подгруппы получается как степень 2. Последовательно умножая на 2, выписываем все квадратичные вычеты:

$$1,\ 2,\ 4,\ 8,\ 16,\ 9,\ 18,\ 13,\ 3,\ 6,\ 12.$$

Мы эти числа уже видели в примере 11.45: это показатели степеней примитивного корня ω 23-й степени из единицы у корней одного из неприводимых делителей степени 11 многочлена

$$x^{23} - 1 = (x - 1)f(x)g(x),$$

где $\deg f = \deg g = 11$. У другого показателями степеней ω будут квадратичные невычеты (то есть все остальные ненулевые вычеты). Будем считать, что квадратичные вычеты появляются как показатели степеней у многочлена f в написанном выше разложении $x^{23} - 1$ на неприводимые.

Обозначим Q_{23} множество квадратичных вычетов по модулю 23, а N_{23} — множество квадратичных невычетов.

Определим многочлен $q(x)$, коэффициенты которого в убывающем порядке задаются строкой (12.3):

$$q(x) = \sum_{i \in Q_{23}} x^i = x^1 + x^2 + x^3 + x^4 + x^6 + x^8 + x^9 + x^{12} + x^{13} + x^{16} + x^{18}.$$

Код Голея — это, идеал в циклическом кольце R_{23}, порождённый многочленом $q(x)$: $G_{23} = ([q])$. Из леммы 12.26 об идеалах в циклических кольцах мы знаем, что этот идеал порождён также каким-то делителем многочлена $x^{23} - 1$.

Поскольку 2 — квадратичный вычет по модулю 23 и умножение на 2 по модулю 23 сохраняет квадратичные вычеты, то

$$q(\omega)^2 = \left(\sum_{i \in Q_{23}} \omega^i \right)^2 = \sum_{i \in Q_{23}} \omega^{2i} = \sum_{i \in Q_{23}} \omega^i = q(\omega)$$

(в первом равенстве мы учли, что характеристика поля равна 2). Но это означает, что $q(\omega) \in \mathbb{F}_2$ (корень уравнения $x^2 - x = 0$). Аналогично для сумм степеней, которые

являются квадратичными невычетами (умножение на 2 их также сохраняет, это же класс смежности по подгруппе). Поэтому

$$A = q(\omega) = \sum_{i \in Q_{23}} \omega^i \in \mathbb{F}_2, \quad B = \bar{q}(\omega) = \sum_{i \in N_{23}} \omega^i \in \mathbb{F}_2, \quad A + B = 1$$

(последнее равенство вытекает из $\omega^{23} = 1$, $\omega \neq 1$). Здесь мы используем ещё один многочлен

$$\bar{q}(x) = \sum_{i \in N_{23}} x^i = x^5 + x^7 + x^{10} + x^{11} + x^{14} + x^{15} + x^{17} + x^{19} + x^{20} + x^{21} + x^{22}.$$

Одно из чисел A, B равно 1, а другое 0. Установить, какое именно равно нулю, невозможно. Вся наша конструкция держится на выборе примитивного корня 23-й степени из единицы. Но если ω — примитивный корень (то есть порождающий группы корней 23-й степени из единицы), то и ω^{-1} также является порождающей группы корней 23-й степени из единицы и примитивным корнем.

Однако -1 является квадратичным невычетом по модулю 23. Поэтому, если $q(\omega) = 0$, то $q(\omega^{-1}) = \bar{q}(\omega) = 1$ и наоборот.

Не ограничивая общности, мы полагаем, что $q(\omega) = 0$ (в противном случае заменим примитивный корень на ω^{-1}). Но минимальный многочлен ω — это $f(x)$ по нашему выбору, то есть $q(x)$ делится на $f(x)$.

Других общих делителей у $q(x)$ и $x^{23} - 1$ нет. Действительно, $q(\omega^{-1}) = \bar{q}(\omega) = 1$, поэтому $q(x)$ не делится на $g(x)$ (так как у последнего есть корень ω^{-1}). Кроме того, $q(1) = 1 \neq 0$ (напомним ещё раз, что характеристика поля равна 2 и потому $11 = 1$). Значит, $q(x)$ не делится на $x - 1$.

Мы перебрали все неприводимые делители многочлена $x^{23} - 1$ и убедились, что $\gcd(q(x), x^{23} - 1) = f(x)$. Значит, как было показано в лемме 12.26, код Голея порождается также многочленом $f(x)$.

Как следствие из общей леммы 12.34 получаем, что параметры $G_{23} = ([f])$ равны $n = 23$, $k = 23 - 11 = 12$. Осталось лишь убедиться, что кодовое расстояние именно 7.

Оценку кодового расстояния кода Голея даёт следующая теорема.

Теорема 12.45. *Количество кодовых слов в G_{23} равно 2^{12}, а кодовое расстояние равно 7.*

Доказательство. Размерность кода мы уже нашли.

Осталось найти кодовое расстояние. Из (12.2) и того, что размерность кода Голея равна 12 следует, что кодовое расстояние не больше 7.

Весом многочлена (точнее, класса вычетов в циклическом кольце R_n) назовём вес Хэмминга вектора его коэффициентов (коэффициентов остатка при делении на $x^n - 1$).

Докажем сначала, что если вес многочлена (точнее соответствующего класса вычетов) в кольце $R_{23} = \mathbb{F}_2/(x^{23} - 1))$ из кода Голея нечетен, то он не меньше 7.

Возьмем элемент $[c(x)] \in G_{23}$ нечетного веса $w = 2s + 1$, считаем не ограничивая общности, что $\deg c(x) < 23$.

Корнями $c(x)$ будут такие степени ω^k примитивного корня ω, для которых $k \in Q_{23}$, то есть k является квадратичным вычетом по модулю 23.

Рассмотрим также класс вычетов $[\bar{c}(x)] = [c(x^{-1})] = [c(x^{22})]$. Многочлен $\bar{c}(x)$ выберем так, что $\deg \bar{c}(x) < 23$ (например, возьмём остаток при делении $c(x^{22})$ на $x^{23} - 1$). В вычислениях удобно будет использовать представление $\bar{c}(x)$ как $c(x^{-1})$. Вес многочлена $\bar{c}(x)$ также равен w.

Контрольный вопрос 12.46. Докажите это утверждение.

Поскольку -1 является квадратичным невычетом по модулю 23, корнями $\bar{c}(x)$ будут такие ω^k, что $k \in N_{23}$ (произведение вычета на невычет является невычетом, поэтому $k \in Q_{23}$ равносильно $-k \in N_{23}$).

Поэтому произведение $c(x)\bar{c}(x)$ делится на взаимно простые многочлены $f(x)$ и $g(x)$, а значит, и на их произведение

$$f(x)g(x) = \frac{x^{23} - 1}{x - 1} = 1 + x + x^2 + \cdots + x^{22}.$$

Но $c(1) = \bar{c}(1) \neq 0$, так как w нечетен, то есть $c(x)\bar{c}(x)$ не делится на $x - 1$. Значит, выполняется равенство

$$c(x)\bar{c}(x) \bmod x^{23} - 1 = f(x)g(x).$$

При перемножении $c(x)$ и $\bar{c}(x)$ получается не более w^2 мономов, из которых w равны 1 (произведения x^k и x^{-k}). Чтобы в результате получилось не меньше 23 мономов, должно выполнятся неравенство

$$w^2 - w + 1 \geqslant 23,$$

из которого получаем $w \geqslant 7$.

Пусть теперь вес w многочлена из кода Голея $[c(x)] \in G_{23}$ чётен.

Докажем, что $w > 4$. Для этого заметим, что

$$f(\omega) = f(\omega^2) = f(\omega^3) = f(\omega^4) = 0,$$

так как 1, 2, 3, 4 являются квадратичными вычетами по модулю 23. Поэтому применимо то же рассуждение, что и в случае кодов БЧХ.

Наконец докажем, что если вес многочлена из кода Голея чётен, то он делится на 4. Это, в частности, исключает слова веса 6. (Рассуждение заимствовано из книги [3].)

Пусть $c(x)$ — многочлен чётного веса w. Тогда $c(1) = 0$, то есть $(x - 1) \mid c(x)$, откуда следует, что $c(x)\bar{c}(x) = 0 \bmod x^{23} - 1$.

Из равенства $c(x)\bar{c}(x) = 0$ получаем равенства для коэффициентов c_i многочлена $c(x)$, рассматриваемых как целые числа: для $r = 0, \ldots, 22$ имеем

$$\sum_{i-j \equiv r \bmod 23} c_i c_j = 2a_r, \quad a_i \text{ — целые числа.}$$

Меняя местами i и j, получаем равенства $a_r = a_{23-r}$ при $r \neq 0$. Поэтому

$$w(w-1) = \sum_{i=0}^{22} \sum_{j \neq i} c_i c_j = 2 \sum_{r=1}^{22} a_r = 4 \sum_{r=1}^{11} a_r,$$

то есть из четности w следует, что w делится на 4. □

Список литературы

[1] *Алексеев В. Б.* Теорема Абеля в задачах и решениях. М.: МЦНМО, 2001.

[2] *Беклемишев Д. В.* Курс аналитической геометрии и линейной алгебры. М.: Наука, 1988.

[3] *Блейхут Р.* Теория и практика кодов, контролирующих ошибки. М.: Мир, 1986.

[4] *Ван дер Варден Б.:Л.* Алгебра. М.: Наука, 1976.

[5] *Винберг Э. Б.* Курс алгебры. М.: Факториал, 1999.

[6] *Виноградов И. М.* Основы теории чисел. М.: Наука, 1972.

[7] *Влэдуц С. Г.*, *Ногин Д. Ю.*, *Цфасман М. А.* Алгеброгеометрические коды. Основные понятия. М.: МЦНМО, 2003.

[8] *Гельфанд И. М.* Лекции по линейной алгебре. М.: Наука, 1971.

[9] *Горенстейн Д.* Конечные простые группы. Введение в их классификацию. М.: Мир, 1985.

[10] *Каргаполов М. И.*, *Мерзляков Ю. И.* Основы теории групп. М.: Наука, 1977.

[11] *Кострикин А. И.* Введение в алгебру. М.: Наука, 1977.

[12] Кострикин А. И. (ред.) Сборник задач по алгебре. М.: Факториал, 1995.

[13] *Кострикин А. И.*, *Манин Ю. И.* Линейная алгебра и геометрия. М.: Наука, 1986.

[14] *Курош А. Г.* Курс высшей алгебры. М.: Наука, 1975.

[15] *Ленг С.* Алгебра. М.: Мир, 1968.

[16] *Леонтьев В. К.* Комбинаторика и информация. Часть I. Комбинаторный анализ. М.: МФТИ, 2015.

[17] *Лидл Р.*, *Нидеррейтер Х.* Конечные поля. М.: Мир, 1989.

[18] *Магнус В., Каррас А., Солитэр Д.* Комбинаторная теория групп. М.: Наука, 1974.

[19] *Мак-Вильямс Ф. Дж., Слоан Н. Дж. А.* Теория кодов, исправляющих ошибки. М.: Связь, 1979.

[20] *Прасолов В. В.* Многочлены. М.: МЦНМО, 1999.

[21] *Прасолов В. В.* Задачи и теоремы линейной алгебры. М.: Наука. Физматлит, 1996.

[22] *Райхштейн З. Б.* Тождества Ньютона и математическая индукция // Математическое просвещение. Серия 3. Вып. 4. 2000. С. 204–205.

[23] *Серр Ж.-П.* Курс арифметики. М.: Мир, 1972.

[24] *Холл М.* Теория групп. М.: ИЛ, 1962.