Алгебра и теория чисел*

Александр Лузгарев

Содержание

1 Наивная теория множеств **5**
1.1 Множества . . . 5
1.2 Операции над множествами . . . 6
1.3 Отображения . . . 8
1.4 Бинарные отношения . . . 12
1.5 Отношения эквивалентности . . . 12
1.6 Метод математической индукции . . . 13
1.7 Операции . . . 14

2 Элементарная теория чисел **17**
2.1 Делимость целых чисел . . . 17
2.2 Наибольший общий делитель и алгорифм Эвклида . . . 18
2.3 Свойства НОД и взаимная простота . . . 19
2.4 Линейные диофантовы уравнения . . . 21
2.5 Основная теорема арифметики . . . 23
2.6 Сравнения и классы вычетов . . . 26
2.7 Классы вычетов, действия над ними . . . 27
2.8 Кольца и поля . . . 27
2.9 Китайская теорема об остатках . . . 31
2.10 Теорема Вильсона . . . 31
2.11 Функция Эйлера . . . 32
2.12 Теорема Эйлера и малая теорема Ферма . . . 33
2.13 Алгоритм шифрования RSA . . . 34

3 Комплексные числа **36**
3.1 Определение комплексных чисел . . . 36
3.2 Комплексное сопряжение и модуль . . . 37
3.3 Тригонометрическая форма записи комплексного числа . . . 39

*Конспект лекций для механиков, 2014–2016; предварительная версия

3.4 Корни из комплексных чисел . . . 41
3.5 Корни из единицы . . . 42
3.6 Экспоненциальная форма записи комплексного числа . . . 43

4 Кольцо многочленов 45
4.1 Определение и первые свойства . . . 45
4.2 Делимость в кольце многочленов . . . 48
4.3 Многочлен как функция . . . 49
4.4 Многочлены над $\mathbb{R}$ и $\mathbb{C}$ . . . 51
4.5 Кратные корни и производная . . . 52
4.6 Интерполяция . . . 56
4.7 НОД и неприводимость . . . 58
4.8 Поля частных . . . 60
4.9 Поле рациональных функций . . . 62

5 Вычислительная линейная алгебра 67
5.1 Системы линейных уравнений и элементарные преобразования . . . 67
5.2 Метод Гаусса . . . 69
5.3 Операции над матрицами . . . 71
5.4 Матрицы элементарных преобразований . . . 76
5.5 Блочные матрицы . . . 80
5.6 Перестановки . . . 81
5.7 Определитель . . . 86
5.8 Дальнейшие свойства определителя . . . 90
5.9 Разложение определителя по строке . . . 93

6 Векторные пространства 98
6.1 Первые определения . . . 98
6.2 Подпространства . . . 100
6.3 Линейная зависимость и независимость . . . 104
6.4 Базис . . . 108
6.5 Размерность . . . 109

7 Линейные отображения 113
7.1 Первые определения . . . 113
7.2 Операции над линейными отображениями . . . 114
7.3 Ядро и образ . . . 116
7.4 Матрица линейного отображения . . . 118
7.5 Изоморфизм . . . 122
7.6 Ранг матрицы . . . 125
7.7 Фактор-пространство . . . 128
7.8 Относительный базис . . . 130

7.9 Матрица перехода 131

8 Жорданова нормальная форма **136**
8.1 Инвариантные подпространства и собственные числа 136
8.2 Собственные числа оператора над алгебраически замкнутым полем 139
8.3 Диагонализуемые операторы 143
8.4 Корневое разложение 145
8.5 Характеристический и минимальный многочлены 149
8.6 Жорданов базис для нильпотентного оператора 151
8.7 Жорданова форма 154
8.8 Комплексификация 157
8.9 Вещественная жорданова форма 162

9 Эвклидовы и унитарные пространства **165**
9.1 Эвклидовы пространства 165
9.2 Унитарные пространства 166
9.3 Норма 168
9.4 Матрица Грама 169
9.5 Процесс ортогонализации Грама–Шмидта 171
9.6 Ортогональные и унитарные матрицы 173
9.7 Ортонормированные базисы 175
9.8 Ортогональное дополнение 176
9.9 Сопряженные отображения 179
9.10 Самосопряженные операторы 181
9.11 Нормальные операторы 183
9.12 Спектральные теоремы 184
9.13 Самосопряженные, кососимметрические, унитарные, ортогональные операторы 189
9.14 Положительно определенные операторы 193

10 Теория групп **197**
10.1 Определения и примеры 197
10.2 Подгруппы 199
10.3 Классы смежности и нормальные подгруппы 201
10.4 Гомоморфизмы групп 204
10.5 Фактор-группы 206
10.6 Циклические группы 207
10.7 Теорема Лагранжа 208
10.8 Прямое произведение 210
10.9 Симметрическая группа 212
10.10Диэдральная группа 214

11 Полилинейная алгебра **217**
11.1 Полилинейные отображения 217
11.2 Тензорное произведение двух пространств 217
11.3 Тензорное произведение нескольких пространств 221
11.4 Двойственное пространство 223
11.5 Канонические изоморфизмы 224
11.6 Тензорное произведение линейных отображений 227
11.7 Тензорные пространства 229
11.8 Тензоры в классических обозначениях 230

Предметный указатель **233**

В начале каждого подраздела указана вспомогательная литература. Обозначения:

- [F] Д. К. Фаддеев, *Лекции по алгебре*, М.: Наука, 1984.
- [K1] А. И. Кострикин, *Введение в алгебру. Часть I. Основы алгебры*, 3-е изд. — М.: ФИЗМАТЛИТ, 2004.
- [K2] А. И. Кострикин, *Введение в алгебру. Часть II. Линейная алгебра*, М.: ФИЗМАТЛИТ, 2000.
- [K3] А. И. Кострикин, *Введение в алгебру. Часть III. Основные структуры*, М.: ФИЗМАТЛИТ, 2004.
- [vdW] Б. Л. ван дер Варден, *Алгебра*, М.: Мир, 1976.
- [Bog] О. В. Богопольский, *Введение в теорию групп*, Москва–Ижевск: Институт компьютерных исследований, 2002.
- [KM] А. И. Кострикин, Ю. И. Манин, *Линейная алгебра и геометрия*, М.: Наука, 1986.
- [V] И. М. Виноградов, *Основы теории чисел*, М., 1952.
- [B] А. А. Бухштаб, *Теория чисел*, М.: Просвещение, 1966.

1 Наивная теория множеств

1.1 Множества

Литература: [K1], гл. 1, § 5, п. 1; [vdW], гл. 1, § 1.

Мы не будем давать точных определений основным понятиям теории множеств, этим занимается аксиоматическая теория множеств. Наш подход к теории множеств совершенно наивен; под множеством мы будем понимать некоторый *набор* (*совокупность*, *семейство*) объектов — *элементов*. Природа этих объектов для нас не очень важна: это могут быть, скажем, натуральные числа, а могут быть другие множества. Множество полностью определяется своими элементами. Иными словами, два множества A и B равны тогда и только тогда, когда они состоят из одних и тех же элементов: $x \in A$ тогда и только тогда, когда $x \in B$.

Как задать множество? Самый простой способ — перечислить его элементы следующим образом: $A = \{1, 2, 3\}$. Сразу отметим, что каждый объект x может либо являться элементом данного множества A (это записывается так: $x \in A$), либо не являться его элементом ($x \notin A$); он не может быть элементом множества A «два раза». Поэтому запись $\{1, 2, 1, 3, 3, 2\}$ задает то же самое множество, что и запись $\{1, 2, 3\}$, и запись $\{2, 3, 1\}$.

Прямое перечисление может задать только конечное множество. Для задания бесконечных множеств можно использовать неформальную запись с многоточием, например, $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ — множество натуральных чисел.

Замечание 1.1.1. Мы будем считать, что 0 является натуральным числом.

В такой записи с многоточием мы предполагаем, что читатель понимает, какие именно элементы имеются в виду. Многоточие может стоять и справа, и слева: например, запись $\{\dots, -4, -2, 0, 2, 4, \dots\}$ призвана обозначать множество четных чисел.

Мы предполагаем также, что нам известны такие множества, изучающиеся в школе, как множество вещественных чисел $\mathbb{R}$, множество рациональных чисел $\mathbb{Q}$, множество целых чисел $\mathbb{Z}$.

Очень важный пример множества — пустое множество $\varnothing$. Это такое множество, что высказывание $x \in \varnothing$ ложно для любого объекта x.

Чуть более строгий способ задания множества: $A = \{s \in S \mid s \text{ удовлетворяет свойству } P\}$; здесь вертикальная черта $\mid$ читается как «таких, что», а P — то, что в математической логике называется *предикатом*, то есть, высказыванием, которое может для каждого объекта s быть истинным или ложным. Для записи предикатов (и вообще высказываний) полезны значки $\forall$ («для любого»), $\exists$ («существует») и $\exists!$ («существует единственный»). Эти значки называются *кванторами* и также имеют строгий смысл, но для нас они будут служить просто сокращениями интуитивно понятных фраз «для любого», «существует» и «существует единственный». Например, $\forall x \in \mathbb{N}, x > -5$ и $\exists! x \in \mathbb{N}, 3x = 15$ — истинные высказывания, а $\forall x \in \mathbb{N}, x < 20$ — ложное.

Теперь мы можем более точным образом описать множество всех четных чисел: $\{x \in \mathbb{Z} \mid \exists y \in \mathbb{Z} : x = 2y\}$. Еще одно полезное сокращение позволяет записать множество четных чисел так: $\{2x \mid x \in \mathbb{Z}\}$. Множество четных чисел мы будем обозначать через $2\mathbb{Z}$.

Обратите внимание, что порядок, в котором идут кванторы в высказывании, чрезвычайно важен: высказывание $\forall x \in \mathbb{Z} \exists y \in \mathbb{Z} : x = y + 1$, очевидно, истинно (из любого целого числа можно вычесть 1). А вот высказывание $\exists y \in \mathbb{Z} \forall x \in \mathbb{Z} : x = y + 1$ означает существование такого загадочного целого числа y, которое на единицу меньше любого целого числа. Понятно, что это высказывание ложно.

На самом деле, запись $\{s \in S \mid s \text{ удовлетворяет свойству } P\}$ задает не просто множество, а *подмножество* множества S. Если множество T таково, что любой элемент множества T является и элементом множества S, то говорят, что T является подмножеством S и пишут $T \subseteq S$. Более строго, $T \subseteq S$ тогда и только тогда, когда из $x \in T$ следует $x \in S$. Конструкцию «из ... следует ...» можно записывать значком $\Rightarrow$; в определении подмножества тогда можно писать $x \in T \Rightarrow x \in S$. Заметим, что стрелочка идет только в одну сторону; если бы было верно и $x \in S \Rightarrow x \in T$, то множества S и T совпадали бы. Таким образом, если $T \subseteq S$ и $S \subseteq T$, то $S = T$, поскольку в этом случае $x \in S \Leftrightarrow X \in T$; множества S и T состоят из одних и тех же элементов.

Примеры: $\mathbb{N} \subseteq \mathbb{Z} \subseteq \mathbb{Q} \subseteq \mathbb{R}$. Кроме того, $2\mathbb{Z} \subseteq \mathbb{Z}$. Более того, $\varnothing \subseteq X$ для любого множества X: пустое множество является подмножеством любого множества. В частности, $\varnothing \subseteq \varnothing$. Не следует путать значки $\subseteq$ и $\in$: так, $\varnothing \notin \varnothing$. К тому же, слева от значка $\in$ может стоять объект любой природы, а слева от значка $\subseteq$ — только множество.

Следующее важное понятие — *мощность* множества. Неформально говоря, это количество элементов в множестве. Мощность множества X обозначается через $|X|$. Четко различаются два случая: когда мощность множества конечна и когда она бесконечна. Если мощность множества конечна, то она измеряется натуральным числом (вообще говоря, это практически является определением натурального числа). Например, $|\varnothing| = 0$, $|\{1, 2, 3\}| = |\{2, 1, 3, 2, 2, 1\}| = 3$. Когда мощность множества X не является натуральным числом, говорят, что X бесконечно: $|X| = \infty$. Если множество X конечно, то любое его подмножество Y также конечно, и $|Y| \leqslant |X|$. Более того, если Y — подмножество конечного множества X, то $|Y| = |X|$ тогда и только тогда, когда $Y = X$. Если же $Y \subseteq X$ и $Y \neq X$ (в этом случае говорят, что Y — *собственное подмножество* X), то $|Y| < |X|$.

1.2 Операции над множествами

Литература: [K1], гл. 1, § 5, п. 1; [vdW], гл. 1, § 1.

Операции над множествами предоставляют массу способов получать новые множества из уже имеющихся. Мы обсудим по крайней мере следующие операции:

- объединение $\cup$,
- пересечение $\cap$,
- разность $\setminus$,
- симметрическая разность Δ,

- (декартово) произведение $\times$,
- несвязное объединение (копроизведение) $\coprod$,
- факторизация $/$.

Пересечение $A \cap B$ множеств A и B состоит из всех элементов, лежащих и в A, и в B. Более формально, $x \in A \cap B$ тогда и только тогда, когда $x \in A$ и $x \in B$.

Объединение $A \cup B$ множеств A и B состоит из всех элементов, лежащих в A или в B (возможно, и в A, и в B). Иначе говоря, $x \in A \cup B$ тогда и только тогда, когда $x \in A$ или $x \in B$.

Разность $A\backslash B$ состоит из элементов A, не лежащих в B: $A\backslash B = \{x \in A \mid x \notin B\}$. Иначе говоря, $x \in A\backslash B$ тогда и только тогда, когда $x \in A$ и $x \notin B$.

Симметрическая разность A и B состоит из элементов, лежащих ровно в одном из этих множеств. Это можно записать, например, так: $A\Delta B = (A \cup B)\backslash(A \cap B)$.

Несвязное объединение $A \coprod B$ предназначено для того, чтобы объединить два множества A и B (которые, возможно, имеют непустое пересечение) так, чтобы в результате элементы из A и из B «не перемешались»: все элементы из A оказались отличными от всех элементов из B. Представьте, что элементы множества A выкрашены в красный цвет, а элементы B — в синий цвет. После этого они стали все различны (их пересечение стало пустым), и мы рассмотрели их объединение. Если множества A и B конечны, то $|A \coprod B| = |A| + |B|$.

Произведение множества A и B — это множество всех упорядоченных пар (a, b), где $a \in A$, $b \in B$. Запись (a, b) означает, что мы заботимся о порядке элементов a, b (в отличие от записи $\{a, b\}$): пара (a, b), вообще говоря, не равна паре (b, a), если $a \neq b$. Более строго, $(a, b) = (a', b')$ тогда и только тогда, когда $a = a'$ и $b = b'$.

Итак, $A \times B = \{(a, b) \mid a \in A, b \in B\}$. Например,

$$\{1, 2, 3\} \times \{x, y\} = \{(1, x), (2, x), (3, x), (1, y), (2, y), (3, y)\}.$$

В школе изучают декартову плоскость, которая фактически представляет собой квадрат вещественной прямой: $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$. Заметим, что $|A \times B| = |A| \times |B|$ для конечных множеств A, B.

Несложно обобщить понятия пересечения и объединения на несколько множеств: $A_1 \cap A_2 \cap \cdots \cap A_n$, $A_1 \cup A_2 \cup \cdots \cup A_n$. Например, $A_1 \cap A_2 \cap A_3 \cap A_4 = ((A_1 \cap A_2) \cap A_3) \cap A_4$; и на самом деле порядок расстановки скобок в таком выражении не имеет значения. Более интересно попробовать обобщить понятие произведения; заметим, что $A_1 \times (A_2 \times A_3)$ не равно $(A_1 \times A_2) \times A_3$. Действительно, первое множество состоит из упорядоченных пар, первый элемент которых лежит в A_1, а второй является упорядоченной парой элементов из A_2 и A_3. В то же время второе множество состоит из совершенно других упорядоченных пар: первый их элемент является упорядоченной парой элементов из A_1 и A_2, а второй элемент лежит в множестве A_3. Но по аналогии с упорядоченной парой можно определить *упорядоченную тройку* и получить множество $A_1 \times A_2 \times A_3 = \{(a_1, a_2, a_3) \mid a_1 \in A_1, a_2 \in A_2, a_3 \in A_3\}$ (не совпадающее ни с $A_1 \times (A_2 \times A_3)$, ни с $(A_1 \times A_2) \times A_3$!). Совершенно аналогично определяется

упорядоченная n*-ка* или *кортеж* из n элементов $(a_1, \ldots, a_n)$, что позволяет определить произведение $A_1 \times A_2 \times \cdots \times A_n$.

Несложно определить пересечение и объединение для произвольного (не обязательно конечного) набора множеств: если $(A_i)_{i \in I}$ — семейство множеств, проиндексированное некоторым индексным множеством I, то $\bigcap_{i \in I} A_i$ — пересечение множеств A_i — состоит из элементов, которые лежат в каждом A_i, а $\bigcup_{i \in I} A_i$ — объединение множеств A_i — состоит из элементов, которые лежат хотя бы в одном из A_i.

С помощью упорядоченных пар мы можем более строго определить несвязное объединение множеств A и B: рассмотрим множества $\{0\} \times A$ и $\{1\} \times B$ (состоящие из «покрашенных элементов» $(0, a)$ для $a \in A$ и $(1, b)$ для $b \in B$). Теперь все элементы $(0, a)$ и $(1, b)$ уж точно различны, и можно положить $A \coprod B = (\{0\} \times A) \cup (\{1\} \times B)$.

1.3 Отображения

Литература: [K1], гл. 1, § 5, п. 2, [vdW], гл. 1, § 2.

Наивное определение: **отображение** $f\colon X \to Y$ сопоставляет каждому элементу $x \in X$ некоторый элемент $y \in Y$. При этом пишут $y = f(x)$ или $x \mapsto y$ и y называют **образом** элемента x при отображении f. Вместе с каждым отображением нужно помнить его **область определения** X и **область значений** Y; например, отображения $\mathbb{N} \to \mathbb{N}$, $x \mapsto x^2$ и $\mathbb{R} \to \mathbb{R}$, $x \mapsto x^2$ — два совершенно разных отображения.

Для каждого множества X можно рассмотреть **тождественное отображение** $\mathrm{id}_X\colon X \to X$, переводящее каждый элемент $x \in X$ в x.

С каждым декартовым произведением $X \times Y$ множеств X и Y связаны отображения $\pi_1\colon X \times Y \to X$ и $\pi_2\colon X \times Y \to Y$, определенные следующим образом: отображение π_1 сопоставляет паре (x, y) элементов $x \in X$, $y \in Y$ элемент x, а отображение π_2 сопоставляет этой паре элемент y. Эти отображения называются **каноническими проекциями**.

Пусть $f\colon X \to Y$ — отображение, и $A \subseteq X$; **образом** подмножества A называется множество образов всех элементов из A: $f(A) = \{y \in Y \mid \exists x \in A\colon f(x) = y\} = \{f(x) \mid x \in A\}$. Если же $B \subseteq Y$, можно посмотреть на все элементы X, образы которых лежат в B. Получаем **(полный) прообраз** подмножества B: $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$. Вообще, говорят, что x является прообразом элемента $y \in Y$, если $f(x) = y$; таким образом, полный прообраз подмножества составлен из всех прообразов всех его элементов.

Если $f\colon X \to Y$ — отображение множеств и $A \subseteq X$, можно определить **ограничение** отображения f на A. Это отображение, которое мы будем обозначать через $f|_A$, из A в Y, задаваемое, неформально говоря, тем же правилом, что и f. Более точно, $f|_A(x) = f(x)$ для всех $x \in A$.

Пусть теперь даны два отображения, $f\colon X \to Y$, $g\colon Y \to Z$. Их **композиция** $g \circ f$ — это новое отображение из X в Z, переводящее элемент $x \in X$ в $g(f(x)) \in Z$. То есть, $(g \circ f)(x) = g(f(x))$ для всех $x \in X$. Обратите внимание, что мы записываем композицию справа налево: в записи $g \circ f$ сначала применяется f, а потом g.

Теорема 1.3.1 (Ассоциативность композиции). *Пусть* X, Y, Z, T *— множества,* $f\colon X \to Y$*,* $g\colon Y \to Z$*,* $h\colon Z \to T$*. Тогда отображения* $(h \circ g) \circ f$ *и* $h \circ (g \circ f)$ *из* X *в* T *совпадают.*

Доказательство. Что значит, что два отображения совпадают? Во-первых, должны совпадать их области определения и значений; и действительно, $(h \circ g) \circ f$ и $h \circ (g \circ f)$ действуют из X в T. Во-вторых, они должны совпадать в каждой точке. Возьмем любой элемент $x \in X$ и проверим, что $((h \circ g) \circ f)(x) = (h \circ (g \circ f))(x)$. Действительно,

$$((h \circ g) \circ f)(x) = (h \circ g)(f(x)) = h(g(f(x)))$$

и

$$(h \circ (g \circ f))(x) = h((g \circ f)(x)) = h(g(f(x))).$$

□

Еще одно полезное свойство композиции: пусть $f\colon X \to Y$ — отображение. Тогда $f \circ \mathrm{id}_X = \mathrm{id}_Y \circ f = f$. Действительно, $(f \circ \mathrm{id}_X)(x) = f(\mathrm{id}_X(x)) = f(x)$ и $(\mathrm{id}_Y \circ f)(x) = \mathrm{id}_Y(f(x)) = f(x)$.

Все отображения из множества X в множество Y образуют множество, которое мы будем обозначать через $\mathrm{Map}(X, Y)$ или через Y^X. Последнее обозначение связано с тем, что для конечных X, Y имеет место равенство $|Y^X| = |Y|^{|X|}$. В частности, если $X = \varnothing$, то существует ровно одно отображение из X в Y: $|Y^{\varnothing}| = 1$. Если же, наоборот, $Y = \varnothing$, то для непустого X отображений из X в $\varnothing$ вообще нет: точке из X нечего сопоставить. Таким образом, $\varnothing^X = \varnothing$ для непустого X. Наконец, для пустого Y, как и для любого другого, существует ровно одно отображение из $\varnothing$ в Y (тождественное), поэтому $|\varnothing^{\varnothing}| = 1$.

Определение 1.3.2. Пусть $f\colon X \to Y$ — отображение.

1. f называется **инъективным отображением**, или **инъекцией**, если из $x_1 \neq x_2$ следует, что $f(x_1) \neq f(x_2)$ для $x_1, x_2 \in X$. Иными словами, у каждого элемента Y не более одного прообраза.

2. f называется **сюръективным отображением**, или **сюръекцией**, если для каждого $y \in Y$ найдется $x \in X$ такой, что $f(x) = y$. Иными словами, у каждого элеента Y не менее одного прообраза.

3. f называется **биективным отображением**, или **биекцией**, если оно инъективно и сюръективно.

Пример 1.3.3. Обозначим через $\mathbb{R}_{\geqslant 0}$ множество неотрицательных вещественных чисел: $\mathbb{R}_{\geqslant 0} = \{x \in \mathbb{R} \mid x \geqslant 0\}$. Рассмотрим четыре отображения

$$f_1\colon \mathbb{R} \to \mathbb{R}, x \mapsto x^2;$$
$$f_2\colon \mathbb{R} \to \mathbb{R}_{\geqslant 0}, x \mapsto x^2;$$
$$f_3\colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}, x \mapsto x^2;$$
$$f_4\colon \mathbb{R}_{\geqslant 0} \to \mathbb{R}_{\geqslant 0}, x \mapsto x^2.$$

Хотя эти отображения задаются одной и той же формулой (возведение в квадрат), их свойства совершенно различны: f_4 биективно; f_3 инъективно, но не сюръективно; f_2 сюръективно, но не инъективно; f_1 не инъективно и не сюръективно.

Определение 1.3.4. Пусть $f\colon X \to Y$ — отображение. Отображение $g\colon Y \to X$ называется **левым обратным** к f, если $g \circ f = \mathrm{id}_X$. Отображение $g\colon Y \to X$ называется **правым обратным** к f, если $f \circ g = \mathrm{id}_Y$. Наконец, g называется **[двусторонним] обратным** к f, если оно одновременно является левым обратным и правым обратным к f. Отображение f называется **обратимым слева**, если у него есть левое обратное, **обратимым справа**, если у него есть правое обратное, и просто **обратимым** (или **двусторонне обратимым**), если у него есть обратное.

Лемма 1.3.5. *Если у отображение $f\colon X \to Y$ есть левое обратное и правое обратное, то они совпадают. Таким образом, отображение обратимо тогда и только тогда, когда оно обратимо слева и обратимо справа.*

Доказательство. Пусть у f есть левое обратное g_L и правое обратное g_R. По определению это означает, что $g_L \circ f = \mathrm{id}_X$ и $f \circ g_R = \mathrm{id}_Y$. Рассмотрим отображение $(g_L \circ f) \circ g_R$. По теореме об ассоциативности композиции 1.3.1 оно равно $g_L \circ (f \circ g_R)$. С другой стороны, $(g_L \circ f) \circ g_R = \mathrm{id}_X \circ g_R = g_R$ и $g_L \circ (f \circ g_R) = g_L \circ \mathrm{id}_Y = g_L$. Поэтому $g_L = g_R$. □

Покажем, что мы на самом деле уже встречали понятия левой, правой и двусторонней обратимости под другими названиями.

Теорема 1.3.6. *Пусть $f\colon X \to Y$ — отображение.*

1. *Пусть X непусто. f обратимо слева тогда и только тогда, когда f инъективно.*
2. *f обратимо справа тогда и только тогда, когда f сюръективно.*
3. *f обратимо тогда и только тогда, когда f биективно.*

Доказательство. 1. Предположим, что f обратимо слева, то есть, $g \circ f = \mathrm{id}_X$ для некоторого $g\colon Y \to X$. Покажем инъективность f: пусть $x_1, x_2 \in X$ таковы, что $f(x_1) = f(x_2)$. Применяя g, получаем, что $g(f(x_1)) = g(f(x_2))$. Но $g(f(x)) = (g \circ f)(x) = \mathrm{id}_X(x) = x$ для всех $x \in X$, поэтому $x_1 = x_2$.

Обратно, предположим, что f инъективно, построим к f левое обратное отображение $g\colon Y \to X$. В силу непустоты X можно выбрать некоторый элемент $c \in X$. Для определения отображения g нам нужно задать его значение для каждого $y \in Y$. Возьмем $y \in Y$; в силу инъективности найдется не более одного элемента $x \in X$ такого, что $f(x) = y$. Если такой элемент (ровно один) есть, положим $g(y) = x$. Если же его нет, положим $g(y) = c$. Проверим, что так определенное отображение g действительно является левым обратным к f. Действительно, для всякого $x_0 \in X$ элемент $f(x_0)$ лежит в Y, и есть ровно один элемент $x \in X$ такой, что $f(x) = f(x_0)$ — это сам x_0. Поэтому в силу нашего определения $g(f(x_0)) = x_0 = \mathrm{id}_X(x_0)$. Мы получили, что для произвольного $x_0 \in X$ справедливо $(g \circ f)(x_0) = \mathrm{id}_X(x_0)$. Поэтому $g \circ f = \mathrm{id}_X$.

2. Предположим, что f обратимо справа, то есть, $f \circ g = \mathrm{id}_Y$ для некоторого $g\colon Y \to X$. Покажем сюръективность f; нужно проверить, что для каждого $y \in Y$ найдется элемент

$x \in X$ такой, что $f(x) = y$. Действительно, положим $x = g(y)$. Тогда $f(x) = f(g(y)) = (f \circ g)(y) = \mathrm{id}_Y(y) = y$.

Обратно, предположим, что f сюръективно. Построим отображение $g\colon Y \to X$ такое, что $f \circ g = \mathrm{id}_Y$. Для этого мы должны определить $g(y)$ для каждого $y \in Y$. В силу сюръективности найдется хотя бы один элемент $x \in X$ такой, что $f(x) = y$. Тогда положим $g(y) = x$. Очевидно, что $f(g(y)) = y$ для всех $y \in Y$.

Замечание 1.3.7. На самом деле тот факт, что мы можем *одновременно* для каждого $y \in Y$ выбрать один какой-нибудь элемент $x \in X$ со свойством $f(x) = y$, и получится корректно заданное отображение, является одной из аксиом теории множеств (она называется **аксиомой выбора**). Фактически, она равносильна как раз тому, что мы доказываем: обратимости справа любого сюръективного отображения. Заметим, что при доказательстве первого пункта теоремы такой проблемы не возникает: там при построении левого обратного отображения мы либо выбираем единственный прообраз, либо (в случае пустого прообраза) отправляем наш элемент в фиксированный элемент c. Здесь же прообраз может быть огромным, и возможность одновременно в огромном количестве прообразов выбрать по одному элементу как раз и гарантируется аксиомой выбора. Мы не обсуждаем строгую формализацию понятия множества, поэтому игнорируем все проблемы, связанные с аксиомой выбора.

3. Пусть f обратимо. Тогда, очевидно, оно обратимо слева и обратимо справа. По доказанному выше, из этого следует, что f инъективно и сюръективно (заметим, что в доказательстве того, что из обратимости слева следует инъективность, мы не использовали предположение о непустоте X). Значит, f биективно.

   Обратно, пусть f биективно, то есть, инъективно и сюръективно. Предположим сначала, что X непусто. Тогда, по доказанному выше, f обратимо слева и обратимо справа. По лемме 1.3.5 из этого следует, что f обратимо. Осталось рассмотреть случай, когда $X = \varnothing$. Покажем, что в этом случае и $Y = \varnothing$. Для этого вспомним, что f сюръективно. По определению это означает, что для каждого $y \in Y$ найдется $x \in X$ такой, что $f(x) = y$. Если Y непусто, то для какого-нибудь элемента $y \in Y$ должен найтись элемент $x \in X$, а это невозможно, поскольку X пусто. Мы показали, что $X = Y = \varnothing$; но в этом случае есть единственное отображение $f\colon X \to Y$ (тождественное), и единственное отображение $g\colon Y \to X$ будет обратным к нему.

□

Если $f\colon X \to Y$ — некоторое отображение, можно рассмотреть его **график**

$$\Gamma_f = \{(x, f(x)) \mid x \in X\} \subseteq X \times Y.$$

Это понятие помогает нам дать точное определение понятию отображения. Нетрудно видеть, что график отображения f однозначно определяет само f. С другой стороны, какие подмножества $X \times Y$ могут быть графиками отображений из X в Y? Нетрудно понять, что над каждой точкой $x \in X$ должна находиться ровно одна точка (x, y) из графика (у каждой точки x есть ровно один образ). Это приводит нас к следующему определению.

Определение 1.3.8. Упорядоченная тройка (X, Y, Γ), где X, Y — множества и $\Gamma \subseteq X \times Y$, называется **отображением** из X в Y, если

1. для любого $x \in X$ из того, что $(x, y_1) \in \Gamma$ и $(x, y_2) \in \Gamma$, следует, что $y_1 = y_2$;
2. для любого $x \in X$ существует $y \in Y$ такое, что $(x, y) \in \Gamma$.

1.4 Бинарные отношения

Литература: [K1], гл. 1, § 6, п. 1.

Определение 1.4.1. Бинарным отношением на множестве S называется подмножество $R \subseteq S \times S$. Если $(x, y) \in S$, говорят, что x **находится в отношении** R **с** y, и пишут xRy.

Примеры 1.4.2. Отношение $\geqslant$ на множестве $\mathbb{R}$: $\geqslant = \{(x, y) \in \mathbb{R} \times \mathbb{R} \mid x \geqslant y\}$. Аналогично — на множестве $\mathbb{Z}$, или на множестве $\mathbb{N}$. Отношения $\leqslant$, $>$, $<$ на тех же множествах. Отношение равенства на $\mathbb{R}$: $\{(x, x) \mid x \in \mathbb{R}\}$ — аналогично на любом множестве. Отношение делимости на целых числах (точное определение будет дано во второй главе). На множестве всех прямых на декартовой плоскости можно ввести отношение параллельности и отношение перпендикулярности.

Для визуализации отношений полезно рисовать их графики — изображать множества точек, координаты которых лежат в данном отношении.

1.5 Отношения эквивалентности

Литература: [K1], гл. 1, § 6, п. 2; [vdW], гл. 1, § 5.

Определение отношения достаточно общее; на практике встречаются отношения, удовлетворяющие некоторым из следующих свойств.

Определение 1.5.1. Пусть $R \subseteq X \times X$ — бинарное отношение на множестве X.

1. R называется **рефлексивным**, если для любого $x \in X$ выполнено xRx.
2. R называется **симметричным**, если для любых $x, y \in X$ из xRy следует yRx.
3. R называется **транзитивным**, если для любых $x, y, z \in X$ из xRy и yRz следует xRz.
4. R называется **отношением эквивалентности**, если оно рефлексивно, симметрично и транзитивно.

Примеры 1.5.2. Посмотрим на примеры 1.4.2. Нетрудно видеть, что отношения $\geqslant$, $\leqslant$, $>$, $<$ на множестве $\mathbb{R}$ транзитивны, но не симметричны. При этом отношения $\geqslant$ и $\leqslant$ рефлексивны. Отношение равенства на любом множестве является отношением эквивалентности. Отношение делимости рефлексивно и транзитивно. Отношение параллельности прямых на плоскости (если учесть, что прямая параллельна самой себе) является отношением эквивалентности. Отношение перпендикулярности симметрично, но не рефлексивно и не транзитивно.

Определение 1.5.3. Пусть $\sim$ — отношение эквивалентности на множестве X. Для элемента $x \in X$ рассмотрим множество $\{y \in X \mid y \sim x\}$. Мы будем обозначать его через $\overline{x}$ или $[x]$ и называть **классом эквивалентности** элемента x.

Теорема 1.5.4 (О разбиении на классы эквивалентности). *Пусть $\sim$ — отношение эквивалентности на множестве X. Тогда X разбивается на классы эквивалентности, то есть, каждый элемент множества X лежит в каком-то классе, и любые два класса либо не пересекаются, либо совпадают.*

Доказательство. Из рефлексивности следует, что $x \in \overline{x}$, поэтому каждый элемент лежит в каком-то классе. Пусть $\overline{x}$ и $\overline{y}$ — два класса эквивалентности и $\overline{x} \cap \overline{y} \neq \varnothing$. Выберем $z \in \overline{x} \cap \overline{y}$; тогда $z \sim x$ и $z \sim y$. Докажем, что на самом деле $\overline{x} = \overline{y}$, проверив включения в обе стороны. Возьмем $t \in \overline{x}$; тогда $t \sim x$, $x \sim z$, $z \sim y$, откуда $t \sim y$, то есть, $t \in \overline{y}$. Поэтому $\overline{x} \subseteq \overline{y}$. Аналогично, $\overline{y} \subseteq \overline{x}$. □

Определение 1.5.5. Пусть $\sim$ — отношение эквивалентности на множестве X. Множество всех классов эквивалентности элементов X называется **фактор-множеством** множества X по отношению $\sim$ и обозначается через $X/\sim$. Отображение $\pi\colon X \to X/\sim$, сопоставляющее каждому элементу $x \in X$ его класс эквивалентности $\overline{x}$, называется **канонической проекцией** множества X на фактор-множество $X/\sim$. Нетрудно видеть, что это отображение сюръективно.

1.6 Метод математической индукции

Литература: [K1], гл. 1, § 7; [vdW], гл. 1, § 3; [B], гл. 1, п. 2.

Пусть $P(n)$ — набор высказываний, зависящий от натурального параметра n. **Принцип математической индукции** гласит, что если $P(0)$ истинно (**база индукции**) и для любого натурального k из истинности $P(k)$ следует истинность $P(k+1)$ (**индукционный переход**), то $P(n)$ истинно для всех натуральных n.

Эквивалентная переформулировка принципа математической индукции гласит, что в любом непустом множестве натуральных чисел есть минимальный элемент. Этот принцип (или какой-то равносильный ему), как правило, принимается за аксиому в современных аксиоматиках натуральных чисел.

Покажем, что если в любом непустом множестве натуральных чисел есть минимальный элемент, то принцип математической индукции выполняется. Будем действовать от противного: предположим, что $P(0)$ истинно, и для любого $k \in \mathbb{N}$ из истинности $P(k)$ следует истинность $P(k+1)$, но, в то же время, $P(n)$ истинно не для всех n. Пусть $A \subseteq \mathbb{N}$ — множество натуральных чисел n, для которых $P(n)$ ложно; оно непусто по нашему предположению. Тогда в A есть минимальный элемент a. Если $a = 0$, то $P(0)$ ложно (поскольку $a \in A$), что противоречит базе индукции. Если же $a > 0$, то $a - 1$ также является натуральным числом, и $a - 1 \notin A$ в силу минимальности. Поэтому $P(a-1)$ истинно. Но тогда из индукционного перехода следует, что и $P(a) = P((a-1)+1)$ истинно — противоречие.

Принципа математической индукции равносилен следующему принципу полной индукции: пусть $P(n)$ — набор высказываний, зависящий от натурального параметра n. Если $P(0)$ истинно

и из истинности $P(0), P(1), \ldots, P(k)$ следует истинность $P(k+1)$, то $P(n)$ истинно для всех натуральных n.

1.7 Операции

Литература: [K1], гл. 4, § 1, п. 1.

Определение 1.7.1. Пусть X — множество. **Бинарной операцией** на множестве X называется отображение $X \times X \to X$.

Примеры 1.7.2. Отображения $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$, задаваемые формулами $(a, b) \mapsto a + b$, $(a, b) \mapsto ab$, $(a, b) \mapsto a - b$, являются бинарными операциями. Отображение $(a, b) \mapsto a^b$ является бинарной операцией на множестве $\mathbb{N}_{\geqslant 0}$ положительных натуральных чисел.

Определение 1.7.3. Пусть $\varphi\colon X \times X \to X$ — бинарная операция на множестве X.

1. Операция φ называется **ассоциативной**, если $\varphi(\varphi(a, b), c) = \varphi(a, \varphi(b, c))$ выполняется для всех $a, b, c \in X$.

2. Операция φ называется **коммутативной**, если $\varphi(a, b) = \varphi(b, a)$ выполняется для всех $a, b \in X$.

Нетрудно видеть, что операции сложения и умножения на множестве вещественных чисел являются ассоциативными и коммутативными, а вот возведение в степень положительных натуральных положительных чисел не является ни ассоциативной, ни коммутативной операцией.

Определение 1.7.4. Пусть $\bullet$ — бинарная операция на множестве X. Элемент $e \in X$ называется **левым нейтральным** (или **левой единицей**) по отношению к операции $\bullet$, если $e \bullet x = x$ для любого $x \in X$. Элемент $e \in X$ называется **правым нейтральным** (или **правой единицей**) по отношению к $\bullet$, если $x \bullet e = x$ для любого $x \in X$. Элемент $e \in X$ называется **нейтральным** (или **единицей**), если он одновременно является левым и правым нейтральным.

Отметим, что бинарная операция возведения в степень на множестве $\mathbb{R}$ обладает правой единицей (это 1: действительно, $a^1 = a$), но не обладает левой единицей.

Лемма 1.7.5. *Если $\bullet\colon X \times X \to X$ — бинарная операция, и в X есть правая единица и левая единица относительно $\bullet$, то они совпадают.*

Доказательство. Действительно, если $e_L \in X$ — левая единица, а $e_R \in X$ — правая единица, то по определению левой единицы выполнено $e_L \bullet e_R = e_R$, а по определению правой единицы выполнено $e_L \bullet e_R = e_L$. Поэтому $e_L = e_L \bullet e_R = e_R$. □

Определение 1.7.6. Пусть $\bullet$ — бинарная операция на множестве X, и в X есть нейтральный элемент e относительно этой операции. Пусть $x \in X$. Элемент $y \in X$ называется **левым обратным** (относительно операции $\bullet$) к x, если $yx = e$. Элемент $y \in X$ называется **правым обратным**

(относительно операции $\bullet$) к x, если $xy = e$. Если $y \in X$ одновременно является левым и правым обратным к x, то он называется просто **обратным** к x. Элемент x называется **обратимым слева**, если у него есть левый обратный, **обратимым справа**, если у него есть правый обратный, и **обратимым**, если у него есть обратный.

Лемма 1.7.7. *Пусть $\bullet$ — бинарная операция на множестве X, и в X есть нейтральный элемент e относительно это операции. Предположим, что операция $\bullet$ ассоциативна. Пусть элемент x обратим слева и обратим справа. Тогда он обратим. Иными словами, если у элемента есть левый и правый обратный относительно ассоциативной операции, то они совпадают.*

Доказательство. Пусть y_L — левый обратный к x, а y_R — правый обратный к x. По определению это означает, что $y_L \bullet x = e$ и $x \bullet y_R = e$. Но тогда

$$y_R = e \bullet y_R = (y_L \bullet x) \bullet y_R = y_L \bullet (x \bullet y_R) = y_L \bullet e = y_L$$

(обратите внимание, что в середине мы воспользовались ассоциативностью операции $\bullet$). □

Пусть на X задана бинарная операция $\bullet$, и $a, b, c \in X$. Выражение $a \bullet b \bullet c$ не определено: для его однозначной интерпретации необходимо расставить скобки, и получится либо $(a \bullet b) \bullet c$, либо $a \bullet (b \bullet c)$. Если операция $\bullet$ ассоциативна, то результат вычисления этих двух выражений одинаков. Пусть теперь $a, b, c, d \in X$. Скобки в выражении $a \bullet b \bullet c \bullet d$ можно расставить уже пятью вариантами:

$$((a \bullet b) \bullet c) \bullet d, \quad (a \bullet (b \bullet c)) \bullet d, \quad (a \bullet b) \bullet (c \bullet d), \quad a \bullet ((b \bullet c) \bullet d), \quad a \bullet (b \bullet (c \bullet d)).$$

Оказывается, что если операция $\bullet$ ассоциативна, то результат вычисления всех этих выражений одинаков. Аналогично, в выаржении любой длины для указания порядка, в котором выполняются операции, необходимо расставить скобки. Оказывается, для ассоциативной операции результат выполнения не зависит от порядка расстановки скобок. Это свойство называется **обобщенной ассоциативностью**. Поэтому для ассоциативных операций ставить скобки в подобных выражениях не обязательно.

Теорема 1.7.8. *Если на множестве X задана ассоциативная операция $\bullet$, то она обладает обобщенной ассоциативностью: результат вычисления выражения $x_1 \bullet x_2 \bullet \cdots \bullet x_n$ не зависит от расстановки в нем скобок.*

Доказательство. Будем доказывать индукцией по n. База $n = 3$ является определением ассоциативности. Пусть теперь $n > 3$, и для всех меньших n теорема уже доказана. Достаточно показать, что результат при любой расстановке скобок совпадает с результатом при следующей расстановке, в которой все скобки «сдвинуты влево»

$$(\ldots((x_1 \bullet x_2) \bullet x_3) \bullet \cdots \bullet x_n).$$

Возьмем произвольную расстановку и посмотрим на действие, которое выполняется последним: оно состоит в перемножении некоторого выражения от $x_1, \dots, x_k$ и некоторого выражения от $x_{k+1}, \dots, x_n$:

$$(\dots x_1 \bullet \dots \bullet x_k \dots) \bullet (\dots x_{k+1} \bullet \dots \bullet x_n \dots).$$

При этом $1 < k < n$.

Предположим сначала, что $k = n - 1$. Тогда последняя операция состоит в перемножении скобки, в которой стоят $x_1, \dots, x_{n-1}$, на x_n. В выражении от $x_1, \dots, x_{n-1}$ мы можем, по предположению индукции, сдвинуть все скобки влево, не меняя результата. Приписывая справа x_n, получаем как раз выражение нужного вида уже от $x_1, \dots, x_n$, и доказательство закончено.

Пусть теперь $k < n - 1$. Заметим, что во второй скобке стоят $x_{k+1}, \dots, x_n$ — здесь хотя бы два элемента, и меньше, чем n. По предположению индукции мы можем расставить в этом выражении скобки нашим выбранным способом, не меняя результата:

$$\underbrace{(\dots x_1 \bullet \dots \bullet x_k \dots)}_{A} \bullet (\underbrace{(\dots (x_{k+1} \bullet x_{k+2}) \bullet \dots \bullet x_{n-1})}_{B} \bullet \underbrace{x_n}_{C})$$

(тут нужно отметить, что рассуждение работает и при $k = n-2$; в этом случае во второй скобке стоит лишь два элемента, и формально мы не можем применить предположение индукции, но в этом нет ничего страшного). Применим теперь ассоциативность к полученному выражению вида $A \bullet (B \bullet C)$ и заменим его на $(A \bullet B) \bullet C$:

$$(\underbrace{\dots x_1 \bullet \dots \bullet x_k \dots}_{A} \bullet \underbrace{\dots (x_{k+1} \bullet x_{k+2}) \bullet \dots \bullet x_{n-1})}_{B} \bullet \underbrace{x_n}_{C})$$

Заметим, что теперь последняя выполняемая операция — умножения некоторого выражения от переменных $x_1, \dots, x_{n-1}$ на x_n. Это означает, что мы свели задачу к уже разобранному случаю $k = n - 1$; теперь можно, как и выше, воспользоваться предположением индукции, расставить скобки в выражении от $x_1, \dots, x_{n-1}$ нужным образом, и мы сразу получим необходимую расстановку. □

2 Элементарная теория чисел

В этой главе мы в основном работаем с множеством целых чисел $\mathbb{Z}$.

2.1 Делимость целых чисел

Литература: [F], гл. I, § 1, пп. 1, 2; [K1], гл. 1, § 9, п. 3; [V], гл. I, § 1; [B], гл. 1, п. 2.

Определение 2.1.1. Пусть x, y — целые числа. Говорят, что x **делит** y (или, что y **делится на** x) если существует такое целое число k, что $y = xk$. Обозначение: $x \mid y$.

Предложение 2.1.2. *Для любых целых x, y, z выполнено:*

1. $x \mid x$, $1 \mid x$, $(-x) \mid x$, $(-1) \mid x$;
2. *если* $x \mid y$ *и* $y \mid z$, *то* $x \mid z$ *(отношение делимости транзитивно);*
3. *если* $x \mid y$ *и* $x \mid z$, *то* $x \mid y + z$;
4. *если* $x \mid y$, *то* $x \mid yz$;
5. *если* $z \neq 0$, *то* $xz \mid yz$ *равносильно* $x \mid y$;
6. $x \mid 0$; *если* $0 \mid x$, *то* $x = 0$.

Доказательство. 1. $x = x \cdot 1 = 1 \cdot x = (-x) \cdot (-1) = (-1) \cdot (-x)$.

2. Если $y = xk$, $z = yl$, то $z = (xk)l = x(kl)$.
3. Если $y = xk$, $z = xl$, то $y + z = x(k + l)$.
4. Если $y = xk$, поэтому $yz = (xk)z = x(kz)$.
5. Если $y = xk$, то $yz = xzk$; обратно, если $yz = xzk$, то $(y - xk)z = 0$. Из $z \neq 0$ теперь следует, что $y - xk = 0$, то есть, $y = xk$.
6. $0 = x \cdot 0$; если $x = 0 \cdot k$, то $x = 0$.

□

Определение 2.1.3. Если $x \mid y$ и $y \mid x$, говорят, что числа x и y **ассоциированы**.

Замечание 2.1.4. Заметим, что это означает, что $y = xk$ и $x = yl$, откуда $x = xkl$. Если $x = 0$, то и $y = 0$; иначе $1 = kl$, поэтому $|k| = |l| = 1$ и либо $k = l = 1$, либо $k = l = -1$. Стало быть, $y = x$ или $y = -x$.

Теорема 2.1.5 (О делении с остатком). *Пусть $a, b \in \mathbb{Z}$, $b \neq 0$. Тогда существуют единственные целые числа q (неполное частное) и r (остаток) такие, что $a = bq + r$ и $0 \leqslant r \leqslant |b| - 1$.*

Доказательство. Предположим сначала, что $b > 0$ и $a \geqslant 0$. Доказываем индукцией по a. База: $a < b$. В этом случае $a = b \cdot 0 + a$ и $0 \leqslant a \leqslant b - 1$. Переход: пусть теперь $a \geqslant b$; посмотрим на число $a - b$, снова $a - b \geqslant 0$ и $a - b < a$, поэтому по предположению индукции найдутся q', r' такие, что $a - b = bq' + r'$ и $0 \leqslant r' \leqslant b - 1$. Но тогда $a = b(q' + 1) + r'$. Пусть теперь $a < 0$; но тогда $-a \geqslant 0$ и, по доказанному, найдутся q', r' такие, что $-a = bq' + r'$, $0 \leqslant r' \leqslant b - 1$. Из этого получаем, что $a = -bq' - r'$. Если $r' = 0$, то $a = b(-q') + 0$, и все доказано. Если же $1 \leqslant r' \leqslant b - 1$, то $a = b(-q') - b + b - r' = b(-q' - 1) + (b - r')$. Заметим, что $-b + 1 \leqslant -r' \leqslant -1$, поэтому $1 \leqslant b - r' \leqslant b - 1$, и все доказано.

Наконец, предположим, что $b < 0$; тогда $-b > 0$ и можно найти q', r' такие, что $a = (-b)q' + r'$ и $0 \leqslant r' \leqslant -b - 1$. Но тогда $a = b(-q') + r'$ и $0 \leqslant r' \leqslant |b| - 1$, что и требовалось.

Осталось доказать единственность. Пусть $a = bq + r = bq' + r'$; тогда $b(q - q') = (r' - r)$. Если $q = q'$, то и $r = r'$. Если же $q \neq q'$, то $|b| \cdot |q - q'| = |r - r'|$ и левая часть $\geqslant |b|$. С другой стороны, $0 \leqslant r, r' \leqslant |b| - 1$, поэтому правая часть не превосходит $|b| - 1$, противоречие. □

2.2 Наибольший общий делитель и алгорифм Эвклида

Литература: [F], гл. I, § 1, пп. 3, 4; [K1], гл. 1, § 9, п. 2; [V], гл. I, § 2; [B], гл. 3, пп. 1, 2.

Определение 2.2.1. Пусть $a, b \in \mathbb{Z}$. Говорят, что целое число d является **общим делителем** a и b, если $d \mid a$ и $d \mid b$.

Определение 2.2.2. Пусть $a, b \in \mathbb{Z}$. Целое число d называется **наибольшим общим делителем** (**НОД**) чисел a и b, если

- d — общий делитель a и b;
- если d' — общий делитель a и b, то $d' \mid d$.

Обозначение: $d = \gcd(a, b)$.

Заметим, что НОД двух целых чисел (если он существует) единственен с точностью до знака. А именно, если d и d' — два наибольших общих делителя чисел a и b, то из определения следует, что $d \mid d'$ и $d' \mid d$, откуда по замечанию 2.1.4 следует, что $d = \pm d'$. Поэтому важно понимать, что выражение $\gcd(a, b)$ не является однозначно определенным целым числом, а лишь обозначает *какой-нибудь* из наибольших общих делителей чисел a и b. Например, если $\gcd(a, b) = d$, то и $\gcd(a, b) = -d$.

Легко видеть, что $\gcd(0, a) = a$; в частности, $\gcd(0, 0) = 0$.

Некоторые авторы называют наибольшим общим делителем не произвольное целое, а *натуральное* число с этими свойствами. При этом наибольший общий делитель становится единственным: действительно, из пары целых чисел d и $-d$ всегда ровно одно является натуральным. Однако, такая точка зрения неудобна, поскольку при обобщении понятия наибольшего общего делителя на другие кольца (например, на кольцо многочленов — см. раздел 4.7) подобного рода единственность невозможно обеспечить.

Предложение 2.2.3. *Наибольший общий делитель двух целых чисел* a, b *существует и представляется в виде* $d = au_0 + bv_0$ *для некоторых целых* u_0, v_0.

Доказательство. Если $a = b = 0$, то мы уже знаем, что $\gcd(a, b) = 0$, и доказывать нечего. Теперь можно считать, что $a \neq 0$. Рассмотрим множество всех натуральных чисел вида $au + bv$ для всевозможных целых u, v и выберем в нем наименьший ненулевой элемент (это множество непусто: например, оно содержит $|a|$). Обозначим его через d; по построению имеем $d = au_0 + bv_0$ для некоторых целых u_0, v_0. Покажем, что d является общим делителем a и b. Поделим a на d с остатком: $a = dq + r = (au_0 + bv_0)q + r$, откуда $r = a(1 - u_0 q) + b(-v_0 q)$. Однако, $r < d$ – натуральное число, а d было наименьшим натуральным числом, представляемым в виде $d = ax + by$. Значит, $r = 0$ и a делится на d. Аналогично, b делится на d.

Докажем теперь, что d — это наибольший общий делитель a и b. Пусть d' — какой-то общий делитель a и b: $d' \mid a$ и $d' \mid b$. Тогда по свойствам делимости $d' \mid au_0$, $d' \mid bv_0$, и $d' \mid au_0 + bv_0 = d$, что и требовалось. □

Выражение $d = au_0 + bv_0$ из предложения 2.2.3 называется **линейным представлением НОД**.

Практический способ для нахождения наибольшего общего делителя — алгорифм Эвклида.

Пусть $a, b \in \mathbb{Z}$. Наша цель — найти $\gcd(a, b)$. Заметим сразу, что $\gcd(a, b) = \gcd(|a|, |b|)$, поэтому можно считать, что $a, b \in \mathbb{N}$. Если одно из чисел a, b равно 0, цель достигнута. Иначе пусть для определенности $a \geqslant b > 0$. Делим с остатком a на b: $a = bq_0 + r_0$. Посмотрим на пару (b, r_0) и применим ту же операцию к ней (теперь мы знаем, что $b > r_0$): $b = r_0 q_1 + r_1$ и так далее: $r_0 = r_1 q_2 + r_2 \ldots$ Заметим, что максимальное число в паре всегда уменьшается; значит, процесс когда-то остановится (остаток станет равен нулю). Мы утверждаем, что последний ненулевой остаток в этой цепочке равен $\gcd(a, b)$. Для доказательства этого факта нам понадобится следующая лемма.

Лемма 2.2.4. *Пусть* $a, b, q, r \in \mathbb{Z}$. *Если* $a = bq + r$, *то* $\gcd(a, b) = \gcd(b, r)$.

Доказательство. Действительно, пусть $d = \gcd(a, b)$ и $d' = \gcd(b, r)$. С одной стороны, $d \mid a$, $d \mid b$, откуда $d \mid (a - bq) = r$, и из определения $d' = \gcd(b, r)$ следует, что $d \mid d'$. Кроме того, $d' \mid b$, $d' \mid r$, откуда $d' \mid bq + r = a$, и из определения $d = \gcd(a, b)$ следует, что $d' \mid d$. Мы получили, что $d \mid d'$ и $d' \mid d$; это означает, что $d = \pm d'$, и потому $\gcd(a, b) = \gcd(b, r)$. □

Поэтому наибольший общий делитель пары, с которой мы работаем в алгорифме Эвклида, не меняется; и как только в паре появился 0, другое число в паре должно быть равно $\gcd(a, b)$.

Более того, алгорифм Эвклида позволяет находить и линейное представление НОД. Действительно, в конце алгорифма мы приходим к паре $(d, 0)$ и линейное представление очевидно: $d = d \cdot 1 + 0 \cdot 0$. На каждом шаге мы переходим от пары (a, b) к паре (b, r), где $a = bq + r$; если мы уже знаем, что $d = bx' + ry'$, то, подставляя $r = a - bq$, имеем $d = bx' + (a - bq)y' = ay' + b(x' - qy')$.

2.3 Свойства НОД и взаимная простота

Литература: [F], гл. I, § 1, п. 5; [V], гл. I, § 2; [B], гл. 3, пп. 1, 3.

Предложение 2.3.1 (Свойства НОД). *1.* $\gcd(x, y) = x$ *тогда и только тогда, когда* $x \mid y$.

2. $\gcd(\gcd(x, y), z) = \gcd(x, \gcd(y, z))$.

3. $\gcd(zx, zy) = z \cdot \gcd(x, y)$.

Доказательство. 1. Если $\gcd(x, y) = x$, то $x \mid y$ по определению. Обратно, пусть $x \mid y$, тогда x — общий делитель x и y, и если d' — какой-то общий делитель x, y, то, в частности, $d' \mid x$. Значит, $\gcd(x, y) = x$.

2. Любой общий делитель $\gcd(x, y)$ и z является общим делителем x, y и z; то же можно сказать про любой общий делитель x и $\gcd(y, z)$. Позже мы распространим определение gcd на несколько элементов и увидим, что и левая, и правая части необходимого равенства равны $\gcd(x, y, z)$.

3. Если $z = 0$, то и слева, и справа стоит 0; доказывать нечего. Пусть $\gcd(x, y) = d$; $d \mid x$, $d \mid y$, откуда $zd \mid zx$ и $zd \mid zy$; поэтому $zd \mid \gcd(zx, zy)$. Обратно, очевидно, что $z \mid zx$, $z \mid zy$, поэтому $z \mid \gcd(zx, zy)$. Запишем $\gcd(zx, zy) = zc$ для некоторого c. Значит, $zc \mid zx$, $zc \mid zy$, откуда после сокращения (с учетом того, что $z \neq 0$) получаем $c \mid x$ и $c \mid y$. Поэтому $c \mid \gcd(x, y) = d$, откуда $zc \mid zd$, то есть, $\gcd(zx, zy) \mid zd$.

□

Определение 2.3.2. Числа a, b называются **взаимно простыми**, если $\gcd(a, b) = 1$. Обозначение: $a \perp b$.

Предложение 2.3.3 (Свойства взаимной простоты). *Пусть* a, b, c *— некоторые целые числа.*

1. Если $a \perp b$ *и* $a \perp c$*, то* $a \perp bc$.

2. $a \perp b$ *тогда и только тогда, когда существуют целые числа* u_0*,* v_0 *такие, что* $au_0 + bv_0 = 1$.

3. Если $c \mid ab$ *и* $a \perp c$*, то* $c \mid b$.

4. Если $b_1 \mid a$*,* $b_2 \mid a$ *и* $b_1 \perp b_2$*, то* $b_1 b_2 \mid a$.

Доказательство. 1.

$$\begin{aligned}\gcd(a, bc) &= \gcd(\gcd(a, ac), bc) \\ &= \gcd(a, \gcd(ac, bc)) \\ &= \gcd(a, c\gcd(a, b)) \\ &= \gcd(a, c) \\ &= 1.\end{aligned}$$

2. если $a \perp b$, то $1 = au_0 + bv_0$ — линейное представление НОД. Обратно, если $au_0 + bv_0 = 1$ и $d = \gcd(a, b)$, то $d \mid au_0$, $d \mid bv_0$, откуда $d \mid au_0 + bv_0 = 1$ и $d = 1$.

3. Запишем $au_0 + cv_0 = 1$ и умножим на b: $abu_0 + cbv_0 = b$. Мы знаем, что $c \mid ab$, поэтому $c \mid abu_0$. Кроме того, очевидно, что $c \mid cbv_0$. Поэтому c делит и их сумму $abu_0 + cbv_0 = b$.

4. $a = b_1k$ делится на b_2, $b_1 \perp b_2$, по предыдущему свойству k делится на b_2: $k = b_2l$, откуда $a = b_1k = b_1b_2l$.

□

2.4 Линейные диофантовы уравнения

Литература: [В], гл. 14, п. 2.

Пусть $a, b, c \in \mathbb{Z}$. Нас интересуют решения (x, y) уравнения $ax + by = c$. Если $a = b = 0$, то при $c = 0$ решение любое, а при $c \neq 0$ решений нет.

Если $b = 0$, $a \neq 0$, получаем уравнение $ax = c$. Если $a \mid c$, то $x = c/a$, y — любое; иначе решений нет.

Обозначим $d = \gcd(a, b)$. Заметим, что $d \mid a$, $d \mid b$, поэтому d должно делить выражение $ax + by$ при всех x, y. Значит, если d не делит c, то решений нет.

Пусть теперь $d \mid c$. Запишем $a = da'$, $b = db'$, $c = dc'$; тогда обе части нашего уравнения можно поделить на d и прийти к эквивалентному уравнению $a'x + b'y = c'$, для которого уже $\gcd(a', b') = 1$ (поскольку $d = \gcd(a, b) = \gcd(da', db') = d\gcd(a', b')$).

Поэтому теперь можно считать, что $\gcd(a, b) = 1$. Мы знаем, что есть линейное представление НОД: $au_0 + bv_0 = 1$. Умножая на c обе части, получаем, что $a(u_0c) + b(v_0c) = c$. Обозначим $x_0 = u_0c$, $y_0 = v_0c$. Мы получили, что у нашего уравнения есть решение (x_0, y_0). Как найти все решения?

Пусть (x, y) — какое-то решение уравнения $ax + by = c$. Вычитая $ax_0 + by_0 = c$ из этого равенства, получаем $a(x - x_0) + b(y - y_0) = 0$, откуда $a(x - x_0) = b(y_0 - y)$. Стало быть, $b \mid a(x - x_0)$; но $a \perp b$, поэтому $b \mid x - x_0$. Запишем $x - x_0 = bt$; тогда $abt = b(y_0 - y)$, откуда $y_0 - y = at$. Получили, что произвольное решение (x, y) нашего уравнения выглядит так: $x = x_0 + bt$, $y = y_0 - at$. Итак, если (x_0, y_0) — какое-то одно решение уравнения $ax + by = c$, то все его решения имеют вид $(x_0 + bt, y_0 - at)$ для $t \in \mathbb{Z}$. Обратно, прямая подстановка показывает, что $(x_0 + bt, y_0 - at)$ действительно является решением нашего уравнения.

Теперь посмотрим на случай нескольких переменных. Для этого нам понадобится расширить понятие НОД на случай нескольких чисел.

Определение 2.4.1. Пусть $a_1, \ldots, a_n \in \mathbb{Z}$. Натуральное число d называется **наибольшим общим делителем** чисел $a_1, \ldots, a_n$, если выполняются следующие условия:

1. d — общий делитель $a_1, \ldots, a_n$ (то есть, d делит каждое a_i);

2. если d' — общий делитель $a_1, \ldots, a_n$, то $d' \mid d$.

Обозначение: $d = \gcd(a_1, \ldots, a_n)$.

Упражнение 2.4.2. *Докажите следующие свойства НОД:*

1. $\gcd(a_1, \ldots, a_n) = \gcd(\gcd(a_1, a_2), a_3, \ldots, a_n)$*;*

2. gcd *не зависит от порядка аргументов;*

3. $\gcd(za_1, za_2, \ldots, za_n) = |z| \gcd(a_1, \ldots, a_n)$.

Из этого упражнения, в частности, следует, что НОД нескольких чисел существует и единственен.

Теорема 2.4.3 (Критерий разрешимости линейного диофантова уравнения от нескольких переменных). *Пусть* $a_1, \ldots, a_n, c \in \mathbb{Z}$. *Линейное уравнение*

$$a_1x_1 + \cdots + a_nx_n = c$$

разрешимо в целых числах тогда и только тогда, когда $\gcd(a_1, \ldots, a_n)$ *делит* c.

Доказательство. Очевидно, что если это уравнение разрешимо, то каждое слагаемое в левой части делится на $\gcd(a_1, \ldots, a_n)$, поэтому и c на него делится. Докажем теперь, что если c делится на $d = \gcd(a_1, \ldots, a_n)$, то уравнение разрешимо.

Из нашего анализа линейного диофантова уравнения от двух переменных следует, что этот критерий верен для $n = 2$. Это будет базой для индукции по n. Пусть теперь $n \geqslant 3$. Рассмотрим следующее уравнение:

$$\gcd(a_1, a_2)y_1 + a_3y_3 + \cdots + a_ny_n = c.$$

Это линейное диофантово уравнение от $n - 1$ неизвестных $y_1, y_3, \ldots, y_n$. По предположению индукции оно разрешимо тогда и только тогда, когда его правая часть, c, делится на $\gcd(\gcd(a_1, a_2), a_3, \ldots, a_n) = \gcd(a_1, a_2, a_3, \ldots, a_n) = d$. У нас по условию $d \mid c$, поэтому новое уравнение имеет решение $(y_1, y_3, \ldots, y_n)$. Построим теперь решение нашего первоначального уравнения. Посмотрим на еще одно вспомогательное уравнение

$$a_1x_1 + a_2x_2 = \gcd(a_1, a_2)y_1$$

с неизвестными x_1, x_2. Правая часть делится на НОД его коэффициентов, поэтому оно разрешимо. Итак, мы нашли x_1, x_2; положим теперь $x_3 = y_3, \ldots, x_n = y_n$. Тогда

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n &= \gcd(a_1, a_2)y_1 + a_3x_3 + \cdots + a_nx_n \\ &= \gcd(a_1, a_2)y_1 + a_3y_3 + \cdots + a_ny_n \\ &= c, \end{aligned}$$

поэтому $(x_1, \ldots, x_n)$ — решение исходного уравнения.

□

2.5 Основная теорема арифметики

Литература: [F], гл. I, § 1, п. 6; [K1], гл. 1, § 9, п. 1; [V], гл. I, § 5, § 6; [B], гл. 2, п. 1.

Определение 2.5.1. Натуральное число p, отличное от 0 и 1, называется **простым**, если из того, что $p = xy$ для некоторых целых x, y, следует, что x ассоциировано с p или y ассоциировано с p.

При этом, если x ассоциировано с p, то y ассоциировано с 1; если же y ассоциировано с p, то x ассоциировано с 1. Альтернативное определение: натуральное число $p > 1$ называется простым, если у него нет натуральных делителей, кроме 1 и p.

Предложение 2.5.2 (Свойства простых чисел). *Пусть p — простое число.*

1. *если n — целое число, и p не делит n, то p и n взаимно просты;*
2. *пусть $a, b \in \mathbb{Z}$; если p делит ab, то p делит a или p делит b;*
3. *если p делит произведение нескольких целых чисел, то p делит хотя бы одно из них;*
4. *всякое целое число, большее 1, делится по крайней мере на одно простое;*
5. *простых чисел бесконечно много;*
6. *если p_1 и p_2 — два различных простых числа, то они взаимно просты.*

Доказательство. 1. Предположим, что p не делит n, и пусть $d = \gcd(n, p)$. При этом $d \mid p$, поэтому d либо ассоциировано с p, либо ассоциировано с 1. Заметим, что d также делит n, поэтому если d ассоциировано с p, то p делит n — противоречие. Значит, d ассоциировано с 1, откуда $n \perp p$.

2. Пусть p делит ab, но не делит a. По предыдущему свойству $a \perp p$, и по свойству взаимно простых чисел получаем, что $p \mid b$.

3. Индукция по n; база — пункт (2). $p \mid (a_1a_2)a_3 \dots a_n$, поэтому либо a_1a_2, либо какое-то из a_i (при $i > 2$) делится на p; если a_1a_2 делится на p, то либо a_1, либо a_2 делится на p.

4. Пусть $n > 1$. Если n простое, доказывать нечего. Если же n не простое, то $n = m_1n_1$ для некоторых целых чисел n_1, m_1, причем $1 < n_1 < n$ и $1 < m_1 < n$. Посмотрим теперь на n_1: оно либо простое, либо нет; если оно не простое, можно снова записать $n_1 = m_2n_2$, и так далее. Заметим, что $n > n_1 > n_2 > \dots$, поэтому бесконечно долго этот процесс продолжаться не может — все эти числа натуральные. Значит, на каком-то шаге мы получим простое число n_k; нетрудно видеть, что n на него делится.

5. Предположим обратное; пусть $\{p_1, \dots, p_k\}$ — множество всех простых чисел. Рассмотрим число $n = p_1 \cdot p_2 \cdot \dots \cdot p_k + 1$. По предыдущему свойству n делится на какое-то простое число p; при этом если $p = p_i$ для некоторого i, то $1 = n - p_1 \cdot p_2 \cdot \dots \cdot p_k$ делится на p_i, чего быть не может. Значит, число p не входит в множество $\{p_1, \dots, p_k\}$.

6. Пусть p_1 и p_2 не взаимно просты; тогда по пункту (1) имеем $p_1 \mid p_2$ и $p_2 \mid p_1$, то есть, они равны.

□

Теорема 2.5.3 (Основная теорема арифметики). *Каждое натуральное число, большее нуля, может быть представлено в виде произведения простых чисел, и два таких разложения могут отличаться только порядком следования сомножителей.*

Доказательство. Существование разложения для натурального числа n докажем индукцией по n. База: если $n = 1$, доказывать нечего — произведение пустого множества простых чисел равно 1. Переход: пусть теперь $n > 1$. По свойству (4) предложения 2.5.2 мы знаем, что $n = p_1 n_1$ для некоторого простого p_1. Теперь $n_1 < n$ и мы можем применить предположение индукции к n_1: $n_1 = p_2 \cdots p_k$ для некоторых простых $p_2, \ldots, p_k$. Отсюда $n = p_1 p_2 \cdots p_k$ — произведение простых чисел.

Докажем единственность разложения. Для этого снова проведем индукцию по n. В случае $n = 1$ снова доказывать нечего. Пусть $n = p_1 \cdots p_k = q_1 \cdots q_l$. Видим, что произведение $p_1 \cdots p_k$ делится на q_1. По свойству 3 простых чисел (2.5.2) один из сомножителей $p_1, \ldots, p_k$ делится на q_1. Пусть это p_i: $q_1 \mid p_i$. Но по свойству 6 простых чисел (2.5.2) из этого следует, что $p_i = q_1$. Поделим теперь обе части равенства $p_1 \cdots p_k = q_1 \cdots q_l$ на $p_i = q_1$: $p_1 \cdots \widehat{p_i} \cdots p_k = q_1 \cdots q_l$ (здесь крышечка над p_i означает, что соответствующий множитель пропущен). Полученное произведение меньше n; по предположению индукции, разложения в левой и правой частях отличаются лишь порядком следования простых сомножителей. Значит, и первоначальные разложения $p_1 \cdots p_k = q_1 \cdots q_l$ отличаются лишь порядком сомножителей. □

Определение 2.5.4. Пусть n — натуральное число, большее 0. Сгруппируем одинаковые простые числа в разложении n вместе, расположим их в порядке возрастания и запишем $n = p_1^{k_1} \cdots p_s^{k_s}$, где $p_1 < \cdots < p_s$ — простые, и $k_1, \ldots, k_s > 0$ — натуральные числа. Такая (очевидно, однозначная) запись называется **каноническим разложением** натурального числа n на простые множители.

Замечание 2.5.5. На практике полезно допускать в каноническом разложении и нулевые показатели $k_1, \ldots, k_s$ (конечно, при этом потеряется однозначность записи). К примеру, мы будем пользоваться тем, что если m, n — два ненулевых натуральных числа, то можно записать их в виде $m = p_1^{k_1} \ldots p_s^{k_s}$, $n = p_1^{l_1} \ldots p_s^{l_s}$ для некоторых *общих* простых $p_1, \ldots, p_s$ и натуральных $k_1, \ldots, k_s, l_1, \ldots, l_s$: если какие-то простые множители, скажем, есть в каноническом разложении m, но отсутствуют в разложении n, можно дописать их в разложение n с нулевыми показателями.

Приведем несколько примеров использования канонического разложения. Пусть m, n — ненулевые натуральные числа. Как по каноническому разложению m и n определить, делится ли m на n? Запишем (пользуясь замечанием 2.5.5) $m = p_1^{k_1} \cdots p_s^{k_s}$ и $n = p_1^{l_1} \cdots p_s^{l_s}$ для некоторых простых $p_1, \cdots, p_s$. Если m делит n, можно записать $n = mr$. Пусть $r = q_1 \cdots q_t$ —

какое-то разложение r на простые множители. Тогда равенство $n = mr$ превращается в равенство

$$p_1^{l_1} \cdots p_s^{l_s} = p_1^{k_1} \cdots p_s^{k_s} q_1 \cdots q_t. \tag{1}$$

Можно посмотреть на это равенство как на два разложения числа m в произведение простых. По основной теореме арифметики (2.5.3) они должны совпадать с точностью до перестановки множителей. Стало быть, если в разложении m встретилось $p_i^{k_i}$ для $k_i > 0$, то справа в равенстве 1 простой сомножитель p_i встретился как минимум k_i раз; значит, и слева он должен встретиться как минимум k_i раз. Однако слева показатель при l_i равен l_i. Значит, $k_i \leqslant l_i$. Если же $k_i = 0$ для какого-то i, то неравенство $k_i \leqslant l_i$ выполнено автоматически. Обратно, если $k_i \leqslant l_i$ для всех $i = 1, \ldots, s$, то $n = m \cdot p_1^{l_1 - k_1} \cdots p_s^{l_s - k_s}$. Мы доказали следующее предложение:

Предложение 2.5.6. *Пусть* $m = p_1^{k_1} \cdots p_s^{k_s}$, $n = p_1^{l_1} \cdots p_s^{l_s}$ *для некоторых простых* $p_1, \ldots, p_s$. m *делит* n *тогда и только тогда, когда* $k_i \leqslant l_i$ *для всех* $i = 1, \ldots, s$.

Теперь нетрудно посчитать количество всех натуральных делителей числа по его каноническому разложению.

Предложение 2.5.7. *Пусть* $n = p_1^{l_1} \cdots p_s^{l_s}$ — *каноническое разложение числа* n. *Тогда количество всех натуральных делителей* n *равно* $(1 + l_1) \cdots (1 + l_s)$.

Доказательство. По предложению 2.5.6 каждый делитель n имеет вид $p_1^{k_1} \cdots p_s^{k_s}$ для некоторых k_i таких, что $0 \leqslant k_i \leqslant l_i$, и по основной теореме арифметики (2.5.3) различные наборы (k_i) приводят к различным делителям. Значит, количество натуральных делителей n равно количеству таких наборов. Заметим, что у нас имеется $1 + l_i$ вариантов для выбора натурального k_i с условием $0 \leqslant ka_i \leqslant l_i$, и все эти выборы независимы друг от друга, поэтому простой комбинаторный подсчет показывает, что количество наборов (k_i) равно $(1 + l_1) \cdots (1 + l_s)$. □

Выразим теперь каноническое разложение наибольшего общего делителя чисел m и n через канонические разложения m и n.

Предложение 2.5.8. *Если* $m = p_1^{k_1} \cdots p_s^{k_s}$, $n = p_1^{l_1} \cdots p_s^{l_s}$ *для некоторых простых* $p_1 < \cdots < p_s$ *и* $d = \gcd(m, n)$, *то* $d = p_1^{\min(k_1, l_1)} \cdots p_s^{\min(k_s, l_s)}$.

Доказательство. Проверим, что d является общим делителем m и n. Действительно, $k_i \geqslant \min(k_i, l_i)$, поэтому $m = d \cdot p_1^{k_1 - \min(k_1, l_1)} \cdots p_s^{k_s - \min(k_s, l_s)}$ и $d \mid m$. Аналогично, $d \mid n$. Теперь пусть d' — какой-то общий делитель m и n. Заметим, что все простые множители d' тогда должны содержаться среди $p_1, \ldots, p_s$. Значит, можно записать $d' = p_1^{r_1} \cdots p_s^{r_s}$ для некоторых натуральных $r_1, \ldots, r_s$. Поскольку $d' \mid m$, по предложению 2.5.6 получаем, что $k_i \geqslant r_i$ для всех i; аналогично, $l_i \geqslant r_i$ для всех i. Но тогда и $\min(k_i, l_i) \geqslant r_i$, откуда получаем, что $d \mid d'$, рассуждая так же, как в начале доказательства. □

2.6 Сравнения и классы вычетов

Литература: [F], гл. I, § 2, п. 1; [V], гл. III, §§ 1–5; [B], гл. 8, п. 1.

Определение 2.6.1. Пусть m — ненулевое натуральное число. Говорят, что целые числа a и b **сравнимы по модулю** m, если m делит $a - b$. Обозначение: $a \equiv b \pmod{m}$, $a \equiv_m b$.

Предложение 2.6.2 (Свойства сравнений). *Пусть $m > 0$ — натуральное число.*

1. $a \equiv a \pmod{m}$;
2. *если* $a \equiv b \pmod{m}$, *то* $b \equiv a \pmod{m}$;
3. *если* $a \equiv b \pmod{m}$ *и* $b \equiv c \pmod{m}$, *то* $a \equiv c \pmod{m}$;
4. *если* $a_1 \equiv a_2 \pmod{m}$ *и* $b_1 \equiv b_2 \pmod{m}$, *то* $a_1 + b_1 \equiv a_2 + b_2 \pmod{m}$ *и* $a_1 b_1 \equiv a_2 b_2 \pmod{m}$;
5. *каждое целое число сравнимо по модулю m ровно с одним из чисел* $0, 1, \ldots, m-1$;
6. *если* $ac \equiv bc \pmod{m}$ *и* $c \perp m$, *то* $a \equiv b \pmod{m}$;
7. *сравнение* $ax \equiv 1 \pmod{m}$ *разрешимо (относительно x) тогда и только тогда, когда* $a \perp m$.

Доказательство. 1. m делит $a - a = 0$.

2. Если m делит $a - b$, то m делит $b - a = -(a - b)$.

3. Если m делит $a - b$ и $b - c$, то m делит и $a - c = (a - b) + (b - c)$.

4. Если m делит $a_1 - a_2$ и $b_1 - b_2$, то m делит $(a_1 + b_1) - (a_2 + b_2) = (a_1 - a_2) + (b_1 - b_2)$ и $a_1 b_1 - a_2 b_2 = (a_1 - a_2)b_1 + a_2(b_1 - b_2)$.

5. Пусть $n \in \mathbb{Z}$. Поделим n на m с остатком: $n = mq + r$, где $0 \leqslant r \leqslant m-1$; тогда $n - r = mq$ делится на m, поэтому $n \equiv r \pmod{m}$. С другой стороны, если $n \equiv r_1 \pmod{m}$ и $n \equiv r_2 \pmod{m}$ и $0 \leqslant r_1, r_2 \leqslant m - 1$, то $r_1 \equiv r_2$ (по уже доказанным свойствам 2 и 3), откуда $m \mid r_1 - r_2$. Но $|r_1 - r_2| \leqslant m - 1$, поэтому $r_1 = r_2$.

6. Если m делит $ac - bc = (a - b)c$, и $c \perp m$, то по свойству 3 из 2.3.3 получаем, что m делит $a - b$.

7. Если $a \perp m$, то $1 = au_0 + mv_0$ для некоторых целых u_0, v_0, откуда $au_0 - 1 = -mv_0$ делится на m, и $au_0 \equiv 1 \pmod{m}$. Обратно, если $ax_0 \equiv 1 \pmod{m}$ для некоторого x_0, то $m \mid ax_0 - 1$, значит, $ax_0 - 1 = mq$ для некоторого q, откуда $ax_0 - mq = 1$. По свойству 2 взаимной простоты (2.3.3) получаем, что $a \perp m$.

□

Замечание 2.6.3. Первые три свойства в 2.6.2 показывают, что $\equiv_m$ является отношением эквивалентности на множестве целых чисел.

2.7 Классы вычетов, действия над ними

Литература: [F], гл. I, § 2, пп. 2, 3; [K1], гл. 4, § 3, пп. 1, 2; [B], гл. 8, п. 2.

Мы знаем, что отношение сравнимости по модулю m является отношением эквивалентности на множестве целых чисел (см. 2.6.3). Значит, можно рассмотреть фактор-множество множества $\mathbb{Z}$ по этому отношению эквивалентности (см. 1.5.5).

Определение 2.7.1. Фактор-множество $\mathbb{Z}/\equiv_m$ мы будем обозначать через $\mathbb{Z}/m\mathbb{Z}$. Элементы этого множества называются **классами вычетов** по модулю m. Класс эквивалентности элемента a в $\mathbb{Z}/m\mathbb{Z}$ мы будем обозначать через $\overline{a}$ или $[a]_m$.

Замечание 2.7.2. По свойству 5 сравнений (2.6.2) каждое целое число попадает в один класс с ровно одним из чисел $0, 1, \ldots, m-1$. Это означает, что $\mathbb{Z}/m\mathbb{Z} = \{\overline{0}, \overline{1}, \ldots, \overline{m-1}\}$. В частности, получаем, что $|\mathbb{Z}/m\mathbb{Z}| = m$.

Сейчас мы определим на множестве $\mathbb{Z}/m\mathbb{Z}$ операции сложения $+$ и умножения $\cdot$. Чтобы сложить два класса вычетов, нужно выбрать в каждом из них какой-нибудь элемент (такой элемент называется *представителем* класса вычетов), сложить эти выбранные элементы и посмотреть, в какой класс попадет результат. Совершенно аналогично поступаем и с умножением. Остается проверить, что результат этой операции не зависит от выбора представителей. Эту независимость обычно называют *корректностью* определения операции.

Итак, если даны два класса $\overline{x}, \overline{y} \in \mathbb{Z}/m\mathbb{Z}$ (то есть, $x, y \in \mathbb{Z}$ — представители этих двух классов), положим $\overline{x}+\overline{y} = \overline{x+y}$ и $\overline{x}\cdot\overline{y} = \overline{xy}$. Проверим, что эти операции корректно определены: пусть теперь x', y' — другие представители тех же классов, то есть, $x' \in \overline{x}$, $y' \in \overline{y}$ (или, что то же самое, $\overline{x'} = \overline{x}$ и $\overline{y'} = \overline{y}$). По определению классов эквивалентности (1.5.3) это означает, что $x' \equiv x \pmod m$, $y' \equiv y \pmod m$. Почему же $\overline{x+y}$ совпадает с $\overline{x'+y'}$, а $\overline{xy}$ совпадает с $\overline{x'y'}$? Это в точности свойство 4 сравнений (2.6.2): $x'+y' \equiv x+y \pmod m$ и $x'y' \equiv xy \pmod m$.

2.8 Кольца и поля

Литература: [F], гл. I, § 3, п. 2; [K1], гл. 4, § 3, пп. 2, 4; [vdW], гл. 3, § 11.

В предыдущем разделе мы построили новую структуру, элементы которой могут складываться и умножаться. Эти элементы очень похожи на числа, поскольку сложение и умножение обладает фактически «теми же» свойствами, что и обычные числовые системы — множества $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$. Сейчас мы сформулируем несколько базовых свойств сложения и умножения, из которых, при желании, можно вывести аналоги большинства алгебраических тождеств, изучаемых в средней школе. Множество с операциями сложения и умножения, которые ведут себя как «настоящие» сложение и умножение, называется *кольцом*

Определение 2.8.1. Пусть R — множество, на котором заданы две бинарные операции $+$ и $\cdot$ (называемые, соответственно, *сложением* и *умножением*). Предположим, что выполняются следующие свойства:

1. $a + (b + c) = (a + b) + c$ для любых $a, b, c \in R$ (*ассоциативность сложения*).

2. существует элемент $\overline{0} \in R$ такой, что $\overline{0} + a = a = a + \overline{a}$ для всех $a \in R$ (то есть, $\overline{0}$ — *нейтральный элемент относительно сложения*; он называется **нулем** и часто обозначается просто через 0);

3. для любого $a \in R$ существует элемент $a' \in R$ такой, что $a + a' = \overline{0} = a' + a$ (то есть, a' — [двусторонний] *обратный к* a *относительно сложения*; такой элемент обычно обозначается через $-a$ и называется **противоположным** к a);

4. $a + b = b + a$ для любых $a, b \in R$ (*коммутативность сложения*);

5. $a \cdot (b + c) = a \cdot b + a \cdot c$ и $(b + c) \cdot a = b \cdot a + c \cdot a$ для любых $a, b, c \in R$ (*дистрибутивность сложения относительно умножения*).

6. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ для любых $a, b, c \in R$ (*ассоциативность умножения*);

7. существует элемент $\overline{1} \in R$ такой, что $\overline{1} \cdot a = a = a \cdot \overline{1}$ для любого $a \in R$ (то есть, $\overline{1}$ — *нейтральный элемент относительно умножения*; он называется **единицей** и часто обозначается просто через 1);

8. $a \cdot b = b \cdot a$ для любых $a, b \in R$ (*коммутативность умножения*);

Тогда R (с этими двумя операциями) называется **ассоциативным коммутативным кольцом с единицей**. Тяжеловесность этого названия связана с тем, что обычно множество с операциями, удовлетворяющее свойствам (1)–(5), называют **кольцом**, а при наложении условий (6), (7), (8) (в различных комбинациях) добавляют к слову «кольцо» эпитеты «ассоциативное», «с единицей», «коммутативное». В нашем курсе большинство встречающихся колец (во всяком случае, до пятой главы) будут обладать всеми указанными свойствами, поэтому мы часто будем называть ассоциативное коммутативное кольцо с единицей просто *кольцом*, а при необходимости говорить о *некоммутативных кольцах* или, скажем, *кольцах без единицы*.

Обратите внимание, что свойства (1), (2), (4) для сложения совершенно параллельны свойствам (6), (7), (8). Однако, свойство (3) утверждает, что сложение обладает еще одним свойством, которое не требуется от умножения. Чуть ниже мы назовем кольцо, в котором аналогичное свойство (с небольшой модификацией) выполнено для умножения, *полем*. Свойство (5) — единственное, которое связывает две операции; в каждое из остальных входит либо сложение, либо умножение по отдельности.

Примеры 2.8.2. Совершенно очевидно, что множества $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$ являются кольцами относительно обычных операций сложения и умножения; в каждом из них нейтральный элемент по сложению — это 0, а нейтральный элемент по умножению — это 1.

Предложение 2.8.3. *Пусть* m *— натуральное число,* $m \geqslant 1$. *Множество* $\mathbb{Z}/m\mathbb{Z}$ *с операциями* $+$ *и* $\cdot$*, введенными в разделе 2.7, является ассоциативным коммутативным кольцом с* 1.

Доказательство. Проверим свойство (1). Пусть x, y, z — представители классов a, b, c соответственно, то есть, $a = \overline{x}$, $b = \overline{y}$, $c = \overline{z}$. Тогда $a + (b + c) = \overline{x} + (\overline{y} + \overline{z}) = \overline{x} + \overline{y + z} = \overline{x + (y + z)}$ и $(a + b) + c = (\overline{x} + \overline{y}) + \overline{z} = \overline{x + y} + \overline{z} = \overline{(x + y) + z}$. Полученные элементы равны, поскольку сложение целых чисел ассоциативно. Остальные свойства доказываются совершенно аналогично с помощью соответствующих свойств сложения и умножения целых чисел. Заметим, что в качестве нейтрального элемента по сложению в свойстве (2) следует взять класс нуля $\overline{0}$, а в качестве нейтрального элемента по умножению в свойстве (7) — класс единицы $\overline{1}$. Наконец, если $a = \overline{x}$, то в свойстве (3) в качестве противоположного элемента нужно взять $a' = \overline{-x}$. □

Определение 2.8.4. Кольцо $\mathbb{Z}/m\mathbb{Z}$, описанное в предложении 2.8.3, называется **кольцом классов вычетов по модулю** m.

Определение 2.8.5. Множество, состоящее из одного элемента, единственным образом снабжается структурой ассоциативного коммутативного кольца с единицей. Обычно мы называем этот элемент *нулем*, а полученное кольцо $R = \{0\}$ **нулевым кольцом**, и обозначаем это кольцо через 0 (если это не вызывает путаницы в обозначениях).

Лемма 2.8.6. *Пусть* R *— кольцо.*

1. $a \cdot \overline{0} = \overline{0}$ *для всех* $a \in R$*;*
2. *если в* R *элементы* $\overline{0}$ *и* $\overline{1}$ *совпадают, то это нулевое кольцо;*
3. *если у элемента* $\overline{0} \in R$ *есть обратный по умножению, то* R *— нулевое кольцо;*

Доказательство.

1. Из определения $\overline{0}$ следует, что $\overline{0} + \overline{0} = \overline{0}$. Домножая обе части на a, получаем, что $a \cdot (\overline{0} + \overline{0}) = a \cdot \overline{0}$. Воспользуемся дистрибутивностью: $a \cdot \overline{0} + a \cdot \overline{0} = a \cdot \overline{0}$. Прибавляя к обеим частям полученного равенства противоположный элемент к $a \cdot \overline{0}$, получаем, что $a \cdot \overline{0} = \overline{0}$, что и требовалось.
2. Пусть $\overline{0} = \overline{1}$ и $a \in R$. Тогда $a \cdot \overline{0} = a \cdot \overline{1}$. Но мы только что показали, что левая часть равна $\overline{0}$, в то время как правая часть равна a. Поэтому $a = \overline{0}$, и кольцо R состоит из одного элемента.
3. Пусть $\overline{0}^{-1}$ — обратный по умножению к 0; тогда $\overline{0}^{-1} \cdot \overline{0} = \overline{1}$; с другой стороны, левая часть равна $\overline{0}$ по уже доказанному. Поэтому $\overline{0} = \overline{1}$, и R — нулевое кольцо.

□

Лемма 2.8.6 показывает, что не очень разумно ожидать, что у *каждого* элемента кольца окажется обратный по умножению: из этого тут же следовало бы, что это кольцо нулевое. Однако, если потребовать существования обратного у каждого *ненулевого* элемента, то получится разумная структура, которая называется *полем*.

Определение 2.8.7. Ассоциативное коммутативное кольцо R с единицей называется **полем**, если $R \neq 0$ и у каждого ненулевого элемента R имеется обратный по умножению. Иными словами, ненулевое кольцо R называется полем, если для любого $x \in R$ найдется $x^{-1} \in R$ такое, что $x \cdot x^{-1} = 1 = x^{-1} \cdot x$.

Примеры 2.8.8. Кольца $\mathbb{Q}$ и $\mathbb{R}$ из примера 2.8.2 являются полями, а кольцо $\mathbb{Z}$ — нет.

Множество всех обратимых элементов кольца мы будем обозначать через R^*. Так, $\mathbb{R}^* = \mathbb{R}\backslash\{0\}$, $\mathbb{Z}^* = \{-1, 1\}$.

Сейчас мы выясним, какие из колец вида $\mathbb{Z}/m\mathbb{Z}$ являются полями.

Определение 2.8.9. Пусть R — кольцо. Элемент $x \in R$ называется **делителем нуля**, если найдется ненулевой элемент $y \in R$ такой, что $xy = 0$. Делитель нуля называется **тривиальным**, если он равен нулю, и **нетривиальным**, если он не равен нулю. Кольцо R называется **областью целостности**, если $R \neq 0$ и в R нет нетривиальных делителей нуля. Иными словами, ненулевое кольцо R называется областью целостности, если из равенства $xy = 0$ следует, что $x = 0$ или $y = 0$.

Лемма 2.8.10. *Произведение обратимых элементов кольца* R *обратимо.*

Доказательство. Если $x, y \in R$ обратимы, то $y^{-1}x^{-1}$ — обратный элемент к xy. Действительно, $(xy)(y^{-1}x^{-1}) = x(yy^{-1})x^{-1} = xx^{-1} = 1$, и $(y^{-1}x^{-1})(xy) = y^{-1}(x^{-1}x)y = y^{-1}y = 1$. □

Лемма 2.8.11. *Любое поле является областью целостности.*

Доказательство. Пусть R — поле. Если в R есть нетривиальный делитель нуля $x \neq 0$, то найдется $y \neq 0$ такой, что $xy = 0$. В поле все ненулевые элементы обратимы, в том числе x и y. По лемме 2.8.10 и их произведение $xy = 0$ обратимо, и по лемме 2.8.6 кольцо R нулевое — противоречие. □

Заметим, что обратное утверждение к лемме 2.8.11 неверно: например, $\mathbb{Z}$ является областью целостности, но не полем.

Лемма 2.8.11 показывает, например, что кольцо $\mathbb{Z}/6\mathbb{Z}$ не является полем, поскольку в нем есть делители нуля. Действительно, $\overline{2} \cdot \overline{3} = \overline{6} = \overline{0}$ в $\mathbb{Z}/6\mathbb{Z}$.

Предложение 2.8.12. *Пусть* $m > 0$ *— натуральное число,* $a \in \mathbb{Z}$*. Класс* $\overline{a}$ *обратим в* $\mathbb{Z}/m\mathbb{Z}$ *тогда и только тогда, когда* $a \perp m$*.*

Доказательство. Заметим, что $\overline{x}$ является обратным к $\overline{a} \Leftrightarrow \overline{a} \cdot \overline{x} = \overline{1}$ Leftrightarrow $\overline{ax} = \overline{1}$ $\Leftrightarrow ax \equiv 1 \pmod{m}$. По предложению 2.6.2 это сравнение разрешимо относительно x тогда и только тогда, когда $a \perp m$. □

Предложение 2.8.13. *Кольцо* $\mathbb{Z}/m\mathbb{Z}$ *является полем тогда и только тогда, когда* m *— простое число.*

Доказательство. Пусть m — простое и $\overline{x} \in \mathbb{Z}/m\mathbb{Z}$ таков, что $\overline{x} \neq \overline{0}$. Стало быть, x не делится на m. По свойству 1 простых чисел (2.5.2) получаем, что $x \perp m$, и по предложению 2.8.12 класс $\overline{x}$ обратим. Обратно, если m не простое, можно записать $m = kl$ для некоторых натуральных k, l, причем $1 < k, l < m$. Тогда $\overline{k} \cdot \overline{l} = \overline{m} = \overline{0}$, и потому в $\mathbb{Z}/m\mathbb{Z}$ есть делители нуля. По лемме 2.8.11 это кольцо не может быть полем. □

2.9 Китайская теорема об остатках

Литература: [V], гл. IV, § 3.

Теорема 2.9.1 (Китайская теорема об остатках). *Пусть* $m, n \geqslant 1$ — *натуральные числа,* $m \perp n$, a, b — *целые числа. Тогда существует целое* x *такое, что* $x \equiv a \pmod{m}$, $x \equiv b \pmod{n}$. *Кроме того, целое* x' *удовлетворяет сравнениям* $x' \equiv a \pmod{m}$, $x' \equiv b \pmod{n}$ *тогда и только тогда, когда* $x' \equiv x \pmod{mn}$.

Доказательство. Воспользуемся свойством (7) сравнений (2.6.2) и найдем $x_1, x_2 \in \mathbb{Z}$ такие, что $nx_1 \equiv 1 \pmod{m}$, $mx_2 \equiv 1 \pmod{n}$. Теперь положим $x = anx_1 + bmx_2$. Мы утверждаем, что это x удовлетворяет свойствам из формулировки теоремы. Действительно, $x = anx_1 + bmx_2 \equiv a(nx_1) \equiv a \pmod{m}$ и $x = anx_1 + mbx_2 \equiv b(mx_2) \equiv b \pmod{n}$. Теперь пусть x' — целое число такое, что $x' \equiv a \pmod{m}$ и $x' \equiv b \pmod{n}$, то $x - x' \equiv a - a \equiv 0 \pmod{m}$ и $x - x' \equiv b - b \equiv 0 \pmod{n}$. Это означает, что $x - x'$ делится на m и n. Но m и n взаимно просты, поэтому по свойству 4 взаимной простоты (2.3.3) получаем, что $mn \mid x - x'$, откуда $x \equiv x' \pmod{mn}$. Обратно, если $x \equiv x' \pmod{m}n$, то $x - x'$ делится на m и на n, поэтому $x' \equiv x \equiv a \pmod{m}$ и $x' \equiv x \equiv b \pmod{n}$. □

Иными словами, система сравнений

$$\begin{cases} x \equiv a \pmod{m}, \\ y \equiv b \pmod{n} \end{cases}$$

всегда имеет решение, и это решение единственно с точностью до сравнимости по модулю mn.

2.10 Теорема Вильсона

Литература: [V], гл. IV, § 4; [B], гл. 15, п. 3.

Теорема 2.10.1 (Вильсона). *Пусть* $p \in \mathbb{N}$, $p > 1$. *Число* p *является простым тогда и только тогда, когда* $(p-1)! \equiv -1 \pmod{p}$.

Доказательство. Пусть p — простое. Посмотрим на класс $\overline{(p-1)!}$ в $\mathbb{Z}/p\mathbb{Z}$:

$$\overline{(p-1)!} = \overline{1} \cdot \overline{2} \cdot \dots \cdot \overline{(p-1)}. \tag{2}$$

В произведении справа выписаны все ненулевые элементы $\mathbb{Z}/p\mathbb{Z}$. По предложению 2.8.13 все они обратимы. Разобьем их на пары, поставив каждому классу в пару обратный к нему. Нетрудно проверить, что у каждого класса только один обратный (если a', a'' —обратные к a, то $a' = a' \cdot (a \cdot a'') = (a' \cdot a) \cdot a'' = a''$), и что $(a^{-1})^{-1} = a$.

Проблемы с разбиением на пары возникают только тогда, когда класс обратен сам себе (в этом случае получается вырожденная «пара» из одного элемента). Но таких класса только два: $\overline{1}$ и $\overline{-1}$. Действительно, если $\overline{x} \in \mathbb{Z}/p\mathbb{Z}$ таков, что $\overline{x} \cdot \overline{x} = \overline{1}$, то $x^2 \equiv 1 \pmod{p}$, откуда $p \mid x^2 - 1$, то есть, $p \mid (x-1)(x+1)$, и по свойству 2 простых чисел (2.5.2) из этого следует, что $p \mid x \pm 1$, то есть, что $x \equiv \pm 1 \pmod{p}$.

Поэтому все классы, кроме $\overline{1}$ и $\overline{-1}$ разбиваются на пары взаимно обратных, и произведение классов в каждой паре равно $\overline{1}$. Остается только домножить произведение всех классов из пар на $\overline{1}$ и $\overline{-1}$; получаем, что общее произведение, стоящее в правой части (2), равно $\overline{-1}$.

Теперь покажем, что если p не является простым, то $(p-1)!$ не сравнимо с -1 по модулю p. Пусть $p = kl$ — нетривиальное разложение p на множители. Тогда $(p-1)!$ делится на k, поскольку среди чисел $1, \dots, p-1$ встретится k. Если все-таки $(p-1)! \equiv -1 \pmod{p}$, то $p \mid (p-1)! + 1$, откуда $(p-1)! + 1 = ps$ для некоторого $s \in \mathbb{Z}$, откуда $1 = ps - (p-1)!$ делится на k (поскольку p делится на k и $(p-1)!$ делится на k) — противоречие. □

2.11 Функция Эйлера

Литература: [F], гл. I, § 2, п. 3; [V], гл. II, § 4; [B], гл. 10.

Определение 2.11.1. Пусть $n \in \mathbb{N}$, $n > 0$. Количество натуральных чисел, меньших n и взаимно простых с n, обозначается через $\varphi(n)$. Иными словами, $\varphi(n) = |\{x \in \mathbb{N} \mid x < n \text{ и } x \perp n\}|$. Сопоставление $n \mapsto \varphi(n)$ задает функцию $\mathbb{N}\backslash\{0\} \to \mathbb{N}$, которая называется **функцией Эйлера**.

Пример 2.11.2. Прямое вычисление показывает, что $\varphi(1) = 1$, $\varphi(2) = 1$, $\varphi(3) = 2$.

Предложение 2.11.3. *Пусть $n \in \mathbb{N}$, $n > 0$. Тогда $\varphi(n)$ равно количеству обратимых элементов кольца $\mathbb{Z}/n\mathbb{Z}$: $\varphi(n) = |(\mathbb{Z}/n\mathbb{Z})^*|$.*

Доказательство. Пусть $0 \leqslant x < n$; по предложению 2.8.12 $x \perp n$ тогда и только тогда, когда $\overline{x}$ обратим. □

Замечание 2.11.4. Теперь можно посчитать $\varphi(p)$ для простого p: по предложению 2.8.13 кольцо $\mathbb{Z}/p\mathbb{Z}$ является полем, то есть, $(\mathbb{Z}/p\mathbb{Z})^* = (\mathbb{Z}/p\mathbb{Z})\backslash\{\overline{0}\}$, откуда $\varphi(p) = |(\mathbb{Z}/p\mathbb{Z})^*| = p - 1$. Это можно получить и прямым подсчетом: число x, $0 \leqslant x < p$, взаимно просто с p тогда и только тогда, когда оно не делится на p, то есть, когда оно не равно 0.

Прямой подсчет позволяет вычислить и $\varphi(p^k)$, где p — простое, $k > 0$ — натуральное. Действительно, x взаимно прост с p^k тогда и только тогда, когда x взаимно прост p, то есть, x не делится на p. Количество натуральных чисел, меньших p^k и делящихся на p, равно $p^k/p = p^{k-1}$, поэтому $\varphi(p^k) = p^k - p^{k-1} = p^{k-1}(p-1)$.

Для того, чтобы вычислить значение $\varphi(n)$ по каноническому разложению числа n, нам понадобится переформулировка китайской теоремы об остатках.

Теорема 2.11.5. *Пусть натуральные числа $m, n \geqslant 1$ таковы, что $m \perp n$. Рассмотрим отображение $f\colon \mathbb{Z}/mn\mathbb{Z} \to \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$, сопоставляющее классу $\overline{x} = [x]_{mn} \in \mathbb{Z}/mn\mathbb{Z}$ пару классов $([x]_m, [x]_n)$. Это отображение корректно определено и является биекцией.*

Доказательство. Корректная определенность: если $[x]_{mn} = [x']_{mn}$, то $mn \mid x - x'$, поэтому $m \mid x - x'$ и $n \mid x - x'$. Значит, $[x]_m = [x']_m$ и $[x]_n = [x']_n$. По китайской теореме об остатках (2.9.1) для каждой пары $(a, b) \in \mathbb{Z}/m\mathbb{Z} \times \mathbb{Z}/n\mathbb{Z}$ найдется x такой, что $f(\overline{x}) = (a, b)$ и такой x единственный по модулю mn, то есть, задает однозначно определенный элемент $[x]_{mn} \in \mathbb{Z}/mn\mathbb{Z}$. Это и означает биективность f. □

Покажем теперь, что при построенном в теореме 2.11.5 отображении обратимые классы переходят в пары обратимых классов.

Предложение 2.11.6. *Пусть* m, n, f *таковы, как в формулировке теоремы 2.11.5, и пусть* $\overline{x} \in \mathbb{Z}/mn\mathbb{Z}$, $f(\overline{x}) = (a, b)$. *Класс* $\overline{x}$ *обратим в* $\mathbb{Z}/mn\mathbb{Z}$ *тогда и только тогда, когда* a *обратим в* $\mathbb{Z}/m\mathbb{Z}$ *и* b *обратим в* $\mathbb{Z}/n\mathbb{Z}$.

Доказательство. Если $\overline{x'}$ — обратный элемент к $\overline{x}$ в $\mathbb{Z}/mn\mathbb{Z}$ и $f(x') = (a', b')$, то a' обратен к a, а b' обратен к b. Действительно, $a = [x]_m$, $a' = [x']_m$, поэтому $a \cdot a' = [x]_m \cdot [x']_m = [x \cdot x']_m$, но $xx' \equiv 1 \pmod{mn}$, поэтому $xx' \equiv 1 \pmod{m}$. Аналогично, b' является обратным к b.

Обратно, пусть a' — обратный к a, b' — обратный к b. Отображение f биективно, поэтому найдется x' такой, что $f(\overline{x'}) = (a', b')$, то есть, $[x']_m = a'$, $[x']_n = b'$. При этом $[xx']_m = [x]_m \cdot [x']_m = a \cdot a' = [1]_m$ и $[xx']_n = [1]_n$. Значит, $xx' \equiv 1 \pmod{m}$ и $xx' \equiv 1 \pmod{n}$, откуда по свойству 1 взаимно простых чисел (2.3.3) $xx' \equiv 1 \pmod{mn}$ и x обратим. □

Теорема 2.11.7 (Мультипликативность функции Эйлера). *Если* $m, n \geqslant 1$ — *натуральные числа и* $m \perp n$, *то* $\varphi(mn) = \varphi(m)\varphi(n)$.

Доказательство. По предложению 2.11.3, $\varphi(mn) = |(\mathbb{Z}/mn\mathbb{Z})^*|$ и $\varphi(m)\varphi(n) = |(\mathbb{Z}/m\mathbb{Z})^*| \cdot |(\mathbb{Z}/n\mathbb{Z})^*| = |(\mathbb{Z}/m\mathbb{Z})^* \times (\mathbb{Z}/n\mathbb{Z})^*|$ Предложение 2.11.6 утверждает, что f устанавливает биекцию между множествами $(\mathbb{Z}/mn\mathbb{Z})^*$ и $(\mathbb{Z}/n\mathbb{Z})^* \times (\mathbb{Z}/n\mathbb{Z})^*$, поэтому в них поровну элементов. □

Следствие 2.11.8. *Если* $n = p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_s^{k_s}$ — *каноническое разложение натурального числа* n, *то* $\varphi(n) = p_1^{k_1-1}(p_1 - 1) \cdot p_2^{k_2-1}(p_2 - 1) \cdot \dots \cdot p_s^{k_s-1}(p_s - 1)$.

Доказательство. Заметим, что все сомножители вида $p_i^{k_i}$ в каноническом разложении числа n попарно взаимно просты (например, это следует из предложения 2.5.8). Применяя теорему 2.11.7 и замечание 2.11.4, получаем $\varphi(n) = \varphi(p_1^{k_1} \cdot p_2^{k_2} \cdot \dots \cdot p_s^{k_s}) = \varphi(p_1^{k_1}) \cdot \varphi(p_2^{k_2}) \cdot \dots \cdot \varphi(p_s^{k_s}) = p_1^{k_1-1}(p_1 - 1) \cdot p_2^{k_2-1}(p_2 - 1) \cdot \dots \cdot p_s^{k_s-1}(p_s - 1)$, что и требовалось. □

2.12 Теорема Эйлера и малая теорема Ферма

Литература: [F], гл. I, § 2, п. 3; [V], гл. III, § 6; [B], гл. 11, § 1.

Теорема 2.12.1 (Теорема Эйлера). *Пусть* n — *натуральное число,* $a \in \mathbb{Z}$ *и* $a \perp n$. *Тогда* $a^{\varphi(n)} \equiv 1 \pmod{n}$.

Доказательство. Пусть $x_1, x_2, \dots, x_k$ — все обратимые элементы кольца $\mathbb{Z}/n\mathbb{Z}$. По предложению 2.11.3 их ровно $\varphi(n)$, то есть, $k = \varphi(n)$. Пусть $\overline{a}$ — класс числа a в кольце $\mathbb{Z}/n\mathbb{Z}$. По предложению 2.8.12 элемент $\overline{a}$ обратим. Рассмотрим элементы $\overline{a}x_1, \overline{a}x_2, \dots, \overline{a}x_k$. По лемме 2.8.10 каждый из них обратим. С другой стороны, если $\overline{a}x_i = \overline{a}x_j$, то $\overline{a}(x_i - x_j) = \overline{0}$. Домножая это равенство на $\overline{a}^{-1}$, получаем, что $x_i = x_j$. Это означает, что все элементы $\overline{a}x_1, \overline{a}x_2, \dots, \overline{a}x_k$ различны; иными словами, это элементы $x_1, x_2, \dots, x_k$, только, возможно, в другом порядке. Но тогда произведения этих двух наборов элементов совпадают. Значит,

$$x_1x_2 \cdots x_k = \overline{a}x_1 \cdot \overline{a}x_2 \cdot \dots \cdot \overline{a}x_k = \overline{a}^k x_1x_2 \cdots x_k.$$

По лемме 2.8.10 произведение $x_1 x_2 \cdots x_k$ обратимо, поэтому на него можно сократить обе части (более строго — домножить на обратное к нему). Получаем, что $\overline{a}^k = \overline{1}$; это и означает, что $a^k \equiv 1 \pmod{n}$. □

Следствие 2.12.2 (Малая теорема Ферма). *Если* p *— простое число, и* $a \in \mathbb{Z}$ *не делится на* p*, то* $a^{p-1} \equiv 1 \pmod{p}$.

Доказательство. По свойству 1 простых чисел (2.5.2) $a \perp p$; по замечанию 2.11.4 $\varphi(p) = p-1$. Осталось применить теорему Эйлера для $n = p$. □

Приведем несложное следствие малой теоремы Ферма.

Следствие 2.12.3. *Если* p *— простое число, и* $a \in \mathbb{Z}$*, то* $a^p \equiv a \pmod{p}$.

Доказательство. Если $p \mid a$, то $a^p \equiv 0 \pmod{p}$ и $a \equiv 0 \pmod{p}$. В противном случае можно применить малую теорему Ферма 2.12.2: получим, что $a^{p-1} \equiv 1 \pmod{p}$; домножая обе части на a, получаем нужное сравнение. □

2.13 Алгоритм шифрования RSA

Алгоритм шифрования RSA (Rivest, Shamir, Adleman) является одной из простейших криптографических систем с открытым ключом. Он позволяет обмениваться сообщениями по открытым каналам связи без риска быть подслушанным. Пусть Алиса и Боб — два персонажа, и Алиса хочет получить от Боба сообщение, которое сможет прочесть только она. При этом между Алисой и Бобом имеются только общедоступные каналы связи. Алгоритм шифрования RSA говорит, что Алиса должна

- выбрать два случайных различных простых числа (достаточно больших) p и q;
- перемножить их и получить число $n = pq$;
- найти $\varphi(n) = \varphi(pq) = (p-1)(q-1)$;
- выбрать некоторое натуральное число e, взаимно простое с $\varphi(n)$;
- найти число d, являющееся решением сравнения $ed \equiv 1 \pmod{\varphi(n)}$ — существование такого числа гарантируется свойством (7) сравнений (2.6.2). Запишем $ed = 1 + k\varphi(n) = 1 + k(p-1)(q-1)$.

После этого Алиса передает Бобу по открытому каналу связи числа n и e. Мы предполагаем, что *сообщение*, которое Боб хочет передать Алисе, является натуральным числом m таким, что $m < n$. На практике это означает, что Боб должен разрезать длинное сообщение (строку бит) на куски длиной меньше, чем количество цифр в двоичной записи числа n и передавать каждый кусок по отдельности. Для зашифровки Боб вычисляет остаток от деления m^e на n; то есть, целое число $c < n$ такое, что $c \equiv m^e \pmod{n}$. Алиса получает зашифрованное сообщение c от Боба по открытому каналу связи и вычисляет $c^d \equiv (m^e)^d \equiv m^{ed} \pmod{n}$.

Покажем, что $m^{ed} \equiv m \pmod{n}$. Мы хотим показать, что $n \mid m^{ed} - m$. При этом n является произведением двух простых чисел: $n = pq$. По свойству (4) взаимно простых чисел (предложение 2.3.3) для этого достаточно доказать, что $p \mid m^{ed} - m$ и $q \mid m^{ed} - m$. Если $p \mid m$, то $p \mid m^{ed}$, и потому $p \mid m^{ed}$. Если же $p \perp m$, то можно применить малую теорему Ферма 2.12.2:

$$m^{ed} \equiv m^{1+k(p-1)(q-1)} \equiv m \cdot (m^{p-1})^{k(q-1)} \equiv m \cdot 1^{k(q-1)} \equiv m \pmod{p}$$

Аналогично доказывается, что $q \mid m^{ed} - m$.

Отметим, что сравнение $m^{ed} \equiv m \pmod{n}$ сразу следует из теоремы Эйлера, если $m \perp n$. На практике вероятность того, что m имеет общий делитель с n, чрезвычайно мала. Однако нужное сравнение выполнено и в общем случае.

Таким образом, Алиса восстановила исходное сообщение m. При этом все вычисления происходят по модулю n (то есть, само по себе число m^e огромное, но нас интересует только его остаток по модулю n, что значительно упрощает вычисления).

Заметим, что постороннему наблюдателю доступны лишь числа n, e и зашифрованное сообщение c. Для расшифровки необходимо знать обратный к e по модулю $\varphi(n)$ элемент d, для чего необходимо знать $\varphi(n)$. Но для вычисления $\varphi(n)$ необходимо знать разложение n на простые множители. В настоящее время неизвестны эффективные алгоритмы разложения больших чисел на простые множители (в отличие от эффективных тестов на простоту, с помощью которых Алиса и готовит числа p и q).

3 Комплексные числа

3.1 Определение комплексных чисел

Литература: [F], гл. II, § 1, пп. 1–5; [K1], гл. 5, § 1, пп. 1–2.

Комплексные числа представляют собой расширение поля вещественных чисел, обладающее гораздо более приятными алгебраическими свойствами. Наш подход к определению комплексных чисел аксиоматический — мы сначала описываем некоторое множество с операциями, которое оказывается полем, а потом показываем, что оно содержит вещественные числа и задумываемся о мотивации.

Определение 3.1.1. Рассмотрим множество $\mathbb{R} \times \mathbb{R}$ пар вещественных чисел. Введем на нем операции сложения и умножения:

$$(a, b) + (c, d) = (a + c, b + d),$$
$$(a, b) \cdot (c, d) = (ac - bd, ad + bc).$$

Теорема 3.1.2. *Множество с операциями, определенное в 3.1.1, является ассоциативным коммутативным кольцом с единицей.*

Доказательство. Необходимо проверить восемь аксиом из определения 2.8.1.

1. $((a, b)+(c, d))+(e, f) = (a+c, b+d)+(e, f) = ((a+c)+e, (b+d)+f)$, $(a, b)+((c, d)+(e, f)) = (a, b) + (c + e, d + f) = (a + (b + c), d + (e + f))$. Полученные выражения равны, поскольку сложение вещественных чисел ассоциативно.

2. Нейтральным элементом по сложению является пара $(0, 0)$. Действительно, $(a, b)+(0, 0) = (a + 0, b + 0) = (a, b)$, и по коммутативности сложения (аксиома 4) то же верно, если складывать в другом порядке.

3. Противоположным элементом к паре (a, b) является пара $(-a, -b)$. Действительно, $(a, b) + (-a, -b) = (a + (-a), b + (-b)) = (0, 0)$.

4. $(a, b) + (c, d) = (a + c, b + d) = (c + a, d + b) = (c, a) + (d, b)$.

5. $((a, b) \cdot (c, d)) \cdot (e, f) = (ac - bd, ad + bc) \cdot (e, f) = ((ac - bd)e - (ad + bc)f, (ac - bd)f + (ad + bc)e)$. С другой стороны, $(a, b) \cdot ((c, d) \cdot (e, f)) = (a, b) \cdot (ce - df, cf + de) = (a(ce - df) - b(cf + de), a(cf + de) + b(ce - df))$. Раскрытие скобок показывает, что полученные выражения равны.

6. Нейтральным элементом по умножению является пара $(1, 0)$. Действительно, $(a, b) \cdot (1, 0) = (a \cdot -b \cdot 0, a \cdot 0 + b \cdot 1 = (a, b)$, и этого достаточно в силу коммутативности умножения (аксиома 7).

7. $(a, b) \cdot (c, d) = (ac - bd, ad + bc)$ и $(c, d) \cdot (a, b) = (ca - db, cb + da)$.

8. $(a,b)\cdot((c,d)+(e,f)) = (a,b)\cdot(c+e,d+f) = (a(c+e)-b(d+f), a(d+f)-b(c+e))$. С другой стороны, $(a,b)\cdot(c,d)+(a,b)\cdot(e,f) = (ac-bd,ad+bc)+(ae-bf,af+be) = (ac-bd+ae-bf,ad+bc+af+be)$. Раскрытие скобок показывает, что полученные выражения равны; и этого достаточно в силу коммутативности умножения (аксиома 7).

□

Определение 3.1.3. Множество таких пар вещественных чисел с определенными в 3.1.1 операциями обозначается через $\mathbb{C}$; его элементы называются **комплексными числами**.

Замечание 3.1.4. Множество вещественных чисел можно считать подмножеством множества комплексных чисел: число $a \in \mathbb{R}$ можно рассматривать как комплексное число $(a,0)$. При этом введенные нами операции на парах превращаются в обычные операции над комплексными числами: действительно, $(a,0)+(b,0) = (a+b,0)$ и $(a,0)\cdot(b,0) = (ab,0)$; единица $(1,0)$ и нуль $(0,0)$ в множестве комплексных чисел являются вещественными числами 1 и 0. Заметим также, что $a\cdot(c,d) = (a,0)\cdot(c,d) = (ac,ad)$.

Определение 3.1.5. Пусть $z = (a,b)$ — комплексное число; запишем $z = (a,b) = (a,0)+(0,b) = a + b\cdot(0,1)$. Комплексное число $(0,1)$ обозначается через i и называется **мнимой единицей**; основанием этому служит тому, что $i^2 = -1$. Запись $z = a+bi$ называется **алгебраической формой записи комплексного числа**, вещественные числа a и b — **вещественной частью** и **мнимой частью** комплексного числа z соответственно. Обозначения: $a = \mathrm{Re}(z)$, $b = \mathrm{Im}(z)$.

Замечание 3.1.6. Теперь мы можем забыть про интерпретацию комплексного числа как пары вещественных чисел и считать, что комплексное число — это выражение вида $a+bi$ с вещественными a,b. При этом введенные нами в 3.1.1 операцию переписываются в алгебраической форме следующим образом:

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

Иными словами, комплексные числа — это выражения вида $a+bi$, которые складываются и перемножаются согласно обычным правилам обращения с числами с учетом равенства $i^2 = -1$.

3.2 Комплексное сопряжение и модуль

Литература: [F], гл. II, § 1, пп. 3–5, § 2, пп. 1–4; [K1], гл. 5, § 1, п. 3.

Определение 3.2.1. Сопоставим комплексному числу $z = a+bi$ комплексное число $\overline{z} = a-bi$. Полученное отображение $\mathbb{C}\to\mathbb{C}$ называется **сопряжением**, а число $\overline{z}$ — **сопряженным** к числу z.

Предложение 3.2.2 (Свойства сопряжения). *Для любых комплексных чисел $z,w\in\mathbb{C}$ выполняются следующие свойства:*

1. $\overline{z+w} = \overline{z} + \overline{w}$;

2. $\overline{z \cdot w} = \overline{z} \cdot \overline{w}$;

3. $\overline{\overline{z}} = z$;

4. $z = \overline{z}$ *тогда и только тогда, когда* $z \in \mathbb{R}$;

5. $\overline{z} \cdot z = z \cdot \overline{z}$ — *неотрицательное вещественное число; оно равно нулю тогда и только тогда, когда* $z = 0$.

Доказательство. Пусть $z = a + bi$, $w = c + di$.

1. $\overline{(a+bi)+(c+di)} = \overline{(a+c)+(b+d)i} = (a+c)-(b+d)i$, $\overline{a+bi} + \overline{c+di} = (a-bi)+(c-di) = (a+c)-(b+d)i$.

2. $\overline{(a+bi)(c+di)} = \overline{(ac-bd)+(ad+bc)i} = (ac-bd)-(ad+bc)i$, $\overline{a+bi} \cdot \overline{c+di} = (a-bi)(c-di) = (ac-bd)-(ad+bc)i$.

3. $\overline{\overline{z}} = \overline{a-bi} = a+bi$.

4. Если $z \in \mathbb{R}$, то $z = a + 0i$ и $\overline{z} = a - 0i = z$. Обратно, если $a + bi = a - bi$, то $b = -b$, откуда $b = 0$ и $z = a \in \mathbb{R}$.

5. $z \cdot \overline{z} = (a+bi)(a-bi) = (a^2+b^2)+(-ab+ba)i = a^2+b^2 \geqslant 0$, и $a^2+b^2 = 0$ тогда и только тогда, когда $a = b = 0$, то есть, когда $z = 0$.

□

Определение 3.2.3. Поскольку $z \cdot \overline{z}$ — неотрицательное вещественное число, из него можно извлечь (также неотрицательный) квадратный корень. Этот корень называется **модулем** комплексного числа z и обозначается через $|z|$; таким образом, $z \cdot \overline{z} = |z|^2$. Если $z = a + bi$ — алгебраическая форма записи комплексного числа, то $|z| = \sqrt{a^2+b^2}$.

Предложение 3.2.4. *Множество* $\mathbb{C}$ *комплексных чисел является полем.*

Доказательство. После доказательства теоремы 3.1.2 остается проверить наличие обратного по умножению у каждого ненулевого элемента. Пусть $z \in \mathbb{C}$, $z \neq 0$. Тогда $|z| \neq 0$. Рассмотрим число $z' = \frac{1}{|z|^2}\overline{z}$; легко видеть, что $z \cdot z' = z' \cdot z = 1$. □

Замечание 3.2.5. Таким образом, в множестве комплексных чисел можно делить на ненулевые элементы: $z/w = zw^{-1}$. Также определена операция возведения в целую степень: если $n > 0$, то $z^n = \underbrace{z \cdot \dots \cdot z}_{n}$, если $n < 0$ (и $z \neq 0$), то $z^n = \underbrace{z^{-1} \cdot \dots \cdot z^{-1}}_{-n}$, если же $n = 0$, то $z^0 = 1$. Нетрудно видеть, что эта операция удовлетворяет обычным свойствам возведения в степень, типа $z^{m+n} = z^m \cdot z^n$ и $(zw)^n = z^n w^n$.

Предложение 3.2.6 (Свойства модуля комплексных чисел).

1. $|z| \cdot |w| = |z \cdot w|$;

2. *если* $w \neq 0$, *то* $|z|/|w| = |z/w|$.

Доказательство. 1. $|zw| = \sqrt{(zw)(\overline{zw})} = \sqrt{z \cdot w \cdot \overline{z} \cdot \overline{w}} = \sqrt{z\overline{z} \cdot w\overline{w}} = \sqrt{z\overline{z}}\sqrt{w\overline{w}} = |z| \cdot |w|$.

2. Домножая на $|w|$, получаем, что нужно доказать $|z| = |z/w| \cdot |w|$, что следует из первой части.

□

Замечание 3.2.7. Комплексные числа удобно изображать в виде точек плоскости. Рассмотрим декартову систему координат на плоскости и сопоставим комплексному числу $a + bi$ вектор с координатами (a, b) (то есть, радиус-вектор точки (a, b)). Сложение векторов (как и комплексных чисел) происходит покоординатно, поэтому сумма векторов изображает сумму комплексных чисел. Модуль комплексного числа в силу теоремы Пифагора равен длине соответствующего вектора.

Предложение 3.2.8 (Неравенство треугольника). *Для любых комплексных чисел* z_1, z_2, z_3 *выполнено неравенство* $|z_1 - z_2| + |z_2 - z_3| \geqslant |z_3 - z_1|$.

Доказательство. Обозначим $z = z_1 - z_2$, $w = z_2 - z_3$; нужно доказать, что $|z| + |w| \geqslant |z + w|$. Заметим, что если $z + w = 0$, неравенство очевидно. Запишем $1 = \frac{z}{z+w} + \frac{w}{z+w}$. Согласно правилу сложения комплексных чисел, $\operatorname{Re} 1 = \operatorname{Re}(\frac{z}{z+w}) + \operatorname{Re}(\frac{w}{z+w})$. Заметим, что $\operatorname{Re}(z) \leqslant |z|$ для любого комплексного числа z, поэтому $\operatorname{Re} 1 \leqslant |\frac{z}{z+w}| + |\frac{w}{z+w}|$. Домножая на знаменатель, получаем необходимое неравенство. □

3.3 Тригонометрическая форма записи комплексного числа

Литература: [F], гл. II, § 2, пп. 1–6; [K1], гл. 5, § 1, п. 4.

Определение 3.3.1. Пусть $z = a + bi \in \mathbb{C}$ — ненулевое комплексное число. Обозначим через $r = \sqrt{a^2 + b^2}$ модуль числа z. Вещественные числа a/r и b/r таковы, что сумма их квадратов равна 1. Поэтому найдется такой угол φ, что $a/r = \cos(\varphi)$, $b/r = \sin(\varphi)$. Такой угол φ называется **аргументом** комплексного числа z. Заметим, что при этом

$$z = |z| \cdot z/|z| = |z|(\frac{a}{r} + \frac{b}{r}i) = |z|(\cos(\varphi) + i\sin(\varphi)).$$

Выражение $z = r(\cos(\varphi) + i\sin(\varphi))$ называется **тригонометрической формой записи комплексного числа**. Обозначение: $\varphi = \arg(z)$. Как обычно, можно считать, что аргумент (как и любой угол) записывается вещественным числом с точностью до $2\pi k$, $k \in \mathbb{Z}$. Если выбрать представитель в полуинтервале $[0, 2\pi)$, получим то, что называется **главным значением аргумента**, оно обозначается через $\operatorname{Arg}(z)$ Обратно, по модулю r и аргументу φ комплексное число z однозначно восстанавливается: $z = a + bi$, $a = r\cos(\varphi)$, $b = r\sin(\varphi)$.

Обратите внимание на необходимость осторожного обращения с понятием угол. Аргумент комплексного числа z, вообще говоря, является не вещественным числом, а углом (позднее мы придадим этому точный смысл: $\arg(z)$ — элемент *группы углов*, см. пример 10.1.3(7)). Этот угол можно записать вещественным числом, но не однозначным образом: некоторые вещественные числа записывают одинаковые углы. Например, числа 0, 2π, -2π, 4π, -4π,... — это разные формы записи одного и того же угла. При этом два вещественных числа α и β записывают один и тот же угол если и только если они отличаются на целое кратное 2π: $\alpha - \beta = 2\pi k$ для некоторого $k \in \mathbb{Z}$. Это похоже на делимость целых чисел: α и β задают один угол, если их разность «делится» на 2π. Это наводит на мысль, что углы — это классы эквивалентности по описанному отношению «сравнимости по модулю 2π».

Предложение 3.3.2 (Единственность тригонометрической формы записи). *Пусть* r, r' *— положительные вещественные числа,* φ, φ' *— углы,* $z = r(\cos(\varphi) + i\sin(\varphi))$, $z' = r'(\cos(\varphi') + i\sin(\varphi'))$ *Равенство комплексных чисел* $z = z'$ *выполнено тогда и только тогда, когда* $r = r'$ *и* $\varphi = \varphi'$.

Доказательство. Модуль комплексного числа z равен

$$\begin{aligned}\sqrt{(r\cos(\varphi))^2 + (r\sin(\varphi))^2} &= \sqrt{(r^2((\cos(\varphi))^2 + (\sin(\varphi))^2))} \\ &= r;\end{aligned}$$

аналогично, модуль комплексного числа z' равен r'. Если $z = z'$, то $r = r'$, откуда $z/r = z'/r'$. Значит, $\cos(\varphi) + i\sin(\varphi) = \cos(\varphi') + i\sin(\varphi')$, откуда $\cos(\varphi) = \cos(\varphi')$ и $\sin(\varphi) = \sin(\varphi')$. Но если у двух углов совпадают синусы и совпадают косинусы, то они равны. Поэтому и $\varphi = \varphi'$. Обратно, если $r = r'$ и $\varphi = \varphi'$, то очевидно, что $z = z'$. □

Замечание 3.3.3. Таким образом, z можно задавать не парой вещественных чисел, а парой $(|z|, \arg(z))$, состоящей из положительного вещественного числа и угла. Единственное исключение — случай $z = 0$: у нуля модуль равен нулю, а аргумент вообще не определен. Чем полезно такое задание? В алгебраической форме записи комплексные числа легко складывать: вещественные части складываются и мнимые части складываются. Оказывается, в тригонометрической форме записи комплексные числа легко перемножать.

Теорема 3.3.4. *При перемножении комплексных чисел их модули перемножаются, а аргументы складываются. Иными словами, если* $z, w \in \mathbb{C}^*$, *то* $|zw| = |z| \cdot |w|$ *и* $\arg(zw) = \arg(z) + \arg(w)$.

Доказательство. Первое утверждение было доказано в предложении 3.2.6. Обозначим $\varphi = \arg(z)$, $\psi = \arg(w)$. Заметим, что

$$\begin{aligned}zw &= |z|(\cos(\varphi) + i\sin(\varphi))|w|(\cos(\psi) + i\sin(\psi)) \\ &= |z| \cdot |w|(\cos(\varphi)\cos(\psi) - \sin(\varphi)\sin(\psi) + i(\cos(\varphi)\sin(\psi) + \sin(\varphi)\cos(\varphi))) \\ &= |z| \cdot |w|(\cos(\varphi + \psi) + i\sin(\varphi + \psi)).\end{aligned}$$

С другой стороны, $zw = |zw| \cdot (\cos(\arg(zw)) + i\sin(\arg(zw)))$. По предложению 3.3.2 из этого следует, что $|zw| = |z| \cdot |w|$ (что мы знали и раньше) и $\arg(zw) = \varphi + \psi = \arg(z) + \arg(w)$, что и требовалось. □

Следствие 3.3.5. *Для любого ненулевого комплексного числа* $z = r(\cos(\varphi) + i\sin(\varphi))$ *имеем* $z^{-1} = r^{-1}(\cos(-\varphi) + i\sin(-\varphi))$.

Следствие 3.3.6. *При делении комплексных чисел их модули делятся, а аргументы вычитаются.*

Следствие 3.3.7 (Формула де Муавра). *Для любого ненулевого комплексного числа* $z = r(\cos(\varphi) + i\sin(\varphi))$ *и любого целого* n *имеет место равенство* $z^n = r^n(\cos(n\varphi) + i\sin(n\varphi))$.

Доказательство. Для $n = 0$ равенство очевидно; для $n > 0$ следует из теоремы 3.3.4 по индукции, а случай отрицательного n сводится к случаю положительного при помощи равенства $z^n = (z^{-1})^{-n}$ и следствия 3.3.5. □

3.4 Корни из комплексных чисел

Литература: [F], гл. II, § 3, пп. 1–2; [K1], гл. 5, § 1, п. 4.

Пусть n — положительное натуральное число, $w \in \mathbb{C}$. Посмотрим на решения уравнения $z^n = w$. Во-первых, заметим, что если $w = 0$, то и $z = 0$ (иначе из равенства $z^n = 0$ делением на z^n получаем $1 = 0$). Пусть теперь $w \neq 0$. Запишем w и z в тригонометрической форме: $w = r(\cos(\varphi) + i\sin(\varphi))$, $z = |z| \cdot (\cos(\arg(z)) + i\sin(\arg(z)))$. По формуле де Муавра (3.3.7) $z^n = |z|^n \cdot (\cos(n\arg(z)) + i\sin(n\arg(z)))$. Приравнивая z^n к w и пользуясь единственностью тригонометрической записи (3.3.2), получаем, что $|z|^n = r$ и $n\arg(z) = \varphi$. Отсюда следует, что $|z| = r^{1/n}$. Кроме того, равенство углов $n\arg(z) = \varphi$ означает равенство $n\psi = \varphi + 2\pi k$, где ψ — некоторый числовой представитель угла $\arg(z)$, а k — целое число. Значит, $\psi = (\varphi + 2\pi k)/n$.

Теорема 3.4.1. *Пусть* $w = r(\cos(\varphi) + i\sin(\varphi)) \in \mathbb{C}^*$, n *— положительное натуральное число. Существует ровно* n *комплексных чисел* z *таких, что* $z^n = w$*; можно записать их так:*

$$z = r^{1/n}\left(\cos\left(\frac{\varphi + 2\pi k}{n}\right) + i\sin\left(\frac{\varphi + 2\pi k}{n}\right)\right),$$

где $k = 0, 1, \ldots, n-1$.

Доказательство. Выше мы проверили, что решения уравнения $z^n = w$ имеют вид

$$z_k = r^{1/n}\left(\cos\left(\frac{\varphi + 2\pi k}{n}\right) + i\sin\left(\frac{\varphi + 2\pi k}{n}\right)\right).$$

Осталось разобраться с их количеством и устранить неоднозначность: дело в том, что при различных целых k эта формула часто дает одинаковые значения z. А именно, $z_k = z_l$ тогда и только тогда, когда углы $(\varphi + 2\pi k)/n$ и $(\varphi + 2\pi l)/n$ совпадают. А это происходит тогда, когда их числовые значения отличаются на целое кратное 2π: $(\varphi + 2\pi k)/n = (\varphi + 2\pi l)/n + 2\pi t$, откуда $\varphi + 2\pi k = \varphi + 2\pi l + 2\pi t n$ и $k - l = tn$, то есть, $k \equiv l \pmod{n}$. Значит различных значений z столько же, сколько классов вычетов по модулю n, и можно выбрать z_k, соответствующие различным представителям k этих классов вычетов (см. 2.7.2), например, $k = 0, 1, \ldots, n-1$. □

3.5 Корни из единицы

Литература: [F], гл. II, § 4, пп. 1–4.

Пусть n — положительное натуральное число. Посмотрим на решения уравнения $z^n = 1$ в комплексных числах.

Определение 3.5.1. Пусть $n \in \mathbb{N}$, $n \geqslant 1$. Комплексное число $z \in \mathbb{C}$ называется **корнем n-ой степени из 1**, если $z^n = 1$. Множество всех корней степени n из 1 обозначается через μ_n.

Предложение 3.5.2 (Свойства корней n-ой степени из 1). *Для каждого натурального $n \geqslant 1$ существуют ровно n корней степени n из 1; это числа $\varepsilon_0^{(n)}, \varepsilon_1^{(n)}, \ldots, \varepsilon_{n-1}^{(n)}$, где*

$$\varepsilon_k^{(n)} = \cos(\frac{2\pi k}{n}) + i\sin(\frac{2\pi k}{n}).$$

При этом произведение двух корней степени n из 1 является корнем степени n из 1; обратный к корню степени n из 1 является корнем степени n из 1.

Доказательство. Формула для $\varepsilon_k^{(n)}$ немедленно следует из теоремы 3.4.1 (с учетом того, что $|1| = 1$ и $\arg(1) = 0$. Если $z, w \in \mu_n$, то $z^n = 1$, $w^n = 1$, откуда $(zw)^n = z^n \cdot w^n = 1$, поэтому и $zw \in \mu_n$. Кроме того, $(z^{-1})^n = (z^n)^{-1} = 1$, поэтому и $z^{-1} \in \mu_n$. □

Замечание 3.5.3 (Геометрическая интерпретация корней из единицы). Из формулы для $\varepsilon_k^{(n)}$ видно, что модули всех корней степени n из 1 равны единице, а аргументы равны $0, 2\pi/n, 4\pi/n, \ldots, 2(n-1)\pi/n$, то есть, образуют арифметическую прогрессию с разностью $2\pi/n$. Значит, на комплексной плоскости точки $\varepsilon_k^{(n)}$ лежат на окружности с центром в 0 и радиусом 1, и углы $\angle AOB$ для двух соседних точек A, B, равны $2\pi/n$. Из этого следует, что точки $\varepsilon_k^{(n)}$ лежат в вершинах правильного n-угольника с центром в 0. Кроме того, так как $\varepsilon_0^{(n)} = 1$, число 1 является одной из вершин этого n-угольника.

Замечание 3.5.4. Вернемся к уравнению $z^n = w$ для комплексного числа $w \neq 0$. Пусть z_0 — некоторое решение этого уравнения; тогда $z_0^n = w$ и, разделив первоначальное уравнение на это равенство, получаем $z^n/z_0^n = w/w = 1$, откуда $(z/z_0)^n = 1$, то есть, z/z_0 является корнем степени n из 1. Поэтому $z/z_0 = \varepsilon_k^{(n)}$ для некоторого k, и $z = z_0\varepsilon_k^{(n)}$. Таким образом, любое решение уравнения $z^n = w$ отличается от некоторого фиксированного решения z_0 домножением на корень степени n из 1.

Определение 3.5.5. Корень n-ой степени из 1 называется **первообразным**, если он не является корнем из 1 никакой меньшей, чем n, степени. Иными словами, z называется первообразным корнем степени n из 1, если $z^n = 1$ и $z^m \neq 1$ при $0 < m < n$.

Замечание 3.5.6. Заметим, что $\varepsilon_1^{(n)} = \cos(2\pi/n) + i\sin(2\pi/n)$ является первообразным корнем степени n из 1. Действительно, если $(\cos(2\pi/n) + i\sin(2\pi/n))^m = 1$ для некоторого $0 < m < n$, то по формуле Муавра $\cos(2\pi m/n) + i\sin(2\pi m/n) = 1$, откуда $2\pi m/n = 2\pi k$ для некоторого целого k. Получаем $m = kn$, то есть, m делится на n, что невозможно.

Предложение 3.5.7. *Пусть ε — корень степени n из 1. Равносильны:*

1. ε — первообразный корень;

2. все числа $1 = \varepsilon^0, \varepsilon^1, \varepsilon^2, \ldots, \varepsilon^{n-1}$ различны.

Доказательство. (2) $\Leftrightarrow$ (1): если $\varepsilon^m = 1$ для некоторого $0 < m < n$, то среди указанных чисел есть совпадающие. (1) $\Leftrightarrow$ (2): если $\varepsilon^k = \varepsilon^m$ для некоторых k, m, то можно считать, что $k > m$; тогда $\varepsilon^k/\varepsilon^m = \varepsilon^{k-m} = 1$. Из определения первообразного корня следует, что $k = m$. □

Предложение 3.5.8. *Пусть $n \geqslant 1$ — натуральное число, $0 \geqslant k \geqslant n-1$. Корень $\varepsilon_k^{(n)}$ степени n из 1 является первообразным тогда и только тогда, когда $\gcd(k, n) = 1$.*

Доказательство. Обозначим $\varepsilon = \varepsilon_1^{(n)}$. Нетрудно видеть, что $\varepsilon_k^{(n)} = \varepsilon^k$. Если $\gcd(k, n) = d > 1$, то $(\varepsilon_k^{(n)})^{n/d} = (\varepsilon^k)^{n/d} = \varepsilon^{kn/d} = (\varepsilon^n)^{k/d} = 1^{k/d} = 1$ (здесь важно, что k/d — целое число). Это значит, что $\varepsilon_k^{(n)}$ является корнем степени n/d из 1, и, поскольку $n/d < n$, не является первообразным корнем степени n из 1.

Обратно, если $\gcd(k, n) = 1$, покажем, что $\varepsilon_k^{(n)} = \varepsilon^k$ — первообразный корень степени n из 1. Действительно, предположим, что $(\varepsilon^k)^m = \varepsilon^{km} = 1$, где $0 < m < n$. Но $\varepsilon^{km} = (\cos(2\pi/n) + i\sin(2\pi/n))^{km} = (\cos(2\pi km/n) + i\sin(2\pi km/n)) = 1$, откуда $2\pi km/n = 2\pi t$ для некоторого целого t. Это означает, что $km = nt$, то есть, $n \mid km$. Но k и n взаимно просты; по свойству 3 взаимной простоты (2.3.3) теперь $n \mid m$ — противоречие с предположением $0 < m < n$. □

Следствие 3.5.9. *Количество первообразных корней степени n из 1 равно $\varphi(n)$.*

Доказательство. Следует из предложения 3.5.8 и определения функции Эйлера (2.11.1). □

3.6 Экспоненциальная форма записи комплексного числа

Литература: [F], гл. II, § 5, пп. 1–3.

Мы видели, что аргумент комплексного числа ведет себя подобно логарифму: аргумент произведения равен сумме аргументов. Это оправдывает следующее определение.

Определение 3.6.1. Пусть $z = a + bi$ — комплексное число. Положим $e^z = e^a(\cos(b) + i\sin(b))$.

Заметим, что основное свойство экспоненты выполняется при таком определении.

Предложение 3.6.2. $e^{z_1+z_2} = e^{z_1} \cdot e^{z_2}$.

Доказательство. Пусть $z_1 = a_1 + b_1 i$, $z_2 = a_2 + b_2 i$, тогда $z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)i$ и

$$\begin{aligned} e^{z_1} \cdot e^{z_2} &= e^{a_1}(\cos(b_1) + i\sin(b_1)e^{a_2}(\cos(b_2) + i\sin(b_2)) \\ &= e^{a_1+a_2}(\cos(b_1 + b_2) + i\sin(b_1 + b_2) \\ &= e^{z_1+z_2}. \end{aligned}$$

□

При этом $e^{i\varphi} = \cos(\varphi) + i\sin(\varphi)$; в частности, $e^{i\pi} = -1$. Теперь для любого ненулевого комплексного числа $z = r(\cos(\varphi) + i\sin(\varphi))$ можно записать $z = re^{i\varphi} = e^{\ln(r)+i\varphi}$. Эта запись называется **экспоненциальной формой записи комплексного числа**.

Попытаемся теперь определить обратную функцию — логарифм. Основное свойство логарифма должно сохраниться: логарифм должен быть обратной функцией к экспоненте. Заметим, что экспонента переводит сумму в произведение: $e^{a+b} = e^a \cdot e^b$. Поэтому логарифм должен переводить произведение в сумму: $\ln(ab) = \ln(a) + \ln(b)$. Таким образом, если определить логарифм вообще возможно, то для комплексного числа $z = r(\cos(\varphi) + i\sin(\varphi)) = r \cdot e^{i\varphi}$ должно выполняться $\ln(z) = \ln(r) + \ln(e^{i\varphi}) = \ln(r) + i\varphi$. Проблема состоит в том, что аргумент φ комплексного числа z определен не вполне однозначно, а с точностью до прибавления целого кратного числа 2π. Поэтому и логарифм должен быть определен не однозначно, а с точностью до целого кратного числа $2\pi i$. Часто через $\mathrm{Ln}(z)$ обозначают все множество значений, то есть, $\mathrm{Ln}(r(\cos(\varphi) + i\sin(\varphi))) = \{\ln(r) + i\varphi + 2\pi ik \mid k \in \mathbb{Z}\}$. Под записью $\ln(z)$ мы будем понимать *какое-нибудь* значение логарифма, то есть, какой-то элемент множества $\mathrm{Ln}(z)$. При этом из основного свойства экспоненты немедленно следует основное свойство логарифма: $\ln(z_1 z_2) = \ln(z_1) + \ln(z_2)$. Понимать это равенство, конечно, следует с точностью до слагаемого вида $2\pi ik$; например, $\ln(1) = 0$ и $\ln(-1) = \pi i$, но в то же время $\ln(1) = \ln((-1)\cdot(-1)) = \ln(-1) + \ln(-1) = \pi i + \pi i = 2\pi i$.

4 Кольцо многочленов

4.1 Определение и первые свойства

Литература: [F], гл. III, § 1, пп. 1–3; [K1], гл. 5, § 2, п. 1; [vdW], гл. 3, § 14.

Мы воспринимаем многочлен просто как последовательность его коэффициентов: то, что в привычной записи выглядит как $2x^3 - 5x + 4$, для нас является бесконечной последовательностью $(4, -5, 0, 2, 0, 0, \dots)$.

Определение 4.1.1. Пусть R — кольцо (коммутативное, ассоциативное, с 1). **Многочленом над** R (или **многочленом с коэффициентами из** R) называется бесконечная последовательностью элементов R, в которой все элементы, кроме конечного числа, равны нулю. Иными словами — это последовательностью $(a_0, a_1, a_2, \dots)$, где $a_i \in R$ со следующим свойством: существует натуральное $N \in \mathbb{N}$ такое, что $a_i = 0$ для всех $i > N$. Введем следующие операции сложения и умножения на множестве всех многочленов над R: пусть $a = (a_0, a_1, a_2, \dots)$, $b = (b_0, b_1, b_2, \dots)$. Положим $a + b = (a_0 + b_0, a_1 + b_1, a_2 + b_2, \dots)$, $ab = (a_0b_0, a_0b_1 + a_1b_0, a_0b_2 + a_1b_1 + a_2b_2, \dots)$. Формально: $(a + b)_k = a_k + b_k$, $(ab)_k = \sum_{i=0}^{k} a_i b_{k-i}$.

Проверим, что сумма многочленов действительно является многочленом, то есть, что начиная с некоторого места все коэффициенты в $a + b$ равны нулю. Поскольку a является многочленом, найдется натуральное M такое, что $a_i = 0$ при $i > M$. Поскольку b является многочленом, найдется натуральное N такое, что $b_i = 0$ при $i > N$. Но тогда при $i > \max(M, N)$ выполнено и $a_i = 0$, и $b_i = 0$, откуда $(a + b)_i = a_i + b_i = 0$ для всех таких i.

Чуть сложнее строго показать, что произведение многочленов является многочленом. Пусть снова $a_i = 0$ при всех $i > M$, и $b_j = 0$ при всех $j > N$. Мы утверждаем, что при $k > M + N$ коэффициент $(ab)_k$ равен нулю. Действительно, по определению

$$(ab)_k = \sum_{i+j=k} a_i b_j.$$

Заметим, что при $i + j > M + N$ выполнено хотя бы одно из неравенств $i > M$, $j > N$ (иначе, если $i \leqslant M$ и $j \leqslant N$, то $i + j \leqslant M + N$ — противоречие). Значит, каждое слагаемое в сумме, стоящей в правой части, равно нулю, ибо $a_i = 0$ при $i > M$, а $b_j = 0$ при $j > N$. Поэтому и вся сумма $(ab)_k$ равна нулю.

Множество всех многочленов над R с определенными таким образом операциями обозначим через $R[x]$.

Замечание 4.1.2. В обозначении $R[x]$ буква x пока не несет никакого смысла; чуть ниже мы узнаем, что такое каноническая запись многочлена, и x станет вполне определенным элементом $R[x]$. Тем не менее, на ее место можно выбрать любую другую букву.

Теорема 4.1.3. *$R[x]$ является кольцом (ассоциативным, коммутативным, с 1).*

Доказательство. Необходимо проверить восемь аксиом из определения кольца (2.8.1). Сложение в $R[x]$ происходит покомпонентно, поэтому первые четыре аксиомы, отражающие

свойства сложения (ассоциативность и коммутативность, наличие нейтрального элемента и противоположных) сразу следуют из соответствующих свойств сложения в кольце R. Отметим лишь, что роль нейтрального элемента по сложению играет последовательность $(0,0,0,\dots)$, а роль противоположной к последовательности $(a_0,a_1,a_2,\dots)$ играет последовательность $(-a_0,-a_1,-a_2,\dots)$.

Ассоциативность умножения: пусть $a=(a_0,a_1,\dots)$, $b=(b_0,b_1,\dots)$, $c=(c_0,c_1,\dots)$ — элементы $R[x]$. Тогда

$$\begin{aligned}((ab)c)_l &= \sum_{k=0}^{l}(ab)_k c_{l-k} = \sum_{k=0}^{l}\sum_{i=0}^{k} a_i b_{k-i} c_{l-k},\\ (a(bc))_l &= \sum_{i=0}^{l} a_i (bc)_{l-i} = \sum_{i=0}^{l} a_i \sum_{j=0}^{l-i} b_j c_{l-i-j}\\ &= \sum_{i=0}^{l} a_i \sum_{i+j=i}^{l} b_j c_{l-i-j}.\end{aligned}$$

Сделав замену $k=i+j$ в последней сумме, получаем $(a(bc))_l=\sum_{i=0}^{l} a_i \sum_{k=i}^{l} b_{k-i}c_{l-k}$. Теперь видно, что суммы в выражениях для $((ab)c)_l$ и $(a(bc))_l$ равны; можно считать, что суммирования производятся по парам (i,k) таким, что $0\leqslant i\leqslant k\leqslant l$.

Покажем, что элемент $e=(1,0,0,\dots)$ является нейтральным по умножению. Действительно, $(ae)_k=\sum_{i=0}^{k} a_i e_{k-i}=a_k$ и $(ea)_k=\sum_{i=0}^{k} e_i a_{k-i}=a_k$. Умножение коммутативно: $(ab)_k=\sum_{i=0}^{k} a_i b_{k-i}$, $(ba)_k=\sum_{j=0}^{k} b_j a_{k-j}=\sum_{k-j=0}^{k} b_{k-(k-j)}a_{k-j}$, и осталось сделать замену $i=k-j$.

Наконец, проверим дистрибутивность:

$$\begin{aligned}((a+b)c)_k &= \sum_{i=0}^{k}(a+b)_i c_{k-i}\\ &= \sum_{i=0}^{k}(a_i+b_i)c_{k-i}\\ &= \sum_{i=0}^{k}(a_i c_{k-i}+b_i c_{k-i})\\ &= \sum_{i=0}^{k}(a_i c_{k-i})+\sum_{i=0}^{k}(b_i c_{k-i})\\ &= (ac)_k+(bc)_k.\end{aligned}$$

□

Замечание 4.1.4. Можно считать, что кольцо R является подмножеством кольца $R[x]$; действительно, каждому элементу $a\in R$ соответствует многочлен $(a,0,0,\dots)$, и операции на таких элементах в $R[x]$ соответствуют операциям в R. В силу этого, многочлен $(0,0,0,\dots)$, являющийся нейтральным элементом по сложению кольца $R[x]$, мы обозначаем просто через 0, а многочлен

$e = (1, 0, 0, \dots)$ — через 1. Поэтому мы часто будем писать a вместо многочлена $(a, 0, 0, \dots)$ для элементов $a \in R$. При этом, как нетрудно видеть, $a \cdot (b_0, b_1, b_2, \dots) = (ab_0, ab_1, ab_2, \dots)$.

Замечание 4.1.5. Как и в других кольцах, для натурального n и $f \in R[x]$ мы обозначаем через f^n многочлен $\underbrace{f \cdot \dots \cdot f}_{n}$; если $n = 0$, положим $f^0 = 1 \in R[x]$.

Определение 4.1.6. Пусть $a = (a_0, a_1, a_2, \dots)$ — многочлен над кольцом R. **Степенью** многочлена a называется наибольшее d такое, что $a_d \neq 0$. Удобно считать, что степень нулевого многочлена $(0, 0, \dots)$ равна $-\infty$. Если же $a \neq 0$, то степень a — натуральное число. Обозначение: $d = \deg(f)$. Заметим, что многочлены степени 0 — это в точности ненулевые константы из R.

Замечание 4.1.7. Обозначим через x элемент $(0, 1, 0, 0, \dots) \in R[x]$. Нетрудно видеть, что $x^2 = (0, 0, 1, 0, 0, \dots)$, и вообще $x^n = (\underbrace{0, \dots, 0}_{n}, 1, 0, 0, \dots)$ для всякого натурального n. С учетом замечания 4.1.4 любой элемент $a = (a_0, a_1, a_2, \dots) \in R[x]$ можно записать как

$$\begin{aligned} a &= (a_0, a_1, a_2, a_3, \dots) \\ &= (a_0, 0, 0, 0, \dots) + (0, a_1, 0, 0, \dots) + (0, 0, a_2, 0, \dots) + \dots \\ &= a_0 \cdot (1, 0, 0, 0, \dots) + a_1 \cdot (0, 1, 0, 0, \dots) + a_2 \cdot (0, 0, 1, 0, \dots) + \dots \\ &= a_0 + a_1 x + a_2 x^2 + \dots. \end{aligned}$$

Конечно, в полученной сумме лишь конечное число ненулевых слагаемых; если $\deg(a) = d$, можно записать $a = a_0 + a_1 x + \dots + a_d x^d$. Такая запись называется **канонической записью многочлена**.

Теорема 4.1.8. *Пусть R — область целостности. Тогда $\deg(f \cdot g) = \deg(f) + \deg(g)$ для любых $f, g \in R[x]$.*

Доказательство. Пусть $m = \deg(f)$, $n = \deg(g)$. Запишем $f = a_0 + a_1 x + \dots + a_m x^m$, $g = b_0 + b_1 x + \dots + b_n x^n$. По определению степени имеем $a_m \neq 0$ и $b_n \neq 0$. Нетрудно видеть, что $fg = a_0 b_0 + \dots + a_m b_n x^{m+n}$ и $a_m b_n \neq 0$, поскольку R — область целостности. □

Замечание 4.1.9. Заметим, что теорема верна и для случая $f = 0$ или $g = 0$ за счет нашего соглашения $\deg(0) = -\infty$.

Следствие 4.1.10. *Если R — область целостности, то $R[x]$ — область целостности.*

Доказательство. Пусть $fg = 0$; предположим, что $f \neq 0$, $g \neq 0$, тогда $\deg(f)$ и $\deg(g)$ — натуральные числа, поэтому и $\deg(fg)$ — натуральное число. □

Следствие 4.1.11. *Пусть R — область целостности. Многочлен $f \in R[x]$ является обратимым тогда и только тогда, когда он имеет степень 0, то есть является элементом $f = r \in R$, и r обратим в R. Иными словами, $R[x]^* = R^*$.*

Доказательство. Пусть $f \in R[x]^*$ и $g \in R[x]$ — обратный элемент к f: $fg = 1$. При этом $\deg(f) + \deg(g) = \deg(fg) = \deg(1) = 0$. Если одна из степеней f, g равна $-\infty$, то и $\deg(fg)$ равнялась бы $-\infty$; поэтому оба числа $\deg(f)$, $\deg(g)$ натуральны и, следовательно, равны 0. Значит, $f, g \in R$ — константы, произведение которых равно $1 \in R$. Поэтому $f \in R^*$.

Обратно, если $f \in R^*$, обозначим через $g \in R^*$ обратный элемент к f в R. Тогда $fg = 1$, и если рассмотреть f, g как многочлены, получим, что $f \in R[x]^*$. □

4.2 Делимость в кольце многочленов

Литература: [F], гл. VI, § 1, п. 1–2; [K1], гл. 5, § 2, п. 3; § 3, п. 1; [vdW], гл. 3, § 14.

Начиная с этого места мы считаем, что кольцо R является областью целостности (тогда по теореме 4.1.10 и $R[x]$ является областью целостности).

Сейчас мы перенесем основные определения из раздела 2.1 на случай кольца многочленов.

Определение 4.2.1. Пусть $f, g \in R[x]$. Говорят, что многочлен g **делит** многочлен f (или что f **делится на** g), если $f = gp$ для некоторого $p \in R[x]$. Обозначение: $g \mid f$.

Предложение 4.2.2 (Свойства делимости в кольце многочленов). *Пусть* $f, g, h \in R[x]$. *Тогда*

1. $f \mid f$ *и* $f \mid 1$;
2. *если* $h \mid f$, $h \mid g$, *то* $h \mid f + g$;
3. *если* $h \mid f$, *то* $h \mid fg$;
4. *если* $h \mid g$, $g \mid f$, *то* $h \mid f$.

Доказательство. 1. $f = f \cdot 1 = 1 \cdot f$.

2. если $f = hp$, $g = hq$, то $f + g = h(p + q)$.

3. если $f = hp$, то $fg = hgp$.

4. если $g = hp$, $f = gq$, то $f = hpq$.

□

Определение 4.2.3. Два элемента $f, g \in R[x]$ называются **ассоциированными**, если $g \mid f$ и $f \mid g$.

Предложение 4.2.4. *Ассоциированность является отношением эквивалентности.*

Доказательство. Очевидно. □

Предложение 4.2.5. $f, g \in R[x]$ *ассоциированы тогда и только тогда, когда* $f = cg$ *для некоторой обратимой константы* $c \in R^*$.

Доказательство. Если $f = cg$ для $c \in R^*$, то $g \mid f$ и $g = c^{-1}f$, поэтому $f \mid g$. Обратно, из $g \mid f$ следует, что $f = gp$, а из $f \mid g$ следует, что $g = fq$. Поэтому $f = gp = fqp$, откуда $f(1 - pq) = 0$. Заметим, что $R[x]$ — область целостности, поэтому $f = 0$ или $1 - pq = 0$. Если $f = 0$, то и $g = 0$, и доказывать нечего. Иначе получаем, что $1 = pq$, откуда $p \in R[x]^* = R^*$. Значит, p — ненулевая константа, что и требовалось доказать. □

Теорема 4.2.6 (О делении с остатком в кольце многочленов). *Пусть R — область целостности, $f, g \in R[x]$, $g \neq 0$, и старший коэффициент многочлена g обратим. Существуют единственные многочлены $h, r \in R[x]$ такие, что $f = gh + r$ и $\deg(r) < \deg(g)$.*

Доказательство. Сначала докажем существование индукцией по $\deg(f)$. Если $\deg(f) < \deg(g)$, можно записать $f = g \cdot 0 + f$, то есть, взять $h = 0$ и $r = f$.

Пусть теперь $\deg(f) \geqslant \deg(g)$. Запишем $f = a_m x^m + \ldots$, $g = b_n x^n + \ldots$, где $m = \deg(f)$, $n = \deg(g)$. Таким образом, $a_m \neq 0$, $b_n \neq 0$ и $m \geqslant n$. Более того, по нашему предположению коэффициент b_n обратим в R. Рассмотрим многочлен $f_0 = f - g \cdot a_m b_n^{-1} x^{m-n}$. Степень g равна n, степень монома $a_m b_n^{-1} x^{m-n}$ равна $m - n$, поэтому степень многочлена $g \cdot a_m b_n^{-1} x^{m-n}$ равна m, как и степень f. Значит, степень f_0 не превосходит m.

Посмотрим на коэффициент многочлена f_0 при x^m. Он равен разности коэффициентов f и $g \cdot a_m b_n^{-1} x^{m-n}$ при x^m, то есть, $a_m - b_n \cdot a_m b_n^{-1} = 0$. Значит, степень f_0 строго меньше $m = \deg(f)$. Поэтому к f_0 можно применить предположение индукции и записать $f_0 = gh_0 + r_0$, где $\deg(r) < \deg(g)$. Тогда $f = f_0 + g \cdot a_m b_n^{-1} x^{m-n} = gh_0 + r_0 + g \cdot a_m b_n^{-1} x^{m-n} = g(h_0 + a_m b_n^{-1} x^{m-n}) + r_0$. Возьмем $h = h_0 + a_m b_n^{-1} x^{m-n}$ и $r = r_0$; тогда $f = gh + r$ и все еще $\deg(r) = \deg(r_0) < \deg(g)$.

Осталось доказать единственность: предположим, что $f = gh + r$ и $f = g\widetilde{h} + \widetilde{r}$. Тогда $g(h - \widetilde{h}) = \widetilde{r} - r$. Степени многочленов r и $\widetilde{r}$ меньше степени g, поэтому степень правой части равенства меньше степени g; в то же время, степень правой части равна сумме степеней g и $h - \widetilde{h}$. Такое возможно только если степень $h - \widetilde{h}$ равна $-\infty$, то есть, $h = \widetilde{h}$, откуда и $r = \widetilde{r}$. □

Замечание 4.2.7. Заметим, что условие обратимости старшего коэффициента многочлена g автоматически выполняется, если R — поле. Таким образом, над полем можно делить любой многочлен на любой ненулевой.

4.3 Многочлен как функция

Литература: [F], гл. III, § 1, пп. 4–7; [K1], гл. 6, § 1, п. 1–2; [vdW], гл. 5, § 28.

Определение 4.3.1. Пусть $f = a_0 + a_1 x + \cdots + a_n x^n \in R[x]$, $c \in R$. **Значением** многочлена f в точке c называется $f(c) = a_0 + a_1 c + \cdots + a_n c^n = \sum_{i=0}^{\infty} a_i c^i \in R$.

Замечание 4.3.2. Таким образом, с каждым многочленом $f \in R[x]$ связано отображение $\widetilde{f}\colon R \to R$, $c \mapsto f(c)$. Мы называем это отображение **полиномиальной функцией**, заданной многочленом f.

Предложение 4.3.3. *Для любых $f, g \in R[x]$, $c \in R$, выполнено*

1. $(f + g)(c) = f(c) + g(c)$;

2. $(fg)(c) = f(c) \cdot g(c)$;

3. *если* $f = r \in R$, *то* $f(c) = r$

Доказательство. Пусть $f = \sum_{i=0}^{\infty} a_i x^i$, $g = \sum_{i=0}^{\infty} b_i x^i$.

1. $f + g = \sum_{i=0}^{\infty}(a_i + b_i)x^i$, поэтому $(f + g)(c) = \sum_{i=0}^{\infty}(a_i + b_i)c^i = \sum_{i=0}^{\infty}(a_i c^i) + \sum_{i=0}^{\infty}(b_i c^i) = f(c) + g(c)$.

2. $fg = \sum_{m=0}^{\infty}\sum_{i+j=m}(a_i b_j x^m)$, поэтому $f(c)g(c) = (\sum_{i=0}^{\infty} a_i c^i)(\sum_{j=0}^{\infty} b_j c^j) = \sum_{i,j=0}^{\infty}(a_i b_j c^{i+j}) = \sum_{m=0}^{\infty}\sum_{i+j=m}(a_i b_j c^m) = (fg)(c)$.

3. $f(c) = r + 0 \cdot c + \cdots = r$.

□

Определение 4.3.4. Пусть $f \in R[x]$, $c \in R$. Говорят, что c является **корнем** многочлена f, если $f(c) = 0$.

Теорема 4.3.5 (Лемма Безу). *Пусть* $f \in R[x]$, $c \in R$. *Многочлен* f *делится на многочлен* $(x - c)$ *тогда и только тогда, когда* c *является корнем* f. *Более точно, остаток от деления многочлена* f *на* $(x - c)$ *равен* $f(c)$.

Доказательство. Поделим f на $x - c$ с остатком (заметим, что это можно сделать, поскольку старший коэффициент многочлена $x - c$ обратим). $f = (x - c)h + r$. Заметим, что $\deg(r) < \deg(x - c) = 1$, поэтому $r \in R$ — константа. Подставим c в обе части этого равенства:

$$f(c) = ((x - c)h + r)(c) = ((x - c)h)(c) + r(c) = 0 \cdot h(c) + r = r.$$

Если f делится на $x - c$, то $r = 0$, и потому $f(c) = 0$. Обратно, если $f(c) = 0$, то и $r = 0$, и потому f делится на $(x - c)$. □

Предложение 4.3.6. *Пусть* $f \in R[x]$, $f \neq 0$. *Тогда* f *можно записать в виде* $f = (x - c_1)\dots(x - c_m)h$, *где* $c_1, \dots, c_m \in R$ — *все корни* f *(возможно, с повторениями), а* $h \in R[x]$ — *многочлен, у которого нет корней в кольце* R.

Доказательство. Доказываем индукцией по $\deg(f)$. База: $\deg(f) = 0$, то есть, f — ненулевая константа. Это многочлен без корней, поэтому можно взять $m = 0$ и $h = f$. Теперь пусть $\deg(f) > 0$. Если у f нет корней, опять можно взять $m = 0$, $h = f$. Если же c — корень f, то (по теореме 4.3.5) $f = (x - c)f_1$, $\deg(f_1) < \deg(f)$, и к f_1 можно применить предположение индукции. Поэтому f_1 имеет нужное разложение, и, дописывая к нему скобку $(x - c)$, получаем разложение для f.

Теперь мы получили, что $f = (x - c_1)\dots(x - c_m)h$ для некоторых $c_1, \dots, c_m \in R$ и многочлена $h \in R[x]$ без корней. Очевидно, что каждый c_i, $i = 1, \dots, m$, является корнем f. Осталось показать, что среди $c_1, \dots, c_m$ встречаются все корни f. Если c — некоторый корень f, то $0 = f(c) = (c - c_1)\dots(c - c_m)h(c)$. При этом $h(c) \neq 0$, поскольку у h нет корней, значит (поскольку R — область целостности), одна из скобок вида $(c - c_i)$ равна 0, поэтому c содержится среди $c_1, \dots, c_m$. □

Следствие 4.3.7. *Число различных корней ненулевого многочлена над областью целостности не превосходит его степени.*

Доказательство. Посмотрим на разложение из предложения 4.3.6. Все корни c многочлена $f \in R[x]$ содержатся среди $c_1, \ldots, c_m$, поэтому их число не больше m, а $m = \deg(f) - \deg(h) \leqslant \deg(f)$. □

Позже (см. замечание 4.5.3) мы уточним это следствие с помощью понятия *кратности* корня.

Определение 4.3.8. Пусть $f, g \in R[x]$ — многочлены над областью целостности R. Говорят, что многочлен f **функционально равен** многочлену g, если $f(c) = g(c)$ для любого $c \in R$. Иными словами, многочлены функционально равны, если задаваемые ими функции равны: $\widetilde{f} = \widetilde{g}$ (см. замечание 4.3.2). Обычное равенство многочленов при этом иногда называют **формальным равенством**: многочлены f и g формально равны, если $f = g$.

Пример 4.3.9. Пусть $R = \mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}$. Рассмотрим многочлен $f = x^2 - x$. Заметим, что $f(\overline{0}) = f(\overline{1}) = \overline{0}$. Поэтому многочлен f функционально равен многочлену 0, но, конечно, $f \neq 0$. Этот пример обобщается на поле $R = \mathbb{Z}/p\mathbb{Z}$: достаточно взять $f = x^p - x$ и вспомнить малую теорему Ферма (следствие 2.12.2).

Замечание 4.3.10. Очевидно, что из формального равенства многочленов следует функциональное: если $f = g$, то $f(c) = g(c)$ для любого $c \in R$.

Теорема 4.3.11. *Если область целостности R бесконечна, то из функционального равенства многочленов над R следует их формальное равенство.*

Доказательство. Пусть $f, g \in R[x]$ и $f(c) = g(c)$ для всех $c \in R$. Посмотрим на разность $h = f - g \in R[x]$. Для любого $c \in R$ выполнено $h(c) = f(c) - g(c) = 0$, поэтому c — корень h. Если h ненулевой, то по следствию 4.3.7 число корней h не превосходит его степени; с другой стороны, как мы только что видели, любой элемент бесконечного кольца R является корнем h — противоречие. Значит, $h = 0$, поэтому и $f = g$. □

4.4 Многочлены над $\mathbb{R}$ и $\mathbb{C}$

Литература: [F], гл. III, § 1, п. 8; гл. VI, § 1, п. 7; [K1], гл. 6, § 3, п. 1; § 4, п. 1.

Сейчас мы уточним разложение из предложения 4.3.6 для случая многочленов над полями $\mathbb{R}$ и $\mathbb{C}$.

Определение 4.4.1. Поле k называется **алгебраически замкнутым**, если у любого многочлена $f \in k[x]$ степени выше нулевой имеется корень в k.

Пример 4.4.2. Поле комплексных чисел $\mathbb{C}$ является алгебраически замкнутым. Это утверждение называется **основной теоремой алгебры**; в нашем курсе мы будем пользоваться им без доказательства. С другой стороны, поле вещественных чисел $\mathbb{R}$ не алгебраически замкнуто: например, у многочлена $x^2 + 1$ нет вещественных корней.

Теорема 4.4.3 (Разложение многочлена над алгебраически замкнутым полем). *Пусть* k — *алгебраически замкнутое поле. Тогда любой ненулевой многочлен* $f \in k[x]$ *представляется в виде* $f = c_0(x - c_1)\dots(x - c_n)$, *где* $c_0, c_1, \dots, c_n \in k$.

Доказательство. По следствию 4.3.6 можно записать $f = (x - c_1)\dots(x - c_m)h$, где у $h \in k[x]$ нет корней; по определению алгебраической замкнутости из этого следует, что $\deg(h) \leqslant 0$, поэтому $h = c_0 \in k$ — константа. □

Теорема 4.4.4 (Разложение многочлена над полем вещественных чисел). *Пусть* $f \in \mathbb{R}[x]$, $f \neq 0$. *Тогда* f *можно представить в виде* $f = c_0(x - c_1)\dots(x - c_s)(x^2 + a_1x + b_1)\dots(x^2 + a_rx + b_r)$, *где* $c_0, c_1, \dots, c_s, a_1, \dots, a_r, b_1, \dots, b_r \in \mathbb{R}$ *и* $a_i^2 - 4b_i < 0$ *для всех* $i = 1, \dots, r$.

Доказательство. Доказываем индукцией по степени f. Если $\deg(f) = 0$, то $f = c_0$, $s = 0$, $r = 0$. Пусть теперь $\deg(f) > 0$. Рассмотрим f как многочлен над комплексными числами. По основной теореме алгебры у f есть корень $\lambda \in \mathbb{C}$.

Если $\lambda \in \mathbb{R}$, то f делится на $x - \lambda$, и можно записать $f = (x - \lambda)g$. При этом $\deg(g) < \deg(f)$, и по предположению индукции g раскладывается в произведение нужного вида; дописывая к этому разложению скобку $(x - \lambda)$, получаем и разложение для f.

Если же $\lambda \in \mathbb{C}\backslash\mathbb{R}$, рассмотрим $f(\overline{\lambda})$:

$$\begin{aligned} f(\overline{\lambda}) &= a_0 + a_1\overline{\lambda} + \dots + a_n\overline{\lambda}^n \\ &= \overline{a_0} + \overline{a_1\lambda} + \dots + \overline{a_n\lambda^n} \\ &= \overline{f(\lambda)} \\ &= \overline{0} \\ &= 0. \end{aligned}$$

Значит, и λ, и $\overline{\lambda}$ являются корнями f. Поэтому f делится на $(x - \lambda)(x - \overline{\lambda})$. Запишем $f = (x-\lambda)(x-\overline{\lambda})g$. Заметим, что $(x-\lambda)(x-\overline{\lambda}) = x^2-(\lambda+\overline{\lambda})x+\lambda\overline{\lambda} = x^2-(2\operatorname{Re}(\lambda))+|\lambda|^2$ — квадратичный многочлен с вещественными коэффициентами. Поэтому коэффициенты многочлены g также вещественны, $\deg(g) < \deg(f)$ и можно применить предположение индукции. Кроме того, дискриминант квадратичного многочлена $(x - \lambda)(x - \overline{\lambda})$ меньше 0, поскольку у него нет вещественных корней. Поэтому нужное разложение многочлена f получается приписыванием к разложению g указанного квадратичного многочлена. □

4.5 Кратные корни и производная

Литература: [F], гл. VI, § 2, пп. 1, 3; [K1], гл. 6, § 1, п. 3–4; [vdW], гл. 5, §§ 27–28.

Мы возвращаемся к рассмотрению многочленов над произвольной областью целостности R.

Определение 4.5.1. Пусть $f \in R[x]$, $c \in R$. Говорят, что c является корнем многочлена f **кратности** m, если f делится на $(x - c)^m$, но не делится на $(x - c)^{m+1}$. Корень f кратности 1 называют **простым корнем** f, а корень кратности > 1 — **кратным корнем** f.

Лемма 4.5.2. *Пусть* $f \in R[x]$, $c \in R$, $m \geqslant 1$. *Элемент* c *является корнем* f *кратности* m *тогда и только тогда, когда* f *можно представить в виде* $f = (x-c)^m \cdot g$, *где многочлен* $g \in R[x]$ *таков, что* $g(c) \neq 0$.

Доказательство. Если c — корень f кратности m, то $f = (x-c)^m \cdot g$ для некоторого $g \in R[x]$. Если $g(c) = 0$, то по теореме Безу g делится на $(x-c)$, поэтому $g = (x-c)h$ и $f = (x-c)^{m+1}h$, то есть, f делится на $(x-c)^{m+1}$ — противоречие.

Обратно, если $f = (x-c)^m \cdot g$ и $g(c) \neq 0$, то f делится на $(x-c)^m$. Если при этом f делится на $(x-c)^{m+1}$, то $f = (x-c)^{m+1} \cdot h$. Сравнивая два выражения для f,получаем $(x-c)^m \cdot g = (x-c)^{m+1} \cdot h$, откуда $(x-c)^m(g-(x-c)h) = 0$. Так как $R[x]$ — область целостности, получаем $g-(x-c)h = 0$, откуда $g = (x-c)h$ и $g(c) = 0$ — противоречие. □

Замечание 4.5.3. Таким образом, если в выражении для многочлена f из следствия 4.3.6 собрать скобки, соответствующие одинаковым корням, вместе, то скобка $(x-c)$ окажется с показателем, в точности равным кратности c как корня f. В частности, из этого немедленно следует, что сумма кратностей корней многочлена f не превосходит его степени.

Определение 4.5.4. Пусть $f \in R[x]$, $f = \sum_{s=0}^{\infty} a_s x^s$. **Производным многочленом** от многочлена f (или его **производной**) называется многочлен $f' = \sum_{s=1}^{\infty} s a_s x^{s-1}$.

Замечание 4.5.5. Напомним, что для элемента $c \in R$ и натурального числа n можно положить $nc = \underbrace{c + \cdots + c}_{n} = \underbrace{(1 + \cdots + 1)}_{n} \cdot c \in R$.

Предложение 4.5.6 (Свойства производной). *Пусть* $f, g \in R[x]$, $c \in R$, $m \geqslant 1$. *Тогда*

1. $(f+g)' = f' + g'$ *(***аддитивность***);*
2. $(cf)' = cf'$*;*
3. $(fg)' = f'g + fg'$ *(***тождество Лейбница***);*
4. $(g^m)' = mg^{m-1}g'$.

Доказательство. Пусть $f = \sum_{s=0}^{\infty} a_s x^s$, $g = \sum_{s=0}^{\infty} b_s x^s$.

1. $f + g = \sum_{s=0}^{\infty} (a_s + b_s)x^s$, поэтому
$$(f+g)' = \sum_{s=1}^{\infty} s(a_s + b_s)x^{s-1} = \sum_{s=1}^{\infty} (s a_s x^{s-1}) + \sum_{s=1}^{\infty} (s b_s x^{s-1}) = f' + g'.$$
2. $cf = \sum_{s=0}^{\infty} c a_s x^s$, поэтому $(cf)' = \sum_{s=1}^{\infty} s c a_s x^{s-1} = c \sum_{s=1}^{\infty} s a_s x^{s-1} = cf'$.
3. Докажем сначала тождество Лейбница для *мономов* (многочленов вида ax^n): если $f = ax^n$, $g = bx^m$, то $fg = abx^{m+n}$ и $(fg)' = (m+n)abx^{m+n-1}$, в то время как $f' = nax^{n-1}$, $g' = mbx^{m-1}$, откуда $f'g + fg' = nabx^{m+n-1} + mabx^{m+n-1} = (fg)'$. Пусть теперь

f, g произвольны. Запишем их в виде суммы мономов (это можно сделать с любым многочленом): $f = f_1 + \cdots + f_r$, $g = g_1 + \cdots + g_s$. Тогда

$$\begin{aligned} fg &= (f_1 + \cdots + f_r)(g_1 + \cdots + g_s) \\ &= \sum_{\substack{1 \leqslant i \leqslant r \\ 1 \leqslant j \leqslant s}} f_i g_j. \end{aligned}$$

Возьмем производную и воспользуемся уже доказанным свойством аддитивности. Кроме того, заметим, что мы доказали тождество Лейбница для мономов f_i и g_j, поэтому $(f_i g_j)' = f_i' g_j + f_i g_j'$. Получаем:

$$\begin{aligned} (fg)' &= \sum_{\substack{1 \leqslant i \leqslant r \\ 1 \leqslant j \leqslant s}} (f_i g_j)' \\ &= \sum_{\substack{1 \leqslant i \leqslant r \\ 1 \leqslant j \leqslant s}} (f_i' g_j + f_i g_j') \\ &= \sum_{\substack{1 \leqslant i \leqslant r \\ 1 \leqslant j \leqslant s}} (f_i' g_j) + \sum_{\substack{1 \leqslant i \leqslant r \\ 1 \leqslant j \leqslant s}} (f_i g_j') \\ &= (f_1' + \cdots + f_r')(g_1 + \cdots + g_s) + (f_1 + \cdots + f_r)(g_1' + \cdots + g_s') \\ &= (f_1 + \cdots + f_r)'(g_1 + \cdots + g_s) + (f_1 + \cdots + f_r)(g_1 + \cdots + g_s)' \\ &= f'g + fg' \end{aligned}$$

4. Проведем индукцию по m. Для $m = 1$ получаем тождество $g' = g'$. Пусть теперь $m > 1$, тогда $(g^m)' = (g \cdot g^{m-1})' = g' \cdot g^{m-1} + g \cdot (g^{m-1})' = g^{m-1} g' + g \cdot (m-1) g^{m-2} g' = m g^{m-1} g'$, что и требовалось.

□

Предложение 4.5.7 (Связь между корнями многочлена и его производной). *Пусть* $f \in R[x]$, $c \in R$. *Элемент* c *является кратным корнем многочлена* f *тогда и только тогда, когда* c *является корнем и* f, *и* f'.

Доказательство. Если c — кратный корень f, то f делится на $(x-c)^2$. Запишем $f = (x-c)^2 \cdot g$ и посчитаем производную от обеих частей: $f' = ((x-c)^2 \cdot g)' = ((x-c)^2)' g + (x-c)^2 g' = 2(x-c)g + (x-c)^2 g' = (x-c)(2g + (x-c)g')$. Значит, c является и корнем f'.

Обратно, если c корень f и f', запишем $f = (x-c)g$ и $f' = (x-c)h$. При этом $(x-c)h = f' = ((x-c)g)' = (x-c)'g + (x-c)g' = g + (x-c)g'$. Значит, $(x-c)(h-g') = g$, откуда $f = (x-c)g = (x-c)^2(h-g')$, и c — кратный корень f. □

Для исследования более тонких вопросов, касающихся кратностей корней, нам удобно будет предположить, что основное кольцо R является полем.

Определение 4.5.8. Пусть k — поле. **Характеристикой** поля k называется наименьшее число p такое, что $\underbrace{1+\cdots+1}_{p} = 0$ в k, если оно существует; в противном случае говорят, что характеристика k равна 0. Обозначение: $\mathrm{char}(k) = p$.

Примеры 4.5.9. Поля $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ имеют характеристику 0: никакая сумма единиц не равна нулю. Поле $\mathbb{Z}/p\mathbb{Z}$ имеет характеристику p: действительно, $\underbrace{\overline{1}+\cdots+\overline{1}}_{m} = \overline{m}$, причем $\overline{p} = \overline{0}$ и $\overline{m} \neq \overline{0}$ при $1 \leqslant m \leqslant p-1$.

Лемма 4.5.10. *Характеристика поля равна* 0 *или простому числу.*

Доказательство. Заметим, что характеристика поля не может равняться 1, поскольку в поле $1 \neq 0$ (см. определение 2.8.7). Если же $\mathrm{char}(k) = ab$ — составное число $(a, b > 1)$, заметим, что $0 = \underbrace{1+\cdots+1}_{ab} = (\underbrace{1+\cdots+1}_{a})(\underbrace{1+\cdots+1}_{b})$. Поле является областью целостности, поэтому одна из двух получившихся скобок равна 0, но $a, b < ab$, что противоречит минимальности в определении характеристики. □

Теорема 4.5.11. *Пусть* $f \in k[x]$, $c \in k$ — *корень* f, $m \geqslant 1$, *и характеристика поля* k *равна нулю. Если* c *является корнем* f *кратности* m, *то* c *является корнем* f' *кратности* $m-1$. *Обратно, если* c — *корень* f' *кратности* $m-1$, *то* c — *корень* f *кратности* m.

Доказательство. Пусть c — корень f кратности m; по лемме 4.5.2 это означает, что $f = (x-c)^m g$ и $g(c) \neq 0$. Возьмем производную: $f' = (x-c)^m g' + m(x-c)^{m-1} g = (x-c)^{m-1}((x-c)g' + mg)$. Мы утверждаем, что многочлен $(x-c)g' + mg$ в точке c не равен нулю. Действительно, его значение в точке c равно $0 \cdot g'(c) + mg(c) = mg(c)$. При этом $g(c) \neq 0$ и характеристика поля k равна нулю, поэтому $m \neq 0$. Снова применяя лемму 4.5.2, получаем, что c — корень f' кратности $m-1$.

Обратно, если c — корень f' кратности $m-1$, пусть n — кратность c как корня f. По условию c является корнем f, поэтому $n \geqslant 1$. По уже доказанному теперь c является корнем f' кратности $n-1$, поэтому $n-1 = m-1$, откуда $n = m$, что и требовалось. □

Замечание 4.5.12. Теорема 4.5.11 не выполняется для полей положительной характеристики. Пусть, например, $k = \mathbb{Z}/p\mathbb{Z}$ — поле из p элементов. Рассмотрим многочлен $f = x^p(x-1) = x^{p+1} - x^p$. Элемент $c = 0$ является корнем многочлена f кратности p, но у его производной $f' = (p+1)x^p - px^{p-1} = x^p$ элемент c снова является корнем кратности p.

Теорема 4.5.13. *Пусть* $f \in k[x]$, $c \in k$, $m > 1$, *и характеристика поля* k *равна нулю. Элемент* c *является корнем* f *кратности* m *тогда и только тогда, когда* $f(c) = f'(c) = \cdots = f^{(m-1)}(c) = 0$ *и* $f^{(m)}(c) \neq 0$.

Доказательство. Если c является корнем f кратности m, то c является корнем f' кратности $m-1$, ..., корнем $f^{(m-1)}$ кратности 1, и не является корнем $f^{(m)}$.

Обратно, если $f(c) = f'(c) = \cdots = f^{(m-1)}(c) = 0$ и $f^{(m)}(c) \neq 0$, воспользуемся индукцией по m. База $m = 1$: $f(c) = 0$ и $f'(c) \neq 0$ — по теореме 4.5.7 из этого следует, что c — простой корень

f. Многочлен f' таков, что он и его первые $m-2$ производные имеют корень c, а $(m-1)$-ая производная не равна нулю в точке c. По предположению индукции c — корень f' кратности $m-1$. По теореме 4.5.11 тогда c — корень f кратности m, что и требовалось доказать. □

4.6 Интерполяция

Литература: [F], гл. VI, § 4, пп. 1–3; [K1], гл. 6, § 1, п. 2; [vdW], гл. 5, § 29.

Определение 4.6.1. Пусть k — поле, $x_1, \ldots, x_n \in k$ — некоторые попарно различные элементы k, и $y_1, \ldots, y_n \in k$. **Интерполяционной задачей** (или **задачей интерполяции в n точках**) с данными $(x_1, \ldots, x_n; y_1, \ldots, y_n)$ мы будем называть задачу нахождения многочлена $f \in k[x]$ такого, что $f(x_i) = y_i$ для всех $i = 1, \ldots, m$.

Теорема 4.6.2. *Интерполяционная задача имеет не более одного решения среди многочленов степени, не превосходящей $n-1$. Более того, если f, g — два решения одной интерполяционной задачи, то $f - g$ делится на многочлен $(x - x_1) \ldots (x - x_n)$.*

Доказательство. Пусть $f, g \in k[x]$ — два многочлена, являющихся решениями одной интерполяционной задачи с данными $(x_1, \ldots, x_n; y_1, \ldots, y_n)$. Это означает, что $f(x_i) = y_i = g(x_i)$ для всех $i = 1, \ldots, n$. Рассмотрим многочлен $h = f - g$; тогда $h(x_i) = f(x_i) - g(x_i) = 0$ для всех i. Все x_i различны, поэтому у многочлена h есть n различных корней $x_1, \ldots, x_n$. По предложению 4.3.6 из этого следует, что h делится на $(x - x_1) \ldots (x - x_n)$. В частности, если f и g были многочленами степени не выше $n-1$, то и степень h не превосходит $n-1$, откуда $h = 0$ и $f = g$. □

Замечание 4.6.3. У многочлена степени $n-1$ ровно n коэффициентов; неформально говоря, эти n «степеней свободы» фиксируются выбором его значений в n точках.

Сейчас мы покажем, что всякая задача интерполяции в n точках имеет решение, являющееся многочленом степени не выше $n-1$ (и, стало быть, имеет единственное решение среди многочленов такой степени). Мы явно построим по данным интерполяционной задачи нужный многочлен нужной степени, и даже двумя способами: Лагранжа и Ньютона.

Пусть $(x_1, \ldots, x_n; y_1, \ldots, y_n)$ — фиксированная интерполяционная задача. Обозначим

$$\varphi_i = (x - x_1) \ldots \widehat{(x - x_i)} \ldots (x - x_n);$$

здесь знак $\widehat{\ }$ над скобкой означает, что соответствующий множитель нужно пропустить. Более формально,

$$\varphi_i = \prod_{\substack{1 \leqslant j \leqslant n \\ j \neq i}} (x - x_j).$$

Отметим, что φ_i является многочленом степени $n-1$, а его корни — элементы $x_1, \ldots, \widehat{x_i}, \ldots, x_n$.

Посмотрим теперь на многочлен $\varphi_i / \varphi_i(x_i)$. Эта запись имеет смысл, поскольку $\varphi_i(x_i) \neq 0$. Указанный многочлен принимает значение 1 в точке x_i и значения 0 во всех остальных точках из набора $x_1, \ldots, x_n$.

Наконец, рассмотрим сумму $f = \sum_{i=1}^{n} y_i \varphi_i / \varphi_i(x_i)$. При подстановке x_i в многочлен f все слагаемые, кроме $y_i \varphi_i / \varphi_i(x_i)$, обратятся в 0, а указанное слагаемое примет значение y_i. Значит, указанный многочлен является решением нашей интерполяционной задачи. Кроме того, степень f не превосходит $n-1$, поскольку степень каждого φ_i равна $n-1$.

Выпишем его еще раз:

$$f = \sum_{i=1}^{n} y_i \frac{(x-x_1)\dots\widehat{(x-x_i)}\dots(x-x_n)}{(x_i-x_1)\dots\widehat{(x_i-x_i)}\dots(x_i-x_n)}.$$

Многочлен f называется **интерполяционным многочленом Лагранжа**.

Обратимся теперь ко второму способу, который носит название **интерполяционного многочлена Ньютона**. Он решает ту же самую задачу интерполяции в n точках и имеет степень не выше $n-1$; конечно, из единственности решения следует, что он совпадает с интерполяционным многочленом Лагранжа и отличается лишь формой записи. Форма Ньютона удобна, когда добавление новых точек к интерполяционной задаче происходит последовательно.

А именно, мы построим серию многочленов $f_1, f_2, \dots, f_n$ таких, что многочлен f_i имеет степень не выше $i-1$ и решает задачу интерполяции в i точках с данными $(x_1, \dots, x_i; y_1, \dots, y_i)$. Построении будет происходить по индукции: мы опишем, как строить f_1 и как по многочлену f_i строить многочлен f_{i+1}; очевидно, что f_n будет решением исходной интерполяционной задачи.

Задача интерполяции в одной точке проста — в качестве многочлена f_1, принимающего значение y_1 в точке x_1, можно взять константу: $f_1 = y_1$ — это действительно многочлен степени не выше 0, что и требовалось. Предположим теперь, что многочлен f_i построен, то есть, $f_j(x_j) = y_j$ для всех $j = 1, \dots, i$, и $\deg(f_i) \leqslant i-1$. Как построить f_{i+1}? Будем искать его в виде $f_{i+1} = f_i + c_{i+1}(x-x_1)\dots(x-x_i)$, где $c_{i+1} \in k$ — некоторая константа. Это гарантирует нам, что значения f_i в точках $x_1, \dots, x_i$ не испортятся: добавка $c_{i+1}(x-x_1)\dots(x-x_i)$ обращается в 0 в этих точках. Это означает, что $f_{i+1}(x_j) = y_j$ для $j = 1, \dots, i$. Кроме того, степень f_{i+1} не превосходит i. Осталось добиться выполнения условия $f_{i+1}(x_{i+1}) = y_{i+1}$ подбором константы c_{i+1}. То есть, нам нужно, чтобы $f_i(x_{i+1}) + c_{i+1}(x_{i+1}-x_1)\dots(x_{i+1}-x_i) = y_{i+1}$. Отсюда легко находится c_{i+1}:

$$c_{i+1} = \frac{y_{i+1} - f_i(x_{i+1})}{(x_{i+1}-x_1)\dots(x_{i+1}-x_i)}.$$

Заметим, что знаменатель этой дроби — ненулевая константа.

Таким образом, интерполяционный многочлен Ньютона является многочленом f_n в после-

довательности

$$
\begin{aligned}
f_1 &= y_1;\\
f_2 &= f_1 + \frac{y_2 - f_1(x_2)}{x_2 - x_1};\\
f_3 &= f_2 + \frac{y_3 - f_2(x_3)}{(x_3 - x_1)(x_3 - x_2)};\\
&\vdots\\
f_n &= f_{n-1} + \frac{y_n - f_{n-1}(x_n)}{(x_n - x_1)\dots(x_n - x_{n-1})}.
\end{aligned}
$$

4.7 НОД и неприводимость

Литература: [F], гл. VI, § 1, пп. 3–6; [K1], гл. 5, § 3, п. 1–2.

Продолжим построение теории делимости в кольце многочленов, параллельной теории делимости в кольце целых чисел. Начиная с этого места, мы будем рассматривать многочлены над полем k.

Определение 4.7.1. Пусть $f, g \in k[x]$. Многочлен d называется **общим делителем** многочленов f и g, если $d \mid f$ и $d \mid g$.

Определение 4.7.2. Пусть $f, g \in k[x]$. Многочлен d называется **наибольшим общим делителем** многочленов f и g (обозначение: $d = \gcd(f, g)$), если

1. d — общий делитель f и g;
2. если d' — еще какой-нибудь общий делитель f и g, то $d' \mid d$.

Замечание 4.7.3. Сразу же заметим, что если d и d' — два наибольших общих делителя многочленов f и g, то по определению имеем $d \mid d'$ и $d' \mid d$; это означает, что многочлены d и d' ассоциированы, то есть, отличаются домножением на ненулевую константу. В кольце целых чисел у каждого элемента не более двух ассоциированных — он сам и противоположный к нему, и поэтому можно выбрать из них неотрицательный, и считать его наибольшим общим делителем. В кольце многочленов неизвестно, какой из (возможного) множества ассоциированных выбирать; можно, конечно, всегда выбирать многочлен со старшим коэффициентом 1, но мы этого не будем делать, и будем говорить, что gcd многочленов *определен с точностью до ассоциированности.*

Теорема 4.7.4. *Наибольший общий делитель многочленов $f, g \in k[x]$ существует, определен однозначно с точностью до ассоциированности, и может быть представлен в виде $\gcd(f, g) = u_0 f + v_0 g$ для некоторых $u_0, v_0 \in k[x]$*

Доказательство. Заметим, что $\gcd(0, g) = g$, поэтому можно считать, что $f \neq 0$ и $g \neq 0$. Рассмотрим множество I многочленов вида $uf + vg$ для всевозможных $u, v \in k[x]$ и выберем

из них ненулевой многочлен $d = u_0 f + v_0 g$ наименьшей степени (возможно, таких несколько — возьмем любой из них). Мы утверждаем, что d является наибольшим общим делителем f и g. Поделим f на d с остатком: $f = dh + r$, где $\deg(r) < \deg(d)$. Тогда $r = f - dh = f - (u_0 f + v_0 g)h = (1 - u_0 h)f + (-v_0 h)g$ лежит в I и имеет меньшую степень; поэтому $r = 0$, то есть, f делится на d. Аналогично, g делится на d. Это означает, что d — общий делитель f и g. Если же h — какой-то общий делитель f и g, то и $d = u_0 f + v_0 g$ делится на h. □

Замечание 4.7.5. Представление из теоремы 4.7.4 называется, как и в случае целых чисел, **линейным представлением наибольшего общего делителя**.

Совершенно аналогично случаю целых чисел происходит и **алгорифм Эвклида** в кольце многочленов: единственное отличие состоит в том, что при каждом шаге алгорифма убывает не модуль числа, а степень многочлена:

Лемма 4.7.6. *Если* $f = gq + r$ *для* $f, g \in k[x]$, *то* $\gcd(f, g) = \gcd(g, r)$.

Доказательство. Пусть $d = \gcd(f, g)$; тогда $r = f - gq$ делится на d, и если h — некоторый общий делитель g и r, то $f = gq + r$ делится на h, поэтому h является общим делителем f и g, и по определению наибольшего общего делителя должно выполняться $h \mid d$. Поэтому d является и наибольшим общим делителем g и r. □

Теперь для того, чтобы найти $\gcd(f, g)$, можно считать, что $\deg(f) \geqslant \deg(g)$ и $g \neq 0$. Запишем $f = gq_1 + r_1$ и заметим, что $\gcd(f, g) = \gcd(g, r_1)$, причем $\gcd(r_1) < \gcd(g)$, поэтому можно перейти от пары (f, g) к паре (g, r_1) и повторить операцию:

$$\begin{aligned} f &= gq_1 + r_1 \\ g &= r_1 q_2 + r_2 \\ r_1 &= r_2 q_3 + r_3 \\ &\dots \end{aligned}$$

Процесс не может продолжаться бесконечно, поскольку степень остатка убывает. Стало быть, он остановится, когда очередной остаток окажется равным 0; если r_n — последний ненулевой остаток, то $\gcd(f, g) = \gcd(g, r_1) = \gcd(r_1, r_2) = \dots = \gcd(r_{n-1}, r_n) = \gcd(r_n, 0) = r_n$.

Уточним степени многочленов, входящих в линейное представление НОД из теоремы 4.7.4:

Предложение 4.7.7. *Пусть* $f, g \in k[x]$, $d = \gcd(f, g)$, $\deg(f) = m$, $\deg(g) = n$. *Существуют многочлены* $u_0, v_0 \in k[x]$ *такие, что* $\deg(u_0) < n$, $\deg(v_0) < m$, *и* $d = u_0 f + v_0 g$.

Доказательство. Без ограничения общности можно считать, что $m \leqslant n$. По теореме 4.7.4 найдутся *какие-то* $u_0', v_0' \in k[x]$ такие, что $d = u_0' f + v_0' g$. Поделим u_0' с остатком на g: $u_0' = gq + r$. Тогда $d = u_0' f + v_0' g = (gq + r)f + v_0' g = rf + (v_0' - qf)g$. Положим $u_0 = r$, $v_0 = v'0 - qf$. Мы знаем, что $\deg(u_0) < \deg(g) = n$. Наконец, $v_0 g = d - u_0 f$, причем $\deg(d) < \deg(f) = m$ и $\deg(u_0 f) = \deg(u_0) + \deg(f) < n + m$; поэтому $n + m > \deg(v_0 g) = \deg(v_0) + \deg(g) = \deg(v_0) + n$ и $\deg(v_0) < m$, что и требовалось. □

Наконец, определим аналоги простых чисел в кольце многочленов.

Определение 4.7.8. Многочлен $p \in k[x]$ называется **неприводимым**, если p ненулевой, необратимый, и из того, что $p = fg$ для $f, g \in k[x]$, следует, что f ассоциировано с p или g ассоциировано с p.

Лемма 4.7.9. *Пусть* $f, g, p \in k[x]$ *и* p *неприводим. Если* $p \mid fg$, *то* $p \mid f$ *или* $p \mid g$.

Доказательство. Если f не делится на p, то $\gcd(f, p) = 1$. Запишем $1 = u_0 f + v_0 p$ и домножим это равенство на g: $g = u_0 fg + v_0 pg$. По условию fg делится на p, поэтому оба слагаемых в правой части делятся на p, поэтому и g делится на p. □

Теорема 4.7.10. *Любой ненулевой необратимый многочлен* f *из* $k[x]$ *представляется в виде* $f = p_1 \dots p_m$, *где* $p_1, \dots, p_m \in k[x]$ — *неприводимые многочлены. Более того, такое разложение однозначно с точностью до порядка сомножителей и замены их на ассоциированные.*

Доказательство. Для доказательства существования — индукция по степени многочлена f; если f неприводим, доказывать нечего, иначе же запишем $f = gh$ так, чтобы степени g и h были меньше степени f и воспользуемся индукционным предположением.

Доказательство единственности проходит точно так же, как в случае целых чисел (см. теорему 2.5.3), только индукцию снова нужно вести не по модулю числа, а по степени многочлена. □

4.8 Поля частных

Литература: [F], гл. VI, § 3, пп. 1–2; [K1], гл. 5, § 4, п. 1; [vdW], гл. 3, § 13.

Пусть R — область целостности (см. определение 2.8.9). Сейчас мы расширим кольцо R до поля естественным образом. Эта конструкция совершенно аналогична переходу от целых чисел к рациональным: рациональное число можно считать дробью, в числителе и знаменателе которой стоят целые числа. Первая проблема, которую нужно побороть — неоднозначность представления в виде дроби: например, дроби $4/6$, $(-2)/(-3)$ и $2/3$ обозначают одно и то же рациональное число.

Рассмотрим множество $R \times (R \backslash \{0\})$ и введем на нем следующее отношение: пара (a, s) считается эквивалентной паре (b, t) тогда и только тогда, когда $at = bs$ в R. Мы будем использовать обычное обозначение для этого отношения: $(a, s) \sim (b, t)$

Лемма 4.8.1. *Это отношение эквивалентности на* $R \times (R \backslash \{0\})$.

Доказательство. Рефлексивность: $(a, s) \sim (a, s)$, поскольку $as = as$. Симметричность: если $(a, s) \sim (b, t)$, то $at = cs$, откуда $(b, t) \sim (a, s)$. Транзитивность: если $(a, s) \sim (b, t)$ и $(b, t) \sim (c, u)$, то $at = bs$ и $bu = ct$. Поэтому $atu = bsu = cts$, откуда $t(au - cs) = 0$ и, поскольку $t \neq 0$, а R — область целостности, получаем $au = cs$, что означает, что $(a, s) \sim (c, u)$. □

Фактор-множество $R \times (R\backslash\{0\})$ по указанному отношению эквивалентности мы будем обозначать через $\mathrm{Frac}(R)$, а класс пары (a, s) в $\mathrm{Frac}(R)$ будем обозначать через $\frac{a}{s}$ и называть **дробью**. Теперь введем на полученном множестве операции по образу и подобию операций над рациональными числами:

$$\frac{a}{s} + \frac{b}{t} = \frac{at + bs}{st};$$
$$\frac{a}{s} \cdot \frac{b}{t} = \frac{ab}{st}.$$

Как всегда при введении операций на фактор-множестве, эта запись a priori содержит неоднозначность, которую нужно разрешить, проверив *корректность* введенных операций.

Сначала разберемся с произведением: мы определили произведение двух классов $x, y \in \mathrm{Frac}(R)$ с помощью выбора представителей: если (a, s) — представитель класса x, а (b, t) — представитель класса y, мы определили xy как класс, содержащий пару (ab, st). Для начала заметим, что $st \neq 0$ (поскольку R — область целостности), поэтому эта пара действительно лежит в $R \times (R\backslash\{0\})$. Что будет, если мы выберем других представителей? Пусть, действительно, (a', s') — еще одна пара из класса x, а (b', t') — пара из класса y. Это означает, что $(a, s) \sim (a', s')$ и $(b, t) \sim (b', t')$. Верно ли, что пары (ab, st) и $(a'b', s't')$ попали в один класс? Проверим это: нам дано $as' = a's$ и $bt' = b't$, а хочется проверить, что $abs't' = a'b'st$. Для этого нужно лишь перемножить два данных равенства.

Далее, мы определили сумму двух классов x и y так: если (a, s) — представитель класса x, а (b, t) — представитель класса y, мы определили $x + y$ как класс, содержащий пару $(at + bs, st)$. Что будет при выборе других представителей? Пусть снова (a', s') — еще одна пара из класса x, а (b', t') — пара из класса y, то есть, $(a, s) \sim (a', s')$ и $(b, t) \sim (b', t')$. Верно ли, что пары $(at + bs, st)$ и $(a't' + b's', s't')$ попали в один класс? Нам дано нам дано $as' = a's$ и $bt' = b't$, а хочется проверить, что $(at + bs)s't' = (a't' + b's')st$. Но из $as' = a's$ следует $as'tt' = a'stt'$, а из $bt' = b't$ следует $bss't' = b'ss't$, и сложением получаем $as'tt' + bss't' = a'stt' + b'ss't$, то есть, $(at + bs)s't' = (a't' + b's')st$, что и требовалось.

Операции на $\mathrm{Frac}(R)$ определены, осталось проверить, что получилось поле.

Теорема 4.8.2. *Пусть R — область целостности. Множество $\mathrm{Frac}(R)$ с введенными выше операциями является полем.*

Определение 4.8.3. $\mathrm{Frac}(R)$ называется **полем частных** области целостности R.

Доказательство теоремы. 1. Ассоциативность сложения: $(\frac{a}{s} + \frac{b}{t}) + \frac{c}{u} = \frac{at+bs}{st} + \frac{c}{u} = \frac{(at+bs)u+cst}{stu}$, $\frac{a}{s} + (\frac{b}{t} + \frac{c}{u}) = \frac{a}{s} + \frac{bu+ct}{tu} = \frac{atu+(bu+ct)s}{stu}$, что то же самое.

2. Нейтральный элемент по сложению — дробь $\frac{0}{1}$. Действительно, $\frac{a}{s} + \frac{0}{1} = \frac{a \cdot 1 + 0 \cdot s}{s \cdot 1} = \frac{a}{s}$; перемножение в другом порядке можно опустить в силу коммутативности (см. пункт 4). Заметим, что $\frac{0}{1} = \frac{0}{s}$ для любого $s \in R\backslash\{0\}$.

3. Противоположной дробью к $\frac{a}{s}$ будет дробь $\frac{-a}{s}$: $\frac{a}{s} + \frac{-a}{s} = \frac{as+(-a)s}{s \cdot s} = \frac{0}{s \cdot s} = \frac{0}{1}$.

4. Коммутативность сложения: $\frac{a}{s} + \frac{b}{t} = \frac{at+bs}{st}$, $\frac{b}{t} + \frac{a}{s} = \frac{bs+at}{st}$.

5. Ассоциативность умножения: $(\frac{a}{s} \cdot \frac{b}{t}) \cdot \frac{c}{u} = \frac{ab}{st} \cdot \frac{c}{u} = \frac{abc}{stu} = \frac{a}{s} \cdot \frac{bc}{tu} = \frac{a}{s}(\frac{b}{t} \cdot \frac{c}{u})$.

6. Нейтральный элемент по умножению — дробь $\frac{1}{1}$. Действительно, $\frac{a}{s} \cdot \frac{1}{1} = \frac{a \cdot 1}{s \cdot 1} = \frac{a}{s}$. Заметим, что $\frac{1}{1} = \frac{s}{s}$ для любого $s \in R \backslash \{0\}$.

7. Коммутативность умножения: $\frac{a}{s} \cdot \frac{b}{t} = \frac{ab}{st} = \frac{b}{t} \cdot \frac{a}{s}$.

8. Аксиома поля: у каждой дроби $\frac{a}{s} \neq 0$ есть обратный элемент по умножению. Заметим, что если $a = 0$, то $\frac{a}{s} = 0$. Поэтому $a \neq 0$ и можно рассмотреть дробь $\frac{s}{a}$, которая и будет обратной: $\frac{a}{s} \cdot \frac{s}{a} = \frac{as}{as} = \frac{1}{1} = 1$.

Осталось заметить, что в полученном кольце $\mathrm{Frac}(R)$ выполнено условие $0 \neq 1$: условие $\frac{0}{1} = \frac{1}{1}$ означало бы, что $0 \cdot 1 = 1 \cdot 1$ в R, то есть, $0 = 1$, что невозможно, поскольку R — область целостности. □

Отметим теперь, что кольцо R можно считать лежащим в поле $\mathrm{Frac}(R)$: каждому элементу $a \in R$ можно сопоставить дробь $\frac{a}{1}$; при этом разным элементам R сопоставляются разные дроби, поскольку из $\frac{a}{1} = \frac{b}{1}$ следует $a \cdot 1 = b \cdot 1$, то есть, $a = b$. Сложение и умножение полученных дробей выглядит так же, как сложение и умножение в R: $\frac{a}{1} + \frac{b}{1} = \frac{a+b}{1}$, $\frac{a}{1} \cdot \frac{b}{1} = \frac{ab}{1}$. Таким образом, можно считать, что мы расширили R и у каждого ненулевого элемента $s \in R$ в новом кольце $\mathrm{Frac}(R)$ оказался обратный: дробь $\frac{1}{s}$.

Пример 4.8.4. Из конструкции очевидно, что $\mathrm{Frac}(\mathbb{Z}) = \mathbb{Q}$.

4.9 Поле рациональных функций

Литература: [F], гл. VI, § 3, пп. 1–5, 7; [K1], гл. 5, § 2, п. 2–3; [vdW], гл. 5, § 36.

Определение 4.9.1. Применим конструкцию поля частных к кольцу многочленов $k[x]$ над полем k. Полученное поле $\mathrm{Frac}(k[x])$ называется **полем рациональных функций (над k)** и обозначается через $k(x)$.

Таким образом, поле рациональных функций состоит из дробей вида $\frac{f}{g}$, где f, g — многочлены (с учетом отношения эквивалентности), которые складываются и перемножаются как привычные дроби. Исходное кольцо $k[x]$ мы трактуем как подмножество $k(x)$, состоящее из дробей вида $\frac{f}{1}$.

Замечание 4.9.2. Слово «функция» в термине «поле рациональных функций» несколько обманчиво: мы уже убедились, что не стоит отождествлять многочлен $f \in k[x]$ с функцией $k \to k$, $c \mapsto f(c)$. Точно так же, можно попытаться сопоставить рациональной функции $\frac{f}{g} \in k(x)$ отображение $k \to k$, $c \mapsto f(c)/g(c)$, однако она не определена в точках c, для которых $g(c) = 0$; кроме этого, у разных представителей класса дроби f/g будут разные области определения: например, дробь $\frac{1}{x-1}$ не определена в точке 1, а равная ей дробь $\frac{x}{x(x-1)}$ не определена в точках 0 и 1. Может оказаться, что указанное отображение не определено вообще ни в одной точке: для поля $k = \mathbb{Z}/p\mathbb{Z}$ знаменатель дроби $\frac{1}{x^p - x}$, например, обращается в 0 во всех точках $c \in k$. Это показывает, что с подстановкой значений в дроби нужно быть предельно аккуратным.

Определение 4.9.3. Рациональная функция $\frac{f}{g} \in k(x)$ называется **правильной**, если $\deg(f) < \deg(g)$

Лемма 4.9.4. *Это определение корректно, то есть, не зависит от выбора представителей: если* $\frac{f}{g} = \frac{\widetilde{f}}{\widetilde{g}}$, *и* $\deg(f) < \deg(g)$, *то* $\deg(\widetilde{f}) < \deg(\widetilde{g})$.

Доказательство. Если $\frac{f}{g} = \frac{\widetilde{f}}{\widetilde{g}}$, то $f\widetilde{g} = \widetilde{f}g$, поэтому $\deg(f) + \deg(\widetilde{g}) = \deg(\widetilde{f}) + \deg(g)$. □

Лемма 4.9.5. *Сумма, разность и произведение правильных дробей — правильные дроби.*

Доказательство. Пусть $\frac{f}{g}$ и $\frac{\widetilde{f}}{\widetilde{g}}$ — правильные дроби, то есть, $\deg(f) < \deg(g)$ и $\deg(\widetilde{f}) < \deg(\widetilde{g})$. Тогда $\frac{f}{g} + \frac{\widetilde{f}}{\widetilde{g}} = \frac{f\widetilde{g}+\widetilde{f}g}{g\widetilde{g}}$. При этом $\deg(f\widetilde{g}) < \deg(g\widetilde{g})$ и $\deg(\widetilde{f}g) < \deg(g\widetilde{g})$, поэтому и полученная сумма является правильной дробью. Для случая разности достаточно заметить, что противоположная дробь к правильной дроби также является правильной. Наконец, $\deg(f\widetilde{f}) < \deg(g\widetilde{g})$, поэтому и произведение $\frac{f\widetilde{f}}{g\widetilde{g}}$ является правильной дробью. □

Лемма 4.9.6. *Если многочлен равен правильной дроби, то он нулевой.*

Доказательство. Предположим, что $f \in k[x]$ — некоторый многочлен, $\psi = \frac{g}{h} \in k(x)$ — правильная дробь (здесь $g, h \in k[x]$), и $f = \psi$. Равенство $f = \frac{g}{h}$ означает, что $fh = g$, и поэтому $\deg(g) = \deg(f) + \deg(h)$. С другой стороны, по определению правильной дроби $\deg(g) < \deg(h)$. Поэтому $\deg(f) < 0$, то есть, $f = 0$. □

Предложение 4.9.7. *Любую рациональную функцию* $\varphi \in k(x)$ *можно единственным образом представить в виде суммы многочлена и правильной рациональной функции:* $\varphi = f + \psi$, *где* $f \in k[x]$, $\psi \in k(x)$, *и если* $\varphi = \widetilde{f} + \widetilde{\psi}$, *то* $f = \widetilde{f}$ *и* $\psi = \widetilde{\psi}$. *Более того, знаменатель* ψ *можно взять равным знаменателю* φ, *то есть, если* $\varphi = \frac{a}{b}$ *для некоторых* $a, b \in k[x]$, *то* $\psi = \frac{c}{b}$ *для некоторого* $c \in k[x]$.

Доказательство. Запишем $\varphi = \frac{a}{b}$ для некоторых $a, b \in k[x]$, $b \neq 0$. Поделим a на b с остатком: $a = bq + r$, где $q, r \in k[x]$ и $\deg(r) < \deg(b)$. Тогда $\varphi = \frac{a}{b} = \frac{bq+r}{b} = \frac{bq}{b} + \frac{r}{b} = \frac{q}{1} + \frac{r}{b} = q + \frac{r}{b}$, и дробь $\frac{r}{b}$ правильная. Докажем единственность: пусть $f + \psi = \widetilde{f} + \widetilde{\psi}$, тогда $f - \widetilde{f} = \widetilde{\psi} - \psi$. В левой части этого равенства стоит многочлен, в правой — правильная дробь (по лемме 4.9.5); из леммы 4.9.6 следует, что $f - \widetilde{f} = 0$, то есть, $f = \widetilde{f}$ и $\psi = \widetilde{\psi}$. Заметим, наконец, что в нашем построении знаменатель ψ равен знаменателю ϕ. □

Выделение многочлена является первым шагом на пути к выявлению структуры поля рациональных функций.

Определение 4.9.8. Рациональная функция $\psi \in k(x)$ называется **простейшей**, если ее можно представить в виде $\psi = \frac{f}{p^m}$, где $f, p \in k[x]$, p — неприводимый многочлен, $m \geqslant 1$ — натуральное число, и $\deg(f) < \deg(p)$.

Наша цель — доказать, что любая правильная рациональная функция представляется (в некотором смысле единственным образом) в виде суммы простейших.

Лемма 4.9.9. *Пусть* $\frac{f}{gh} \in k(x)$ *— правильная рациональная функция, и многочлены* $g, h \in k[x]$ *взаимно просты:* $\gcd(g,h) = 1$*.. Тогда* $\frac{f}{gh}$ *можно представить в виде* $\frac{f}{gh} = \frac{a}{g} + \frac{b}{h}$*, где* $\frac{a}{g}, \frac{b}{h} \in k(x)$ *— правильные рациональные функции.*

Доказательство. Запишем $ug + vh = 1$. Тогда $\frac{f}{gh} = f \cdot \frac{1}{gh} = f \cdot \frac{ug+vh}{gh} = f \cdot (\frac{ug}{gh} + \frac{vh}{gh}) = f \cdot (\frac{u}{h} + \frac{v}{g}) = \frac{fv}{g} + \frac{uf}{h}$. В силу предложения 4.9.7 можно записать дроби $\frac{fv}{g}$ и $\frac{uf}{h}$ как суммы многочленов и правильных дробей с теми же знаменателями. Соединяя многочлены вместе, получаем $\frac{f}{gh} = c + \frac{a}{g} + \frac{b}{h}$, где $a, b, c \in k[x]$. Наконец, из этого равенство видно, что c является суммой правильных дробей, то есть, по лемме 4.9.5, правильной дробью, и из единственности в предложении 4.9.7, $c = 0$. □

Лемма 4.9.10. *Правильную дробь вида* $\frac{f}{p^m}$ *(здесь* $f, p \in k[x]$*,* $m > 1$*) можно записать в виде суммы* $\frac{a_1}{p} + \frac{a_2}{p^2} + \cdots + \frac{a_m}{p^m}$*, где* $a_i \in k[x]$*,* $\deg a_i < \deg p$*.*

Доказательство. Индукция по m. База $m = 1$ очевидна. Переход: пусть $m > 1$. Поделим f на p с остатком: $f = pq + r$, $\deg(r) < \deg(p)$. Теперь можно записать $\frac{f}{p^m} = \frac{pq+r}{p^m} = \frac{pq}{p^m} + \frac{r}{p^m} = \frac{q}{p^{m-1}} + \frac{r}{p^m}$ и по предположению индукции первую дробь можно записать как сумму дробей, в которых присутствуют знаменатели $p, p^2, \ldots, p^{m-1}$, а числители имеют степень, меньшую степени p. Приписывая слагаемое $\frac{r}{p^m}$, получаем то, что требовалось. □

Наконец, все готово для доказательства главной теоремы.

Теорема 4.9.11. *Пусть* $\frac{f}{g} \in k(x)$ *— правильная дробь,* $g = p_1^{m_1} \ldots p_s^{m_s}$ *— каноническое разложение* g *на неприводимые множители. Тогда* $\frac{f}{g}$ *можно представить в виде суммы простейших дробей, в знаменателях которых стоят* $p_1, p_1^2, \ldots, p_1^{m_1}$*,* $p_2, p_2^2, \ldots, p_2^{m_2}, \ldots,$ $p_s, p_s^2, \ldots, p_s^{m_s}$*. Кроме того, такое представление единственно с точностью до порядка, в котором записаны слагаемые.*

Доказательство. По предложению 4.9.9 можно расщепить знаменатель правильной дроби на два взаимно простых сомножителя; применяя ее несколько раз, получаем, что $\frac{f}{g} = \frac{f_1}{p_1^{m_1}} + \cdots + \frac{f_s}{p_s^{m_s}}$. Далее, по лемме 4.9.10, каждое слагаемое вида $\frac{f_i}{p_i^{m_i}}$ представляется в виде суммы простейших.

Для доказательства единственности предположим, что сумма простейших дробей указанного вида равна другой сумме простейших дробей того же вида. Докажем, что все числители соответствующих дробей в обеих частях этого равенства совпадают. Предположим противное — нашлись различные числители в дробях с одинаковыми знаменателями в левой и правой частях. Без ограничения общности (с точности до нумерации многочленов $p_1, \ldots, p_s$) можно считать, что знаменатели этих дробей — степени многочлена p_1. Посмотрим на все дроби в левой и правой части, знаменатели которых — степени p_1: пусть в левой части стоит $\frac{a_1}{p_1} + \frac{a_2}{p_1^2} + \cdots + \frac{a_{m_1}}{p_1^{m_1}}$, а в правой части — $\frac{b_1}{p_1} + \frac{b_2}{p_1^2} + \cdots + \frac{b_{m_1}}{p_1^{m_1}}$. По нашему предположению, $a_n \neq b_n$ для некоторого n. Рассмотрим максимальное такое n. Тогда $a_{n+1} = b_{n+1}, \ldots, a_{m_1} = b_{m_1}$, поэтому дроби $\frac{a_{n+1}}{p_1^{n+1}}, \ldots, \frac{a_{n+1}}{p_1^{n+1}}$ в левой части равны соответственно дробям $\frac{b_{n+1}}{p_1^{n+1}}, \ldots, \frac{b_{n+1}}{p_1^{n+1}}$ в правой части. Вычеркивая эти дроби, получаем равенство вида

$$\frac{a_1}{p_1} + \frac{a_2}{p_1^2} + \cdots + \frac{a_n}{p_1^n} + A = \frac{b_1}{p_1} + \frac{b_2}{p_1^2} + \cdots + \frac{b_n}{p_1^n} + B,$$

где A и B — суммы дробей, в знаменателях которых стоит степени $p_2, \dots, p_s$. При этом, по предположению, $a_n \neq b_n$. Домножим указанное равенство на $p_1^n p_2^{m_2} \dots p_s^{m_s}$:

$$(a_1 p_1^{n-1} + a_2 p_1^{n-2} + \dots + a_n) p_2^{m_2} \dots p_s^{m_s} + A p_1^n p_2^{m_2} \dots p_s^{m_s} =$$
$$(b_1 p_1^{n-1} + b_2 p_1^{n-2} + \dots + b_n) p_2^{m_2} \dots p_s^{m_s} + B p_1^n p_2^{m_2} \dots p_s^{m_s}.$$

Это уже равенство многочленов (мы избавились от всех знаменателей). Раскроем скобки и заметим, что в левой части лишь одно слагаемое не содержит множитель p_1, а именно, $a_n p_2^{m_2} \dots p_s^{m_s}$. Действительно, по предположению, A не содержит степени p_1 в знаменателях, и остальные слагаемые слева (если они вообще есть) также делятся на p_1. Аналогично, в правой части лишь слагаемое $b_n p_2^{m_2} \dots p_s^{m_s}$ не содержит множитель p_1. Поэтому наше равенство принимает вид

$$a_n p_2^{m_2} \dots p_s^{m_s} + (\dots) \cdot p_1 = b_n p_2^{m_2} \dots p_s^{m_s} + (\dots) \cdot p_1.$$

Значит, разность $a_n p_2^{m_2} \dots p_s^{m_s} - b_n p_2^{m_2} \dots p_s^{m_s} = (a_n - b_n) p_2^{m_2} \dots p_s^{m_s}$ делится на p_1; однако, $p_2, \dots, p_s$ взаимно просты с p_1, поэтому $a_n - b_n$ делится на p_1. Но мы начинали с суммы простейших дробей, то есть, $\deg(a_n) < \deg(p_1)$ и $\deg(b_n) < \deg(p_1)$, откуда $\deg(a_n - b_n) < \deg(p_1)$ и, стало быть, $a_n = b_n$ — противоречие. □

Следствие 4.9.12. *1. Любая правильная дробь из $\mathbb{C}(x)$ представляется в виде суммы дробей вида $\frac{a}{(x-c)^m}$, где $a, c \in \mathbb{C}$, $m \geqslant 1$.*

2. Любая правильная дробь из $\mathbb{R}(x)$ представляется в виде суммы дробей вида $\frac{a}{(x-c)^m}$, где $a, c \in \mathbb{R}$, $m \geqslant 1$, и дробей вида $\frac{cx+d}{(x^2+ax+b)^m}$, где $a, b, c, d \in \mathbb{R}$, $a^2 - 4b < 0$, $m \geqslant 1$.

Доказательство. Напрямую следует из теоремы 4.9.11 и теорем 4.4.3, 4.4.4. □

Теорема 4.9.11 не указывает явного алгоритма нахождения разложения правильной дроби в сумму простейших. Этот алгоритм можно извлечь из доказательства предложения 4.9.9 и леммы 4.9.10, но он несколько замысловат: например, в доказательстве 4.9.9 требуется умение находить коэффициенты в линейном представлении наибольшего общего делителя. На практике для нахождения разложения в сумму простейших хорошо работает метод неопределенных коэффициентов. Кроме того, можно выписать и явные формулы (конечно, если известно разложение знаменателя дроби на неприводимые многочлены). Приведем формулы для простейшего случая: рациональной функции над комплексными числами, знаменатель которой не имеет кратных корней.

Предложение 4.9.13. *Пусть $\frac{f}{g} \in \mathbb{C}(x)$ — правильная дробь, и $g = (x - c_1) \dots (x - c_n)$, где $c_1, \dots, c_n \in \mathbb{C}$ — попарно различные числа. Тогда $\frac{f}{g} = \frac{a_1}{x-c_1} + \dots + \frac{a_n}{x-c_n}$, где $a_i = f(c_i)/g'(c_i)$.*

Доказательство. По теореме 4.9.11 существует разложение вида $\frac{f}{g} = \sum_{i=1}^{n} \frac{a_i}{x-c_i}$; осталось найти коэффициенты a_j для всех j. Домножим это равенство на g:

$$f = \sum_{i=1}^{n} a_i (x - c_1) \dots \widehat{(x - c_i)} \dots (x - c_n)$$

(напомним, что крышечка над множителем означает, что его нужно пропустить в произведении). Подставим c_j; все слагаемые справа, кроме j-го, содержат множитель $(x - c_j)$, поэтому обращаются в нуль. Значит,

$$f(c_j) = a_j(c_j - c_1)\dots\widehat{(c_j - c_j)}\dots(c_j - c_n).$$

Посмотрим теперь на производную многочлена $g = (x - c_1)\dots(x - c_n)$:

$$\begin{aligned} g' &= ((x - c_j)(x - c_1)\dots\widehat{(x - c_j)}\dots(x - c_n))' \\ &= (x - c_j)'(x - c_1)\dots\widehat{(x - c_j)}\dots(x - c_n) + (x - c_j)((x - c_1)\dots\widehat{(x - c_j)}\dots(x - c_n))'. \\ &= (x - c_1)\dots\widehat{(x - c_j)}\dots(x - c_n) + (x - c_j)((x - c_1)\dots\widehat{(x - c_j)}\dots(x - c_n))'. \end{aligned}$$

Наконец, подставим c_j, и второе слагаемое обратится в 0: $g'(c_j) = (c_j - c_1)\dots\widehat{(c_j - c_j)}\dots(c_j - c_n)$. Сравнивая с полученным выше выражением для $f(c_j)$, получаем, что $f(c_j) = a_j g'(c_j)$, откуда $a_j = f(c_j)/g'(c_j)$, что и требовалось. □

5 Вычислительная линейная алгебра

5.1 Системы линейных уравнений и элементарные преобразования

Литература: [F], гл. IV, § 4, п. 5; [K1], гл. 1, § 3, пп. 1, 2.

Пусть R — ассоциативное коммутативное кольцо с единицей. Мы будем называть **системой линейных уравнений** (над R) набор уравнений вида

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \vdots &\quad \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m, \end{aligned}$$

где a_{ij} $(1 \leqslant i \leqslant m,\ 1 \leqslant j \leqslant n)$, b_i $(1 \leqslant i \leqslant m)$ — элементы R, а $x_1, \ldots, x_n$ — неизвестные. **Решением** этой системы линейных уравнений называется набор $(c_1, \ldots, c_n)$ элементов R, при подстановке которого в каждое из m уравнений системы получается верное равенство, то есть, $\sum_{j=1}^{n} a_{ij}c_j = b_i$ для всех $i = 1, \ldots, m$.

В первом приближении линейная алгебра изучает свойства множеств решений систем линейных уравнений. Наша ближайшая цель — указать несколько преобразований, которые не меняют множество решений системы, но, возможно, упрощают ее вид. Чтобы не писать каждый раз значки $+$ и $=$, мы будем пользоваться *матричной формой записи* системы. **Матрицей** указанной системы линейных уравнений называется таблица

$$\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}.$$

Заметим, однако, что матрица системы линейных уравнений содержит не всю информацию о системе: мы нигде не использовали правые части этих уравнений. **Расширенной матрицей** нашей системы линейных уравнений называется таблица

$$\left(\begin{array}{cccc|c} a_{11} & a_{12} & \ldots & a_{1n} & b_1 \\ a_{21} & a_{22} & \ldots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} & b_m \end{array}\right)$$

Вертикальная черта служит для визуального отделения коэффициентов левой части и правой части системы; иногда мы опускаем ее.

Заметим, что в матрице линейной системы с m уравнениями и n неизвестными содержится m строк и n столбцов; на пересечении строки с номером i и столбца с номером j стоит элемент a_{ij}. В расширенной матрице такой системы m строк и $n + 1$ столбец.

Часто мы будем записывать матрицу так: $(a_{ij})_{\substack{1\leqslant i\leqslant m\\ 1\leqslant j\leqslant n}}$: в этой матрице m строк, n столбцов, и на пересечении i-ой строки и j-го столбцы стоит элемент a_{ij}. Если размер матрицы подразумевается известным, мы будем сокращать эту запись до (a_{ij}).

Среди множества преобразований систем линейных уравнений выделяют три несложных типа преобразований, играющих важную роль в нахождении решений.

1. Элементарное преобразование первого типа: прибавить к i-му уравнению j-ое уравнение, умноженное на некоторый элемент $\lambda \in R$. Иными словами, i-ое уравнение

   $$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

   заменяется при этом преобразовании на уравнение

   $$(a_{i1} + \lambda a_{j1})x_1 + (a_{i2} + \lambda a_{j2})x_2 + \cdots + (a_{in} + \lambda a_{jn})x_n = b_i + \lambda b_j,$$

   а все остальные уравнения остаются неизменными.

2. Элементарное преобразование второго типа: поменять местами i-ое уравнение и j-ое уравнение. Остальные уравнения при этом остаются неизменными.

3. Элементарное преобразование третьего типа: домножить i-ое уравнение на обратимый элемент кольца R. Иными словами, для некоторого $\varepsilon \in R^*$ уравнение под номером i

   $$a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n = b_i$$

   заменяется на уравнение

   $$\varepsilon a_{i1}x_1 + \varepsilon a_{i2}x_2 + \cdots + \varepsilon a_{in}x_n = \varepsilon b_i,$$

   а остальные уравнения не меняются.

Несложно понять, как указанные преобразования меняют матрицу системы и расширенную матрицу системы: элементарное преобразование первого типа прибавляет к i-ой строке j-ую, умноженную на $\lambda \in R$; второго типа — меняет местами строки с номерами i и j; третьего типа — домножает все элементы i-ой строки на $\varepsilon \in R^*$.

Мы будем использовать следующие условные обозначения для элементарных преобразований: преобразование первого типа, прибавляющее к i-ой строке j-ую, умноженную на λ, обозначается через $T_{ij}(\lambda)$ (здесь $1 \leqslant i, j \leqslant m$, $i \neq j$, $\lambda \in R$); преобразование второго типа, меняющее местами строки с номерами i и j, обозначается через S_{ij} (здесь $1 \leqslant i, j \leqslant m$, $i \neq j$), а преобразование третьего типа, домножающее i-ую строку на ε, обозначается через $D_i(\varepsilon)$ (здесь $1 \leqslant i \leqslant m$, $\varepsilon \in R^*$). Через некоторое время эти символы превратятся в обозначения совершенно конкретных объектов, связанных с соответствующими преобразованиями.

Сразу же заметим, что каждое элементарное преобразование *обратимо*: это означает, что для каждого элементарного преобразования найдется другое элементарное преобразование (называемое *обратным* такое, что применение двух этих преобразований подряд (в любом

порядке) к системе не меняет ее. Действительно, сразу видно, что для преобразования третьего типа $D_i(\varepsilon)$ обратным является $D_i(\varepsilon^{-1})$, а для преобразования второго типа S_{ij} обратным является оно само. Наконец, несложная выкладка показывает, что для преобразования первого типа $T_{ij}(\lambda)$ обратным является преобразование $T_{ij}(-\lambda)$: последовательное применение этих двух преобразований сначала прибавляет к i-му уравнению исходной системы j-ое, умноженное на λ, а потом прибавляет j-ое, умноженное на $-\lambda$ (или наоборот), поэтому i-ое уравнение в итоге не изменяется (а остальные — тем более).

Лемма 5.1.1. *Элементарные преобразования не меняют множества (всех) решений системы.*

Доказательство. По замечанию выше, каждое элементарное преобразование обратимо; поэтому достаточно доказать, что множество решений системы не уменьшается: если набор $(c_1, \ldots, c_n)$ является решением системы, то он будет являться и решением системы, полученной из нее элементарным преобразованием. Это очевидно для преобразований второго и третьего типов, и несложно проверить для преобразований первого типа. □

5.2 Метод Гаусса

Литература: [F], гл. IV, § 4, п. 5; [K1], гл. 1, § 3, п. 3.

Сейчас мы опишем, как решать произвольную систему линейных уравнений *над полем*. Основная идея состоит в том, чтобы сначала привести систему к удобному для решения виду — *ступенчатому*. Алгоритм приведения произвольной системы к ступенчатому виду называется *методом Гаусса*. Мы дадим строгое определение ступенчатого вида после того, как опишем этот алгоритм.

Как обычно, нам будет удобно работать не с системой линейных уравнений, а с ее [расширенной] матрицей: метод Гаусса состоит в последовательном применении к расширенной матрице системы элементарных преобразований, после чего матрица становится *ступенчатой*, и все решения соответствующей системы легко выписать; по лемме 5.1.1 полученное множество решений будет и множеством решений исходной системы.

Итак, пусть (a_{ij}) — матрица над полем k размера $m \times n$. Мы будем изучать ее столбцы последовательно, слева направо. Возьмем первый столбец. Возможны два варианта: либо он состоит из одних нулей, либо в нем найдется ненулевой элемент. Если столбец состоит из одних нулей, мы пропускаем его и переходим к следующему столбцу, пока не найдем какой-нибудь столбец с ненулевым элементом. Пусть, наконец, в столбце с номером j_1 нашелся ненулевой элемент (если такого столбца нет, то наша матрица нулевая, и алгоритм завершен).

Для начала поставим этот ненулевой элемент на первое место в столбце посредством элементарного преобразования второго типа. Теперь мы сделаем все остальные элементы нашего столбца нулевыми с помощью элементарных преобразований первого типа. Делается это так: теперь мы считаем, что элемент a_{1,j_1} не равен нулю; если какой-нибудь элемент a_{i,j_1} первого столбца также не равен нулю, то прибавим к i-ой строчке первую, умноженную на $-a_{i,j_1}/a_{1,j_1}$. Иными словами, проведем элементарное преобразование $T_{i,j_1}(-a_{i,j_1}/a_{1,j_1})$. При этом

изменится только i-ая строчка, и ее первый элемент станет равным $a_{i,j_1} + a_{1,j_1} \cdot (-a_{i,j_1}/a_{1,j_1}) = 0$. Проделаем это для всех ненулевых элементов первого столбца. Заметим, что здесь мы использовали тот факт, что ненулевой элемент a_{1,j_1} обратим, то есть, что k является полем.

Теперь столбец с номером j_1 нашей матрицы содержит единственный ненулевой элемент a_{1,j_1} (а все столбцы, стоящие слева от него, нулевые). Мысленно забудем про первую строчку нашей матрицы и про все столбцы вплоть до столбца с номером j_1 и повторим нашу операцию: теперь мы берем столбец с номером $j_1 + 1$ и ищем в нем ненулевой элемент, не принимая во внимание первую строчку. Если во всех позициях (кроме, может быть, первой) этого столбцы стоят нули, мы двигаемся дальше вправо, пока не находим, наконец, столбец с номером j_2, в котором стоит какой-нибудь ненулевой элемент не в первой строчке. Посредством элементарного преобразования второго типа можно поставить этот ненулевой элемент на второе место, а затем, с помощью элементарных преобразований первого типа, добиться того, что все элементы ниже его станут нулями. Заметим, что первая строчка в этих преобразованиях уже никак не участвует, поэтому про нее и можно забыть. Кроме того, в столбцах с номерами $1, \ldots, j_1$ стоят нули на тех позициях, которые затрагиваются этими преобразованиями, поэтому они не изменяются. Итак, в столбце с номером j_2 теперь стоит неизвестно что на первой позиции, ненулевой элемент a_{2,j_2} на второй позиции, и 0 на остальных позициях. Далее, конечно, мы продолжаем ту же процедуру, забывая про первый две строчки и про столбцы с номерами $1, \ldots, j_2$. Заметим, что мы обязаны двигаться вправо: $1 \leqslant j_1 < j_2 < j_3 < \ldots$, поэтому этот процесс остановится.

Полученная матрица

$$\begin{pmatrix}
0 & \ldots & 0 & a_{1,j_1} & * & \ldots & * & * & * & \ldots & * & * & * & \ldots & * & * & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & a_{2,j_2} & * & \ldots & * & * & * & \ldots & * & * & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & a_{3,j_3} & * & \ldots & * & * & * & \ldots & * \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & a_{s,j_s} & * & \ldots & * \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 \\
\vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0 & 0 & 0 & \ldots & 0
\end{pmatrix}$$

и называется ступенчатой; теперь мы готовы дать формальное определение.

Определение 5.2.1. Матрица $(a_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$ называется **ступенчатой**, если существует некоторая последовательность индексов $1 \leqslant j_1 < j_2 < \cdots < j_s \leqslant n$ такая, что

- $a_{i,j_i} \neq 0$ для любого $i = 1, \ldots, s$;
- $a_{i,j} = 0$ при $j < j_i$;
- $a_{i,j} = 0$ для любого j при $i > s$.

Иными словами, в ступенчатой матрице имеются строки $1, \ldots, s$ такие, что в строке с номером i первый ненулевой элемент стоит в позиции (i, j_i), а все строки с номерами $s+1, \ldots, m$ — нулевые.

Ненулевые элементы $a_{1,j_1}, a_{2,j_2}, \ldots, a_{s,j_s}$ в ступенчатой матрице (a_{ij}) мы будем называть **ведущими**.

Что же нам дает применение метода Гаусса к расширенной матрице системы линейных уравнений? Напомним, что расширенная матрица системы состоит из m строк и $n+1$ столбца, где m — число уравнений, n — число неизвестных. Самый правый столбец расширенной матрицы несет особый смысл — это правая часть системы. Поэтому сразу рассмотрим особый случай: предположим, что ведущий элемент оказался в последнем столбце. Очевидно, что это может быть только последний ведущий элемент a_{s,j_s}. Тогда уравнение с номером s выглядит так: $0x_1 + \cdots + 0x_n = a_{s,j_s}$, и $a_{s,j_s} \neq 0$. Очевидно, что это уравнение не имеет решений, поэтому и вся система не имеет решений.

Теперь можно считать, что $j_s < n+1$, и всем ведущим элементам соответствуют переменные $x_{j_1}, \ldots, x_{j_s}$. *Все остальные* переменные мы будем называть **свободными**, а переменные $x_{j_1}, \ldots, x_{j_s}$ — **зависимыми**. Теперь мы утверждаем, что множество решений полученной системы выглядит так: свободные переменные могут принимать произвольные значения, и, как только они заданы, значения зависимых переменных определяются однозначным образом.

Действительно, предположим, что мы задали произвольные значения свободных переменных. Пойдем по уравнениям снизу вверх и начнем выражать значения зависимых переменных. Заметим, что уравнения с номерами $s+1, \ldots, m$ фактически имеют вид $0 = 0$, поэтому не влияют на множество решений системы, и их можно выбросить. Последнее уравнение имеет вид $a_{s,j_s}x_{j_s} + \cdots = b_s$, и значения всех переменных в левой части, кроме x_{j_s}, уже заданы. Деля на ненулевой элемент a_{s,j_s} и перенося «многоточие» в правую часть, получаем выражение для зависимой переменной x_{j_s}. Теперь возьмем предпоследнее уравнение: $a_{s-1,j_{s-1}}x_{j_{s-1}} + \cdots = b_{s-1}$; мы уже знаем значения всех переменных в левой части, кроме $x_{j_{s-1}}$, поэтому аналогичным образом получаем выражение для следующей зависимой переменной, $x_{j_{s-1}}$. Продолжая этот процесс, мы дойдем и до первой строчки, выразив значение x_{j_1}.

Итак, если заданы значения свободных переменных, то значения свободных переменных определяются однозначно. С другой стороны, значения свободных переменных могут быть совершенно произвольными, и приведенный алгоритм утверждает, что найдется решение с такими значениями свободных переменных. Иными словами, мы установили взаимно-однозначное соответствие между множеством решений нашей системы и множеством произвольных наборов значений независимых переменных.

5.3 Операции над матрицами

Литература: [F], гл. IV, § 1; [K1], гл. 3, § 3, пп. 1–3.

Определение 5.3.1. Матрицей над кольцом R мы будем называть прямоугольную таблицу, составленную из элементов кольца R. Иными словами, задать матрицу A — значит, задать набор элементов $a_{ij} \in R$ для всех i, j таких, что $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$. Эти элементы называются **коэффициентами** матрицы A и мы пишем $A = (a_{ij})$. При этом мы будем изображать такую матрицу в виде таблицы из m строк и n столбцов, в которой на пересечении i-й строки и j-го столбца стоит элемент a_{ij}. Будем говорить, что A является матрицей $m \times n$; множество

всех матриц $m \times n$ над кольцом R обозначается через $M(m, n, R)$. Если $m = n$ (число строк совпадает с числом столбцов), матрица называется **квадратной**; мы будем писать $M(n, R)$ вместо $M(n, n, R)$. При этом n называется **порядком** квадратной матрицы из $M(n, R)$.

Элемент, стоящий в матрице A на пересечении i-й строки и j-го столбца мы часто будем обозначать через A_{ij}; будем говорить, что в матрице A элемент A_{ij} **стоит на позиции** (i, j).

Введем основные операции над матрицами. Если $A = (a_{ij})$, $B = (b_{ij})$ — две матрицы одинакового размера $m \times n$, определим их сумму $A + B$ как матрицу, у которой на позиции (i, j) стоит $a_{ij} + b_{ij}$. Иными словами, $(A + B)_{ij} = A_{ij} + B_{ij}$ для всех $1 \leqslant i \leqslant m$, $\leqslant i \leqslant n$. Таким образом, сложение матриц происходит *покомпонентно*.

Гораздо интереснее выглядит умножение матриц. Пусть $A \in M(m, n, R)$, $B \in M(n, p, R)$ — обратите внимание, что число столбцов первой матрицы равно числу строк второй матрицы. Тогда их произведением AB называется матрица размера $m \times p$, у которой на позиции (i, k) стоит $\sum_{j=1}^{n} A_{ij}B_{jk}$. Иными словами, $(AB)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk}$. Обратите внимание, что при фиксированных i и k элементы A_{ij} пробегают строку матрицы A с номером i, а элементы B_{jk} пробегают столбец матрицы B с номером k. То есть, для того, чтобы получить элемент, стоящий в матрице AB на позиции (i, k), нужно взять i-ю строку матрицы A, k-й столбец матрицы B, и сформировать сумму произведений соответствующих элементов этой строки и этого столбца; по условию на размер матриц A и B они имеют одинаковую длину.

Определим также результат умножения скаляра (элемента кольца R) на матрицу над R: пусть $\lambda \in R$, $A \in M(m, n, R)$. Рассмотрим матрицу, в которой на позиции (i, j) стоит λA_{ij}; мы будем обозначать ее через λA. То есть, при умножении матрицы A на скаляр λ каждый элемент матрицы A умножается на λ (здесь мы предполагаем, что кольцо R коммутативно, поэтому неважно, с какой стороны происходит умножение).

Наконец, еще одна важная операция — **транспонирование** матрицы. Пусть $A \in M(m, n, R)$. Определим матрицу $A^T \in M(n, m, R)$ так: у нее в позиции (j, i) стоит элемент A_{ij}. Такая матрица называется матрицей, транспонированной к матрице A. Неформально говоря, это матрица, полученная из матрицы A «симметрией» относительно главной диагонали. При этом строки с номерами $1, 2, \ldots, m$ матрицы A становятся столбцами с номерами $1, 2, \ldots, m$ матрицы A^T; аналогично, столбцы матрицы A превращаются в строки матрицы A^T.

Теперь сформулируем свойства введенных операций.

Теорема 5.3.2 (Свойства операций над матрицами). *Следующие тождества выполняются для любых матриц A, B, C над коммутативным кольцом R и для любых $\lambda, \mu \in R$, если определены результаты всех входящих в них операций:*

1. $A + (B + C) = (A + B) + C$ *(ассоциативность сложения);*
2. *пусть 0 — матрица, все коэффициенты которой нулевые; тогда $A + 0 = 0 + A = A$ (нейтральный элемент относительно сложения);*
3. *для любой матрицы A найдется матрица $-A$ такая, что $A + (-A) = (-A) + A = 0$ (противоположный элемент);*

4. $A + B = B + A$ *(коммутативность сложения).*

5. $(AB)C = A(BC)$ *(ассоциативность умножения);*

6. $A(B + C) = AB + AC$ *(левая дистрибутивность);*

7. $(B + C)A = BA + CA$ *(правая дистрибутивность);*

8. $\lambda(A + B) = \lambda A + \lambda B$ *(левая дистрибутивность умножения на скаляр);*

9. $(\lambda + \mu)A = \lambda A + \mu A$ *(правая дистрибутивность умножения на скаляр);*

10. $(\lambda A)B = \lambda(AB) = A(\lambda B)$;

11. $(\lambda\mu)A = \lambda(\mu A)$;

12. $(A + B)^T = A^T + B^T$;

13. $(AB)^T = B^T A^T$.

Поясним формулировку «. . . если определены результаты всех входящих в них операций»: мы можем сложить две матрицы только в том случае, если они имеют одинаковый размер, и перемножить две матрицы только в том случае, если количество столбцов первой матрицы совпадает с количеством строк второй матрицы. Поэтому, скажем, тождество $A + (B + C) = (A + B) + C$ выполняется для любых $A, B, C \in M(m, n, R)$, тождество $(AB)C = A(BC)$ — для любых $A \in M(m, n, R)$, $B \in M(n, p, R)$, $C \in M(p, q, R)$, тождество $A(B + C) = AB + AC$ — для любых $A \in M(m, n, R)$ и $B, C \in M(n, p, R)$, и так далее.

Доказательство. Напоминаем, что через A_{ij} мы обозначаем элемент матрицы A, стоящий в позиции (i, j). Для того, чтобы проверить равенство двух матриц, достаточно проверить, что они имеют одинаковый размер и что элементы, стоящие в соответствующих позициях этих матриц, равны. Мы займемся именно проверкой поэлементного равенства, оставив читателю [тривиальную] проверку равенства размеров.

1. $(A + (B + C))_{ij} = A_{ij} + (B + C)_{ij} = A_{ij} + (B_{ij} + C_{ij}) = (A_{ij} + B_{ij}) + C_{ij} = (A + B)_{ij} + C_{ij} = ((A + B) + C)_{ij}$; здесь мы воспользовались ассоциативностью сложения в кольце R.

2. $(A + 0)_{ij} = A_{ij} + 0_{ij} = A_{ij} + 0 = A_{ij} = 0 + A_{ij} = 0_{ij} + A_{ij} = (0 + A)_{ij}$.

3. Составим матрицу $-A$ из элементов $-A_{ij}$, то есть, положим $(-A)_{ij} = -A_{ij}$. Тогда $(A+(-A))_{ij} = A_{ij}+(-A)_{ij} = A_{ij}-A_{ij} = 0$, откуда $A+(-A) = 0$; аналогично, $(-A)+A = 0$.

4. $(A + B)_{ij} = A_{ij} + B_{ij} = B_{ij} + A_{ij} = (B + A)_{ij}$, поскольку сложение в R коммутативно.

5. Пусть $A \in M(m, n, R)$, $B \in M(n, p, R)$, $C \in M(p, q, R)$. Тогда

$$((AB)C)_{il} = \sum_{k=1}^{p} (AB)_{ik} C_{kl} = \sum_{k=1}^{p} \sum_{j=1}^{n} A_{ij} B_{jk} C_{kl};$$

с другой стороны,

$$(A(BC))_{il} = \sum_{j=1}^{n} A_{ij}(BC)_{jl} = \sum_{j=1}^{n} A_{ij} \sum_{k=1}^{p} B_{jk}C_{kl} = \sum_{j=1}^{n}\sum_{k=1}^{p} A_{ij}B_{jk}C_{kl}.$$

Получившиеся суммы отличаются только изменением порядка суммирования.

6. Пусть $A \in M(m, n, R)$, $B \in M(n, p, R)$. Тогда

$$(A(B + C))_{ik} = \sum_{j=1}^{n} A_{ij}(B + C)_{jk} = \sum_{j=1}^{n}(A_{ij}B_{jk} + A_{ij}C_{jk})$$

и

$$(AB + AC)_{ik} = (AB)_{ik} + (AC)_{ik} = \sum_{j=1}^{n} A_{ij}B_{jk} + \sum_{j=1}^{n} A_{ij}C_{jk} = \sum_{j=1}^{n}(A_{ij}B_{jk} + A_{ij}C_{jk}).$$

7. Доказательство совершенно аналогично доказательству предыдущего пункта.

8. $(\lambda(A + B))_{ij} = \lambda(A + B)_{ij} = \lambda(A_{ij} + B_{ij}) = \lambda A_{ij} + \lambda B_{ij} = (\lambda A)_{ij} + (\lambda B)_{ij} = (\lambda A + \lambda B)_{ij}$.

9. $((\lambda + \mu)A)_{ij} = (\lambda + \mu)A_{ij} = \lambda A_{ij} + \mu A_{ij} = (\lambda A)_{ij} + (\mu A)_{ij} = (\lambda A + \mu A)_{ij}$.

10. Заметим, что $((\lambda A)B)_{ik} = \sum_j((\lambda A)_{ij}B_{jk}) = \sum_j(\lambda A_{ij}B_{jk})$; кроме того,

$$(A(\lambda B))_{ik} = \sum_j(A_{ij}(\lambda B)_{jk}) = \sum_j(A_{ij}\lambda B_{jk}) = \sum_j(\lambda A_{ij}B_{jk})$$

и

$$(\lambda(AB))_{ik} = \lambda(AB)_{ik} = \lambda\sum_j(A_{ij}B_{jk}) = \sum_j(\lambda A_{ij}B_{jk}).$$

11. $((\lambda\mu)A)_{ij} = (\lambda\mu)A_{ij} = \lambda\mu A_{ij} = \lambda(\mu A_{ij}) = \lambda(\mu A)_{ij} = (\lambda(\mu A))_{ij}$.

12. $((A + B)^T)_{ij} = (A + B)_{ji} = A_{ji} + B_{ji} = (A^T)_{ij} + (B^T)_{ij} = (A^T + B^T)_{ij}$.

13. $((AB)^T)_{ik} = (AB)_{ki} = \sum_j(A_{kj}B_{ji}) = \sum_j((A^T)_{jk}(B^T)_{ij}) = \sum_j((B^T)_{ij}(A^T)_{jk}) = B^TA^T$.

□

Определение 5.3.3. Рассмотрим матрицу размера $n \times n$, у которой в позиции (i, j) стоит 1, если $i = j$, и 0, если $i \neq j$. Такая матрица называется **единичной матрицей** и обозначается через E_n (и часто мы будем обозначать ее просто через E, если размер ясен из контекста). Эта матрица действительно играет роль нейтрального элемента относительно умножения, как показывает следующее утверждение.

Предложение 5.3.4. *Пусть* $A \in M(m, n, R)$. *Тогда* $E_m \cdot A = A \cdot E_n = A$.

Доказательство. Заметим, что $(E_m \cdot A)_{ik} = \sum_j (E_m)_{ij} A_{jk}$. В получившейся сумме матричный элемент $(E_m)_{ij}$ равен 0 для всех j, кроме $j = i$. Поэтому от суммы остается одно слагаемое, соответствующее случаю $j = i$, и равное A_{ik}. Это выполнено для всех i, k, поэтому $E_m \cdot A = A$. Второе равенство доказывается аналогично. □

Замечание 5.3.5. Заметим, что для квадратных матриц фиксированного размера (то есть, для элементов $M(n, R)$) свойства 1–7 из теоремы 5.3.2 и свойство единичных матриц из предложения 5.3.4 означают, что эти матрицы образуют ассоциативное кольцо с единицей. Это кольцо $M(n, R)$ называется **кольцом квадратных матриц** порядка n. Отметим, что это кольцо не является коммутативным при $n \geqslant 2$:

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

Напомним, что элемент a произвольного ассоциативного кольца A с единицей называется *обратимым*, если найдется элемент $b \in A$ такой, что $ab = ba = 1$ в A. Такой элемент b обозначается через a^{-1} и называется *обратным* к a. В полном соответствии с этим, квадратная матрица $A \in M(n, R)$ называется **обратимой**, если найдется матрица, обозначаемая через $A^{-1} \in M(n, R)$, такая, что $A \cdot A^{-1} = A^{-1} \cdot A = E_n$. При этом, как и в произвольном ассоциативном кольце с единицей, для обратимой матрицы A выполнено $(A^{-1})^{-1} = A$, а для набора обратимых матриц $A_1, \ldots, A_s$ выполнено $(A_1 \cdot A_2 \cdot \cdots \cdot A_s)^{-1} = A_s^{-1} \cdot \cdots \cdot A_2^{-1} \cdot A_1^{-1}$.

Упомянем еще одно важное свойство, связывающее обратимость и транспонирование.

Предложение 5.3.6. *Если матрица $A \in M(n, R)$ обратима, то и матрица A^T обратима, причем $(A^T)^{-1} = (A^{-1})^T$.*

Доказательство. Пользуясь свойством (13) из теоремы 5.3.2, получаем $A^T \cdot (A^{-1})^T = (A^{-1} \cdot A)^T = (E_n)^T$. Осталось заметить, что $(E_n)^T = E_n$, поскольку из определения единичной матрицы легко видеть, что $(E_n)_{ij} = (E_n)_{ji}$ для всех i, j. Равенство $(A^{-1})^T \cdot A^T = E_n$ проверяется аналогично. □

Замечание 5.3.7. Кольцо матриц $M(n, R)$ не является полем при $n \geqslant 2$, поскольку в нем есть делители нуля. Например, пусть $A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \in M(2, R)$; тогда $A \cdot A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$. Поэтому матрица A никак не может быть обратимой в $M(2, R)$. Нетрудно придумать аналогичный пример в $M(n, R)$ для любого $n \geqslant 2$.

Удобно конструировать матрицы из маленьких кусочков: обозначим через e_{ij} матрицу из $M(m, n, R)$, у которой в позиции (i, j) стоит 1, а во всех остальных позициях стоит 0. Заметим, что m и n в наше обозначение e_{ij} не входят — мы подразумеваем, что всегда из контекста ясно, какого размера матрицы рассматриваются (если это вообще важно). Любую матрицу $A = (a_{ij}) \in M(m, n, R)$ тогда можно представить в виде $A = \sum_{i,j} a_{ij} e_{ij}$. Например, для единичной матрицы имеем $E_n = e_{11} + e_{22} + \cdots + e_{nn}$. Матрицы e_{ij} называются **матричными единицами** (не путать с *единичными матрицами*!)

Как перемножаются матричные единицы? В произведении $e_{ij} \cdot e_{kl}$ ненулевые элементы могут стоять только в i-ой строчке (поскольку все строчки матрицы e_{ij}, кроме i-ой, нулевые), и только в l-ом столбце (поскольку все столбцы матрицы e_{kl}, кроме l-го, нулевые). Поэтому произведение $e_{ij} \cdot e_{kl}$ может отличаться от нуля только в позиции e_{il}. Внимательное рассмотрение произведения i-ой строчки матрицы e_{ij} на l-й столбец матрицы e_{kl} показывает, что

$$e_{ij} \cdot e_{kl} = \begin{cases} e_{il}, & \text{если } j = k; \\ 0, & \text{если } j \neq k. \end{cases}$$

Наконец, докажем полезный критерий равенства двух матриц.

Предложение 5.3.8. *Пусть* $A, B \in M(m, n, R)$. *Следующие утверждения равносильны:*

1. $A = B$;
2. $uA = uB$ *для всех* $u \in M(1, m, R)$;
3. $Av = Bv$ *для всех* $v \in M(n, 1, R)$;
4. $uAv = uBv$ *для всех* $u \in M(1, m, R)$, $v \in M(n, 1, R)$.

Доказательство. Пусть $A = (a_{ij})$, $B = (b_{ij})$. Очевидно, что из первого утверждения следуют остальные. Докажем, что (2) $\Rightarrow$ (1). Возьмем в качестве u матрицу $e_{1,i}$. Тогда $uA = \begin{pmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{pmatrix}$, $uB = \begin{pmatrix} b_{i1} & b_{i2} & \dots & b_{in} \end{pmatrix}$, и из их равенства следует равенство i-х строчек матриц A и B. Подставляя $i = 1, \dots, m$, получаем, что $A = B$.

Совершенно аналогично доказывается, что (3) $\Rightarrow$ (1). Наконец, покажем, что (4) $\Rightarrow$ (1). Достаточно заметить, что если $u = e_{1,i}$ и $v = e_{j,1}$ то $uAv = a_{ij}$ и $uBv = b_{ij}$; подставляя всевозможные пары (i, j), получаем, что $A = B$. □

5.4 Матрицы элементарных преобразований

Литература: [K1], гл. 1, § 3, п. 6.

В качестве первого применения операций над матрицами мы истолкуем элементарные преобразования, введенные в разделе 5.1, как домножения на матрицы определенного вида.

Для $i \neq j$ $(1 \leqslant i, j \leqslant n)$ и $\lambda \in R$ определим $T_{ij}(\lambda) = E_n + \lambda e_{ij}$. Это матрица, которая отличается от единичной матрицы лишь в одной позиции (i, j), в которой стоит λ. Напомним, что по этим же данным i, j, λ мы определили элементарное преобразование первого типа как прибавление к i-й строке матрицы ее j-ой строки, умноженной на λ. Оказывается, проведение этого элементарного преобразования над матрицей $A \in M(n, m, R)$ равносильно умножению матрицы A слева на $T_{ij}(\lambda)$. Действительно, пусть $A = (a_{ij}) \in M(n, m, R)$. Посмотрим на матрицу $T_{ij}(\lambda)A$. Поскольку матрица T_{ij} отличается от матрицы E_n только в i-й строке, произведение $T_{ij}(\lambda)A$ отличается от матрицы A только в i-й строке. Значит, нам осталось только перемножить i-ю строку матрицы $T_{ij}(\lambda)$ на A, и записать результат в i-ю строку результата. В i-й строке матрицы $T_{ij}(\lambda)$ лишь два элемента отличны от нуля: элемент в

позиции i равен 1, а элемент в позиции j равен λ. При умножении на k-й столбец матрицы A, получаем следующее:

$$\begin{pmatrix} 0 & \cdots & 1 & \cdots & \lambda & \cdots & 0 \end{pmatrix} \cdot \begin{pmatrix} a_{1k} \\ \vdots \\ a_{ik} \\ \vdots \\ a_{jk} \\ \vdots \\ a_{nk} \end{pmatrix} = a_{ik} + \lambda a_{jk}$$

Это происходит в каждом столбце матрицы A; поэтому i-я строка произведения $T_{ij}(\lambda)$ равна $(a_{i1} + \lambda a_{j1} \quad \cdots \quad a_{in} + \lambda a_{jn})$, то есть, равна сумме i-й строки матрицы A и j-й строки матрицы A, умноженной на λ.

Теперь разберемся с элементарными преобразованиями второго типа. Для индексов $i \neq j$ рассмотрим матрицу $S_{ij} \in M(n, R)$, которая отличается от единичной матрицы E_n перестановкой строк с номерами i и j. Таким образом, S_{ij} отличается от E_n в четырех позициях: в позициях (i, i) и (j, j) стоят 0 (вместо 1), а в позициях (i, j) и (j, i) стоят 1 (вместо 0). Иными словами, $S_{ij} = E_n - e_{ii} - e_{jj} + e_{ij} + e_{ji}$. Покажем, что умножение матрицы A на S_{ij} слева равносильно элементарному преобразованию второго типа матрицы A — перестановке i-ой и j-ой строчки. Действительно, произведение $S_{ij}A$ отличается от матрицы A только в строчках с номерами i и j: i-ая строчка равна произведению строчки $(0 \quad \cdots \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0)$ (где 1 стоит на j-м месте) на матрицу A, то есть, j-ой строчке матрицы A. Аналогично, j-ая строчка произведения $S_{ij}A$ равна произведению строчки $(0 \quad \cdot \quad 0 \quad 1 \quad 0 \quad \cdots \quad 0)$ (где 1 стоит на i-м месте) на матрицу A, то есть, i-ой строчке матрицы A.

Наконец, для индекса i и обратимого элемента $\varepsilon \in R^*$ рассмотрим матрицу $D_i(\varepsilon) \in M(n, R)$, которая отличается от единичной матрицы E_n лишь в позиции (i, i), где стоит ε. То есть, $D_i(\varepsilon) = E_n + (\varepsilon - 1)e_{ii}$. Покажем, что умножение матрицы A на $D_i(\varepsilon)$ слева равносильно элементарному преобразованию третьего типа матрицы A — умножению i-ой строчки на ε. Действительно, матрица $D_i(\varepsilon) \cdot A$ отличается от A только в i-й строчке, и i-ая строчка матрицы $D_i(\varepsilon) \cdot A$ равна произведению $(\begin{pmatrix} 0 & \cdots & \varepsilon & \cdots & 0 \end{pmatrix}) \cdot A = \varepsilon(\begin{pmatrix} 0 & \cdots & 1 & \cdots & 0 \end{pmatrix}) \cdot A$, что равно произведению ε и i-ой строчки матрицы A.

Таким образом, мы истолковали элементарные преобразования над строками матрицы как домножения слева на несложные матрицы $T_{ij}(\lambda)$, S_{ij} и $D_i(\varepsilon)$:

- умножение на $T_{ij}(\lambda)$ слева соответствует прибавлению к i-ой строчке j-ой строчки, умноженной на λ;
- умножение на S_{ij} слева соответствует перестановке i-ой и j-ой строчек;
- умножение на $D_i(\varepsilon)$ слева соответствует умножению i-ой строчки на ε.

Применяя транспонирование (с учетом свойства $(AB)^T = B^T A^T$), получаем, что элементарные преобразования над *столбцами* матрицы соответствуют домножения *справа* на эти же матрицы: действительно, при транспонировании строки матриц превращаются в столбцы, и $(T_{ij}(\lambda))^T = T_{ji}(\lambda)$, $(S_{ij})^T = S_{ij}$, $(D_i(\varepsilon))^T = D_i(\varepsilon)$. Поэтому

- умножение на $T_{ij}(\lambda)$ справа соответствует прибавлению к j-ому столбцу i-ого столбца, умноженного на λ;
- умножение на S_{ij} справа соответствует перестановке i-ого и j-ого столбцов;
- умножение на $D_i(\varepsilon)$ справа соответствует умножению i-ого столбца на ε.

Заметим, что обратимость элементарных преобразований соответствует тому факту, что любая матрица элементарного преобразования обратима. Так, $(T_{ij}(\lambda))^{-1} = T_{ij}(-\lambda)$, $(S_{ij})^{-1} = S_{ij}$ и $(D_i(\varepsilon))^{-1} = D_i(\varepsilon^{-1})$. Теперь это можно проверить непосредственным матричным перемножением.

Теперь мы можем истолковать метод Гаусса как некоторый матричный факт. Напомним, что метод Гаусса говорит, что с помощью элементарных преобразований строк можно любую матрицу привести к ступенчатому виду. В терминах матриц это означает, что для любой матрицы $A \in M(m, n, k)$ над полем k найдутся матрицы элементарных преобразований $P_1, \ldots, P_s \in M(m, k)$ такие, что матрица $P_s P_{s-1} \ldots P_1 A$ является ступенчатой.

Проведем после этого некоторые элементарные преобразования над *столбцами*. Посмотрим на первую строчку ступенчатой матрицы $A = (a_{ij})$.

$$\begin{pmatrix} 0 & \ldots & 0 & 1 & * & \ldots & * \\ 0 & \ldots & 0 & 0 & * & \ldots & * \\ \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \ldots & 0 & 0 & * & \ldots & * \end{pmatrix}$$

Здесь 1 стоит в позиции $(1, j_1)$, и $a_{1,j} = 0$ при $j < j_1$. Для каждого $j > j_1$ прибавим к j-му столбцу столбец с номером j_1, умноженный на $-a_{1,j}$. После этого в позиции $(1, j)$ окажется $a_{1,j} - a_{1,j} = 0$. То есть, после таких прибавлений первая строчка нашей матрицы будет иметь только один ненулевой элемент — 1 в позиции $(1, j_1)$. Продолжим эту операцию: посмотрим на вторую строчку нашей матрицы. Если она отличается от нулевой, то там стоит 1 в некоторой позиции $(2, j_2)$. Прибавим к j-му столбцу столбец с номером j_2, умноженный на $-a_{2,j}$. При этом первая строчка нашей матрицы уже никак не изменится, а во второй останется лишь один ненулевой элемент — 2 в позиции $(2, j_2)$. Совершив аналогичное действие для всех строк нашей матрицы, мы можем добиться того, что наша матрица отличается от нулевой лишь в позициях $(1, j_1), (2, j_2), \ldots (r, j_r)$, где стоят единицы. После этого перестановкой столбцов можно добиться того, что эти единицы будут стоять в позициях $(1, 1), (2, 2), \ldots (r, r)$. Полученная матрица называется **окаймленной единичной** матрицей. Можно изобразить ее в блочной форме следующим образом:

$$\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

(здесь E_r — единичная матрица размера $r \times r$, а нулевые блоки имеют размеры $r \times (n-r)$, $(m-r) \times r$ и $(m-r) \times (n-r)$). Конечно, возможно, что $r = 0$ и наша матрица нулевая.

Сформулируем то, что было сделано, на матричном языке. Как мы знаем, элементарные перестановки столбцов соответствуют домножениям нашей матрицы на матрицы элементарных преобразований справа. Поэтому на самом деле мы только что доказали следующую теорему:

Теорема 5.4.1. *Для любой матрицы $A \in M(m,n,k)$ над полем k найдутся матрицы элементарных преобразований $P_1, \ldots, P_t, Q_1, \ldots, Q_s$ такие, что*

$$P_t P_{t-1} \ldots P_1 A Q_1 \ldots Q_{s-1} Q_s = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$$

для некоторого r.

Следствие 5.4.2. *Для любой матрицы $A \in M(m,n,k)$ над полем k существуют обратимые матрицы $P \in M(m,k)$, $Q \in M(n,k)$ такие, что $A = PDQ$, где $D = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} \in M(m,n,k)$ — окаймленная единичная матрица. Более того, матрицы P и Q являются произведениями матриц элементарных преобразований.*

Доказательство. По теореме 5.4.1 можно записать $P_t P_{t-1} \ldots P_1 A Q_1 \ldots Q_{s-1} Q_s = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$. Обозначим правую часть через D — это окаймленная единичная матрица. Все матрицы P_i, Q_j обратимы, поэтому можно последовательно домножить на обратные к ним с соответствующих сторон и получить равенство $A = P_1^{-1} \ldots P_t^{-1} D Q_s^{-1} \ldots Q_1^{-1}$. Положим теперь $P = P_1^{-1} \ldots P_t^{-1}$, $Q = Q_s^{-1} \ldots Q_1^{-1}$; матрицы P и Q обратимы, поскольку они являются произведениями обратимых матриц. Получим $A = PDQ$, что и требовалось. □

Заметим, что набор матриц $P_1, \ldots, P_s, Q_1, \ldots, Q_t$ из теоремы не является однозначно определенным. В то же время (хотя мы этого пока не доказали) натуральное число r, полученной по матрице A, определено однозначно: если взять другие матрицы элементарных преобразований, после домножения на которые матрица A превратится в окаймленную единичную, то размер этой единичной матрицы все равно окажется равным r. Это число r является важной характеристикой матрицы A и называется ее *рангом*. Пока что отметим, что для квадратной матрицы A обратимость равносильна тому, что окаймленная единичная матрица, к которой приводится матрица A, на самом деле является единичной:

Следствие 5.4.3. *Пусть квадратная матрица $A \in M(n,k)$ над полем k представлена в виде $A = P_s P_{s-1} \ldots P_1 \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q_1 \ldots Q_{t-1} Q_t$, где P_i, Q_i — матрицы элементарных преобразований. Тогда обратимость матрицы A равносильна тому, что $r = n$.*

Иными словами, матрица A обратима тогда и только тогда, когда ее можно представить в виде произведения матриц элементарных преобразований.

Доказательство. Если $r = n$, то в середине разложения A стоит единичная матрица, которую можно вычеркнуть, и получится, что A является произведением матриц элементарных преобразований. Каждая из матриц элементарных преобразований обратима, а произведение обратимых элементов кольца обратимо (лемма 2.8.10).

Обратно, предположим, что A обратима. Из равенства

$$A = P_s P_{s-1} \dots P_1 \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q_1 \dots Q_{t-1} Q_t$$

получаем, что

$$P_1^{-1} \dots P_{s-1}^{-1} P_s^{-1} A Q_t^{-1} Q_{t-1}^{-1} \dots Q_1^{-1} = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}.$$

Опять же, в левой части стоит произведение обратимых матриц, поэтому и матрица в правой части должна быть обратимой. Но матрица вида $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ может быть обратимой только при $r = n$. Действительно, если $r < n$, то у нее последняя строка равна нулю, и в любом произведении этой матрицы на другую последняя строка также нулевая; поэтому это произведение не может быть единичной матрицей. □

5.5 Блочные матрицы

При работе с большими матрицами часто удобно разбивать их на кусочки поменьше. Мы видели это в теореме 5.4.1: окаймленная единичная матрица размера $m \times n$ и ранга r имеет вид $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$. Вообще, пусть $m = m_1 + \dots + m_s$, $n = n_1 + \dots + n_t$ — разбиения чисел m и n в сумму s и t слагаемых, соответственно. Тогда матрица $A \in M(m, n, R)$ разбивается на st матриц с размерами $m_i \times n_j$: мы группируем первые m_1 строк, следующие m_2 строк, и так далее; а также первые n_1 столбцов, следующие n_2, и так далее. Обозначим эти блоки через $x_{ij} \in M(m_i, n_j, R)$ для $i = 1, \dots, s$, $j = 1, \dots, t$. Матрица с выбранными разбиениями множеств строк и столбцов называется **блочной матрицей** указание разбиений строк и столбцов называется **блочной структурой**. Например, в приведенном выше примере окаймленная единичная матрица имеет вид $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$. в соответствии с разбиениями $m = r + (m - r)$, $n = r + (n - r)$.

Пусть теперь $B \in M(m, n, R)$ — еще одна матрица того же размера, что и A, и пусть для B выбраны те же разбиения $m = m_1 + \dots + m_s$, $n = n_1 + \dots + n_t$; таким образом, у матрицы B есть блоки $y_{ij} \in M(m_i, n_j, R)$. Посмотрим на сумму $A + B$. Это снова матрица из $M(m, n, R)$. Можно и ее разбить на блоки тем же образом и получить блоки $z_{ij} \in M(m_i, n_r, R)$. Нетрудно понять, что $z_{ij} = x_{ij} + y_{ij}$ для всех $i = 1, \dots, s$, $j = 1, \dots, t$. Иными словами, блочные матрицы с одной и той же блочной структурой складываются «поблочно».

Посмотрим теперь, как перемножаются блочные матрицы. Пусть $A \in M(m, n, R)$, $B \in M(n, p, R)$, и пусть выбраны разбиения чисел m, n, p: $m = m_1 + \dots + m_s$, $n = n_1 + \dots + n_t$, $p = p_1 + \dots + p_u$. Тогда A является блочной матрицей с блоками, скажем, $x_{ij} \in M(m_i, n_j, R)$, а

B — блочной матрицей с блоками $y_{jk} \in M(n_j, p_k, R)$. Их произведение AB лежит в $M(m, p, R)$, и его можно рассмотреть как блочную матрицу в соответствии с указанными разбиениями чисел m и p. Блоки матрицы AB обозначим через $z_{ik} \in M(m_i, p_k, R)$. Как блок z_{ik} связан с блоками матриц A и B? Оказывается

$$z_{ik} = x_{i1}y_{1k} + \cdots + x_{it}y_{tk} = \sum_{j=1}^{t} x_{ij}y_{jk}.$$

Таким образом, блочные матрицы можно перемножать «поблочно», и формула для каждого блока в произведении выглядит точно так же, как формула для элемента в произведении матриц. Обратите внимание, однако, что теперь в этом произведении элементы x_{ij} и y_{jk} являются матрицами, так что мы должны следить за порядком, в котором они перемножаются.

5.6 Перестановки

Литература: [F], гл. IV, § 2, п. 2.

Нам необходимо на время отвлечься от линейной алгебры, чтобы ввести важное понятие *группы перестановок*. Пусть X — некоторое множество. **Перестановкой** на множестве X называется биекция $X \to X$. Заметим, что любая биекция обратима: если $\pi\colon X \to X$ — биекция, то существует и обратное отображение $\pi^{-1}\colon X \to X$, также являющееся биекцией, такое, что $\pi \circ \pi^{-1}$ и $\pi^{-1} \circ \pi$ тождественны. Напомним также, что композиция отображений ассоциативна.

Определение 5.6.1. Множество G с бинарной операцией $\circ\colon G \to G$ называется **группой**, если выполняются следующие свойства:

- $a \circ (b \circ c) = (a \circ b) \circ c$ для всех $a, b, c \in G$; (**ассоциативность**);
- существует элемент $e \in G$ (**единичный элемент**) такой, что для любого $a \in G$ выполнено $a \circ e = e \circ a = a$;
- для любого $a \in G$ найдется элемент $a^{-1} \in G$ (называемый **обратным** к a) такой, что $a \circ a^{-1} = a^{-1} \circ a = e$.

Определение 5.6.2. Множество всех биекций из X в X обозначается через $S(X)$ и называется **группой перестановок** множества X. Тождественное отображение $\mathrm{id}_X\colon X \to X$ называется **тождественной перестановкой**.

Как мы заметили выше, $S(X)$ действительно является группой в смысле определения 5.6.1 относительно операции композиции, которая еще называется **умножением** перестановок.

Зачастую нам не важна природа элементов множества X, а важно лишь их количество, особенно если X конечно. Поэтому для каждого натурального n можно рассматривать группу перестановок какого-нибудь выделенного множества из n элементов, например, множества $\{1, \ldots, n\}$. Эта группа обозначается через S_n: $S(\{1, \ldots, n\} = S_n$. Элемент π группы S_n можно

записывать в виде таблицы из двух строк, в первой строке которой стоят числа $1, \ldots, n$ (как правило, в порядке возрастания), а под каждым из них стоит его образ $\pi(1), \ldots, \pi(n)$:

$$\pi = \begin{pmatrix} 1 & 2 & \ldots & n \\ \pi(1) & \pi(2) & \ldots & \pi(n) \end{pmatrix}.$$

Понятно, что по такой записи однозначно восстанавливается элемент π, и обратно, если есть таблица, в первой строке которой стоят числа $1, \ldots, n$, а во второй — те же самые числа в каком-то порядке, то она задает некоторый элемент S_n. Такая запись называется **табличной записью** перестановки. Например, группа S_1 состоит из одного (тождественного) элемента $\begin{pmatrix} 1 \\ 1 \end{pmatrix}$. Группа S_2 состоит из двух элементов: один из них тождественный, $\begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$, а другой переставляет местами 1 и 2: $\begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$. Группа S_3 состоит из шести элементов:

$$S_3 = \left\{ \begin{pmatrix} 1 & 2 & 3 \\ 1 & 2 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}, \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix} \right\}.$$

Несложное комбинаторное рассуждение показывает, что количество элементов в S_n равно $n!$. Действительно, образом элемента 1 может быть любой из n элементов множества $\{1, \ldots, n\}$, образом элемента 2 — любой из оставшихся $n-1$, и так далее; всего получаем $n \cdot (n-1) \cdot \ldots \cdot 1 = n!$ различных вариантов.

Табличная запись позволяет визуализировать перемножение перестановок: для того, чтобы перемножить перестановки π и ρ, нужно записать друг под другом табличные записи π и ρ, переставить столбцы в таблице ρ так, чтобы в первой строке оказалась *вторая* строка таблицы π, и сформировать ответ из первой строки верхней таблицы и второй строки нижней таблицы — это будет табличной записью перестановки $\rho \circ \pi$. Обратите внимание на порядок! Напомним, что мы записываем композицию отображений *справа налево*: запись $\rho \circ \pi$ означает, что мы сначала применяем отображение π, а затем — отображение ρ. Это важно, поскольку при $n \geqslant 3$ умножение в группе S_n некоммутативно. Действительно, рассмотрим перестановки $\pi = \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix}$ и $\rho = \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$. Перемножим их по описанному выше способу:

$$\rho \circ \pi\colon \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \begin{pmatrix} 1 & 3 & 2 \\ 2 & 1 & 3 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 2 & 1 & 3 \end{pmatrix}$$

$$\pi \circ \rho\colon \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 1 & 3 & 2 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 \\ 3 & 2 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 2 & 3 \\ 3 & 2 & 1 \end{pmatrix}$$

Мы получили, что $\rho\circ\pi = \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}$, $\pi\circ\rho = \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}$, и видно, что это разные перестановки: $\rho\circ\pi \neq \pi\circ\rho$.

Сейчас мы покажем, что любая перестановка представляется в виде произведения перестановок простейшего вида. Интуитивно ясно, что простейшей [нетождественной] перестановкой является та, которая лишь меняется местами два элемента, а остальные оставляет на своих местах.

Определение 5.6.3. Пусть $1 \leqslant i, j \leqslant n$ и $i \neq j$. Обозначим через τ_{ij} следующую перестановку:

$$\begin{cases}\tau_{ij}(i) &= j,\\ \tau_{ij}(j) &= i,\\ \tau_{ij}(k) &= k \text{ при } k \neq i, j.\end{cases}$$

Ее табличная запись выглядит так:

$$\begin{pmatrix}\dots & i & \dots & j & \dots\\ \dots & j & \dots & i & \dots.\end{pmatrix}$$

(подразумевается, что все столбики с многоточиями отвечают *неподвижным* элементам). Такая перестановка называется **транспозицией**. Перестановка вида $\tau_{i,i+1}$ (при $1 \leqslant i \leqslant n-1$) называется **элементарной транспозицией**.

Очевидно, что любая транспозиция τ_{ij} совпадает с τ_{ji} и является обратной к себе самой: $\tau_{ij} = \tau_{ji}$, $\tau_{ij}\circ\tau_{ij} = \mathrm{id}$. Посмотрим, что происходит при умножении перестановки на транспозицию: сравним табличные записи перестановок π и $\pi\circ\tau_{ij}$. Нетрудно видеть, что они различаются только в столбцах с номерами i и j (поскольку τ_{ij} совпадает с тождественной в остальных точках). А именно,

$$\pi = \begin{pmatrix}\dots & i & \dots & j & \dots\\ \dots & \pi(i) & \dots & \pi(j) & \dots\end{pmatrix}, \quad \pi\circ\tau_{ij} = \begin{pmatrix}\dots & i & \dots & j & \dots\\ \dots & \pi(j) & \dots & \pi(i) & \dots\end{pmatrix}.$$

Иными словами, домножение на τ_{ij} справа соответствует перестановке i-ой и j-ой позиций в нижней строке табличной записи перестановки.

Предложение 5.6.4. *Любая перестановка является произведением транспозиций.*

Доказательство. Пусть $\pi \in S_n$. Начнем с тождественной перестановки id и покажем, что последовательным домножением на транспозиции справа можно получить перестановку π. Сначала добьемся того, чтобы на первом месте в нижней строке табличной записи нашей перестановки стояло то, что нужно — то есть, $\pi(1)$. Для этого нужно переставить местами первый столбик с тем, в котором стоит $\pi(1)$ (Конечно, если $\pi(1) = 1$, ничего переставлять и не нужно). После этого поставим на второе место в нижней строке $\pi(2)$: так как π является перестановкой, то $\pi(1) \neq \pi(2)$, поэтому где-то справа от первого столбца есть столбец с $\pi(2)$.

Поменяем его со вторым. И так далее: на k-шаге мы добиваемся того, что первые k чисел в нижней строке нашей перестановки выглядели так: $\pi(1), \pi(2), \ldots, \pi(k)$. В конце концов (дойдя до $k = n$) мы получим перестановку π путем домножения id на транспозиции, что и требовалось. □

Предложение 5.6.5. *Любая транспозиция является произведением нечетного числа элементарных транспозиций.*

Доказательство. Неформально задача выглядит так: нам разрешено менять местами любые два соседних элемента в строке, а хочется поменять местами два элемента, стоящих далеко друг от друга. Как этого добиться? Очень просто: сначала «продвинуть» последовательно левый из этих элементов направо до второго, поменять их там местами, а потом второй элемент «отогнать» обратно на место левого. При этом наши элементы поменяются местами, а все остальные элементы останутся на своих местах: любой элемент между нашими мы затронем ровно два раза: на пути «туда» и на пути «обратно»; сначала он сдвинется на шаг влево, а потом — на шаг вправо. Ну, а любой элемент, стоящий не между нашими, и подавно останется на своем месте. Аккуратный подсчет показывает, что мы совершили нечетное число операций.

Формально же это рассуждение выражается в виде формулы

$$\tau_{ij} = \tau_{i,i+1} \circ \tau_{i+1,i+2} \circ \cdots \circ \tau_{j-2,j-1} \circ \tau_{j-1,j} \circ \tau_{j-2,j-1} \circ \ldots \tau_{i+1,i+2} \circ \tau_{i,i+1}$$

(здесь мы считаем, что $i < j$). Это равенство несложно проверить напрямую, и оно представляет транспозицию τ_{ij} в виде произведения $2(j - i) - 1$ элементарных транспозиций. □

Определение 5.6.6. Пусть $\pi \in S_n$. Говорят, что пара индексов (i, j) образует **инверсию** для перестановки π, если $i < j$ и $\pi(i) > \pi(j)$. Количество пар индексов от 1 до n, образующих инверсию для π, называется **числом инверсий** перестановки π и обозначается через $\mathrm{inv}(\pi)$.

Неформально говоря, число инверсий измеряет «отклонение» перестановки от тождественной: если $\pi = \mathrm{id}$, то для $i < j$ всегда выполнено $\pi(i) = i < j = \pi(j)$, поэтому $\mathrm{inv}(\mathrm{id}) = 0$. Число инверсий — это количество пар элементов, стоящих в «неправильном» порядке. Важнейшей характеристикой перестановки является *четность* ее числа инверсий, которая называется *знаком*:

Определение 5.6.7. Пусть $\pi \in S_n$. Число $(-1)^{\mathrm{inv}(\pi)}$ называется **знаком** перестановки π и обозначается через $\mathrm{sgn}(\pi)$. Иными словами, $\mathrm{sgn}(\pi) = 1$, если $\mathrm{inv}(\pi)$ четно, и $\mathrm{sgn}(\pi) = -1$, если $\mathrm{inv}(\pi)$ нечетно. Перестановка называется **четной**, если $\mathrm{sgn}(\pi) = 1$, и **нечетной**, если $\mathrm{sgn}(\pi) = -1$.

Пример 5.6.8. Единственный элемент в S_1 является четной перестановкой. Одна из двух перестановок в S_2 (тождественная) является четной, а другая — нечетной. Среди шести перестановок в S_3 имеется три четных и три нечетных: четными являются id, $\begin{pmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{pmatrix}$ и $\begin{pmatrix} 1 & 2 & 3 \\ 3 & 1 & 2 \end{pmatrix}$, а нечетными — транспозиции τ_{12}, τ_{13} и τ_{23}.

Оказывается, если перестановка представлена в виде произведения транспозиций, то четность числа этих транспозиций всегда совпадает с четностью перестановки (хотя понятно, что у перестановки может быть много различных представлений в виде произведения транспозиций). Для доказательства этого нам необходимо посмотреть на то, что происходит со знаком при домножении перестановки на транспозицию.

Предложение 5.6.9. *Пусть* $\pi \in S_n$, $\tau_{ij} \in S_n$ — *транспозиция. Тогда* $\operatorname{sgn}(\pi) = -\operatorname{sgn}(\pi \circ \tau_{ij})$.

Доказательство. Посмотрим, как меняется число инверсий перестановки при домножении на *элементарную транспозицию*. Сравним перестановки

$$\pi = \begin{pmatrix} \dots & i & i+1 & \dots \\ \dots & \pi(i) & \pi(i+1) & \dots \end{pmatrix} \text{ и } \pi \circ \tau_{i,i+1} = \begin{pmatrix} \dots & i & i+1 & \dots \\ \dots & \pi(i+1) & \pi(i) & \dots \end{pmatrix}.$$

Заметим, что вне столбцов с номерами i и $i+1$ эти перестановки совпадают, поэтому число инверсий для индексов вне множества $\{i, i+1\}$, у них одинаковое. Далее, если для некоторого $j \notin \{i, i+1\}$ индексы i и j образуют инверсию для π (например, мы имели $j < i$ и $\pi(j) > \pi(i)$), то $i+1$ и j образуют инверсию для $\pi \circ \tau_{i,i+1}$, (поскольку $(\pi \circ \tau_{i,i+1})(i+1) = \pi(i) < \pi(j) = (\pi \circ \tau_{i,i+1})(j)$ и $j < i+1$), и наоборот. Аналогично, если $i+1$ и j образуют инверсию для π, то i и j образуют инверсию для $\pi \circ \tau_{i,i+1}$, и наоборот. Поэтому среди всех пар индексов, кроме пары (i, j), количество инверсий у π и $\pi \circ \tau_{i,i+q}$ одинаковое. Но если $(i, i+1)$ является инверсией для π, то $(i, i+1)$ не является инверсией для $\pi \circ \tau_{i,i+1}$, поскольку значения π и $\pi \circ \tau_{i,i+1}$ на i и $i+1$ поменялись местами. Обратно, если пара $(i, i+1)$ не была инверсией для π, она станет инверсией для $\pi \circ \tau_{i,i+1}$. Значит, число инверсий $\pi \circ \tau_{i,i+1}$ отличается от числа инверсий $\tau_{i,i+1}$ ровно на единицу: $\operatorname{inv}(\pi \circ \tau_{i,i+1}) = \operatorname{inv}(\pi) \pm 1$. Поэтому эти числа имеют разную четность.

Это означает, что при домножении на элементарную транспозицию перестановка меняет знак. По предложению 5.6.5 любую транспозицию можно записать как произведение нечетного числа элементарных, поэтому при домножении на любую транспозицию перестановка меняет знак нечетное число раз — то есть, меняет знак. □

Следствие 5.6.10. *Пусть* $\pi = \tau_1 \circ \dots \circ \tau_s$, *где* $\tau_1, \dots, \tau_s$ — *транспозиции. Тогда* $\operatorname{sgn}(\pi) = (-1)^s$.

Доказательство. Запишем $\pi = \mathrm{id} \circ \tau_1 \circ \dots \circ \tau_s$ и посмотрим на это произведение так: мы начали с тождественной перестановки и s раз домножили на транспозиции справа. Тождественная перестановка является четной, и при каждом домножении знак меняется на противоположный, поэтому итоговый знак равен $(-1)^s$. □

Следствие 5.6.11. *При* $n \geqslant 2$ *в группе* S_n *поровну (по* $n!/2$*) четных и нечетных перестановок.*

Доказательство. Рассмотрим отображение $f\colon S_n \to S_n$, $\pi \mapsto \pi \circ \tau_{12}$. Нетрудно видеть, что это биекция (обратным к этому отображению является оно само: $(f \circ f)(\pi) = f(f(\pi)) = (\pi \circ \tau_{12}) \circ \tau_{12} = \pi$, поэтому $f \circ f = \mathrm{id}_{S_n}$). При этом по предложению 5.6.9 f переводит четные перестановки в нечетные, а нечетные — в четные. Поэтому f устанавливает биекцию между подмножеством четных перестановок и подмножеством нечетных перестановок в S_n. Всего перестановок $n!$, поэтому и четных, и нечетных по $n!/2$. □

Теперь несложно показать, что знак ведет себя мультипликативно:

Теорема 5.6.12. *Пусть* $\pi, \rho \in S_n$*; тогда* $\operatorname{sgn}(\pi \circ \rho) = \operatorname{sgn}(\pi) \cdot \operatorname{sgn}(\rho)$.

Доказательство. Представим π и ρ в виде произведения транспозиций: $\pi = \sigma_1 \circ \cdots \circ \sigma_s$, $\rho = \tau_1 \circ \cdots \circ \tau_t$. По следствию 5.6.10 имеем $\operatorname{sgn}(\pi) = (-1)^s$ и $\operatorname{sgn}(\rho) = (-1)^t$. При этом $\pi \circ \rho = \sigma_1 \circ \cdots \circ \sigma_s \circ \tau_1 \circ \cdots \circ \tau_t$ есть произведение $s + t$ транспозиций, поэтому $\operatorname{sgn}(\pi \circ \rho) = (-1)^{s+t} = (-1)^s \cdot (-1)^t = \operatorname{sgn}(\pi) \cdot \operatorname{sgn}(\rho)$. □

Следствие 5.6.13. *Пусть* $\pi \in S_n$*; тогда* $\operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\pi)$.

Доказательство. Заметим, что $\pi \circ \pi^{-1} = \mathrm{id}$, поэтому $\operatorname{sgn}(\pi) \cdot \operatorname{sgn}(\pi^{-1}) = \operatorname{sgn}(\mathrm{id}) = 1$. □

5.7 Определитель

Литература: [F], гл. IV, § 2, пп. 1, 3, 4; [K1], гл. 3, § 1; [vdW], гл. 4, § 25.

Теперь все готово, чтобы ввести интересный инвариант квадратной матрицы.

Определение 5.7.1. Пусть $A = (a_{ij}) \in M(n, k)$ — квадратная матрица над полем k. Ее **определителем** (или **детерминантом**) называется следующий элемент поля k:

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \cdot a_{1,\pi(1)} \cdot a_{2,\pi(2)} \cdot \cdots \cdot a_{n,\pi(n)} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i,\pi(i)}.$$

Мы будем также использовать обозначение $|A| = \det(A)$.

Примеры 5.7.2.

- Определитель матрицы 1×1: в этом случае в сумме из определения $\det(A)$ всего одно слагаемое, и знак тождественной перестановки равен 1, поэтому $\det(\begin{pmatrix} a_{11} \end{pmatrix}) = a_{11}$.
- Определитель матрицы 2×2: $S_2 = \{\mathrm{id}, \tau_{12}\}$, причем $\operatorname{sgn}(\mathrm{id}) = 1$, $\operatorname{sgn}(\tau_{12}) = -1$, поэтому

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

- Определитель матрицы 3×3:

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{31}a_{22} - a_{11}a_{23}a_{32}.$$

Выясним простейшие свойства определителя.

Предложение 5.7.3. *Пусть* $A \in M(n, k)$*; тогда* $\det(A^T) = \det(A)$.

Доказательство. Посмотрим на формулу для определителя матрицы $A = (a_{ij})$. В слагаемом, соответствующем перестановке π, перемножаются элементы вида $a_{i,\pi(i)}$, то есть, элементы вида a_{ij} для $j = \pi(i)$. Заметим, что $j = \pi(i)$ тогда и только тогда, когда $\pi^{-1}(j) = i$. Иными словами, в рассматриваемом слагаемом перемножаются элементы вида $a_{\pi^{-1}(j),j}$ для всех $j = 1, \dots, n$. Поэтому мы можем записать

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i,\pi(i)} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{j=1}^{n} a_{\pi^{-1}(j),j} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{j=1}^{n} a_{\pi(j),j}.$$

В последнем равенстве мы воспользовались тем фактом, что если π пробегает всю группу S_n, то и π^{-1} пробегает всю S_n; кроме того, $\operatorname{sgn}(\pi) = \operatorname{sgn}(\pi^{-1})$, поэтому можно заменить суммирование по всем π на суммирование по всем π^{-1}. Но последнее выражение совпадает с формулой для $\det(A^T)$: элемент матрицы A, стоящий в позиции $(\pi(j), j)$ — это в точности элемент матрицы A^T, стоящий в позиции $(j, \pi(j))$. □

Следующие свойства определителя касаются его зависимость от различных операций над строками. Пусть $A = (a_{ij}) \in M(n, k)$ — квадратная матрица, $(a'_{i1}, a'_{i2}, \dots, a'_{in})$ — некоторая строка. Рассмотрим матрицу A', полученную заменой i-ой строки матрицы A на строку $(a'_{i1}, a'_{i2}, \dots, a'_{in})$, и матрицу A'', полученную заменой i-ой строки матрицы A на строку $(a_{i1} + a'_{i1}, a_{i2} + a'_{i2}, \dots, a_{in} + a'_{in})$. Схематично мы будем изображать это так:

$$A = \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}, A' = \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ a'_{i1} & a'_{i2} & \dots & a'_{in} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix},$$
$$A'' = \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ a_{i1} + a'_{i1} & a_{i2} + a'_{i2} & \dots & a_{in} + a'_{in} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}.$$

Здесь многоточия символизируют тот факт, что все три матрицы A, A', A'' совпадают за пределами i-й строки. Оказывается, что определитель ведет себя **аддитивно** по отношению к строкам матрицы: $\det(A'') = \det(A) + \det(A')$. Иными словами, если представить какую-нибудь строку матрицы в виде суммы двух строк, то определитель исходной матрицы будет равен сумме определителей матриц, в которых эта строка заменена на строки-слагаемые. Нам будет удобнее записывать это следующим образом: обозначим $u = (a_{i1}, a_{i2}, \dots, a_{in})$, $v = (a'_{i1}, a'_{i2}, \dots, a'_{in})$ (таким образом, $u, v \in M(1, n, k)$ — две строки длины n). Тогда

$$\begin{vmatrix} \vdots \\ u + v \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ u \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ v \\ \vdots \end{vmatrix}$$

(здесь $u + v$ обозначает [покомпонентную] сумму строк u и v, и снова подразумевается, что в остальных позициях эти три матрицы совпадают).

Посмотрим на формулу для определителя матрицы A'':

$$\det(A'') = \sum_{\pi\in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \dots (a_{i,\pi(i)} + a'_{i,\pi(i)}) \dots a_{n,\pi(n)}$$

(здесь мы воспользовались тем, что в i-ой строке матрицы A'' стоят суммы соответствующих элементов i-х строк матриц A и A'). Каждое слагаемое выписанной суммы в силу дистрибутивности распадается на два слагаемых, в одно из которых входит $a_{i,\pi(i)}$, а в другое — $a'_{i,\pi(i)}$:

$$\begin{aligned}\det(A'') &= \sum_{\pi\in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a_{i,\pi(i)} \dots a_{n,\pi(n)} + \operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a'_{i,\pi(i)}) \dots a_{n,\pi(n)}\right) \\ &= \sum_{\pi\in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a_{i,\pi(i)} \dots a_{n,\pi(n)}\right) + \sum_{\pi\in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a'_{i,\pi(i)} \dots a_{n,\pi(n)}\right).\end{aligned}$$

Первое из полученных слагаемых в точности равно $\det(A)$, а второе равно $\det(A')$, поэтому $\det(A'') = \det(A) + \det(A')$, что и требовалось.

Кроме того, если все элементы некоторой строки умножить на $\lambda \in k$, то и определитель матрицы умножится на λ. Точнее, рассмотрим матрицу $A = (a_{ij}) \in M(n,k)$ и заменим в ней i-ю строку $(a_{i1}, a_{i2}, \dots, a_{in})$ на строку $(\lambda a_{i1}, \lambda a_{i2}, \dots, \lambda a_{in})$. Обозначим полученную матрицу через A'. Тогда $\det(A') = \lambda \det(A)$. Действительно, определитель матрицы A' равен

$$\det(A') = \sum_{\pi\in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \dots (\lambda a_{i,\pi(i)}) \dots a_{n,\pi(n)}\right).$$

В каждом слагаемом полученной суммы присутствует множитель λ. После вынесения его за скобки получаем

$$\det(A') = \lambda \left(\sum_{\pi\in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a_{i,\pi(i)} \dots a_{n,\pi(n)}\right) = \lambda \det(A).$$

Доказанные два свойства в совокупности называют **линейностью** определителя по строкам. Кроме того, определитель обладает **кососимметричностью** по строкам: если две строки матрицы $A = (a_{ij}) \in M(n,k)$ совпадают, то ее определитель равен нулю. То есть, если найдутся такие индексы $i \neq j$, что $a_{il} = a_{jl}$ для всех $l = 1, \dots, n$, то $\det(A) = 0$. Конечно, кососимметричность имеет смысл только при $n \geqslant 2$.

Для доказательства кососимметричности заметим сначала, что отображение $f\colon S_n \to S_n$, $\pi \mapsto f \circ \tau_{ij}$ является биекцией и меняет четность перестановок. Мы уже видели такое отображение в доказательстве следствия 5.6.11 для частного случая $\{i,j\} = \{1,2\}$. Значит, ограничив должным образом отображение f, мы получаем биекцию между множеством всех четных и множеством всех нечетных перестановок. Обозначим множество всех четных перестановок из S_n через A_n, и для краткости будем писать τ вместо τ_{ij}. Получаем биекцию $A_n \to S_n \backslash A_n$, $\pi \mapsto f \circ \tau$, которую мы обозначим также через f. Теперь вернемся к нашей

матрице $A = (a_{ij}) \in M(n,k)$, в которой i-ая строка совпадает с j-ой. Запишем определитель матрицы A:

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a_{i,\pi(i)} \dots a_{j,\pi(j)} \dots a_{n,\pi(n)}.$$

Теперь при помощи биекции f разобьем все слагаемые на пары, поставив в одну пару слагаемые, соответствующие перестановкам $\pi \in A_n$ и $f(\pi) = \pi \circ \tau \in S_n \backslash A_n$:

$$\det(A) = \sum_{\pi \in A_n} \big(\operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a_{i,\pi(i)} \dots a_{n,\pi(n)} + \operatorname{sgn}(\pi \circ \tau) a_{1,(\pi\circ\tau)(1)} \dots a_{i,(\pi\circ\tau)(i)} \dots a_{j,(\pi\circ\tau)(j)} \dots a_{n,(\pi\circ\tau)(n)} \big).$$

Осталось заметить, что $\operatorname{sgn}(\pi \circ \tau) = -\operatorname{sgn}(\pi)$, $a_{i,(\pi\circ\tau)(i)} = a_{i,\pi(j)} = a_{j,\pi(j)}$, $a_{j,(\pi\circ\tau)(j)} = a_{j,\pi(i)} = a_{i,\pi(i)}$ и $a_{k,(\pi\circ\tau)(k)} = a_{k,\pi(k)}$ для всех $k \neq i,j$. Поэтому сумма двух слагаемых в каждой паре равна 0, а с ней и весь $\det(A)$.

Стало быть, нами доказана следующая теорема.

Теорема 5.7.4. *Определитель линейно и кососимметрично зависит от строк матрицы. Иными словами,*

$$\begin{vmatrix} \vdots \\ u+v \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ u \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ v \\ \vdots \end{vmatrix}, \quad \begin{vmatrix} \vdots \\ \lambda u \\ \vdots \end{vmatrix} = \lambda \begin{vmatrix} \vdots \\ u \\ \vdots \end{vmatrix}, \quad \begin{vmatrix} \vdots \\ u \\ \vdots \\ u \\ \vdots \end{vmatrix} = 0.$$

Кроме того, определитель линейно и кососимметрично зависит от столбцов матрицы.

Доказательство. Утверждение для строк доказано выше; утверждение для столбцов получается транспонированием матрицы. □

Теперь нетрудно понять, как меняется определитель при элементарных преобразованиях строк и столбцов.

Теорема 5.7.5. *Определитель матрицы не меняется при элементарном преобразовании (строк или столбцов) первого типа, меняет знак при элементарном преобразовании второго типа, и умножается на ε при элементарном преобразовании $D_i(\varepsilon)$ третьего типа. На матричном языке:*

$$|T_{ij}(\lambda)A| = |AT_{ij}(\lambda)| = |A|, \quad |S_{ij}A| = |AS_{ij}| = -|A|, \quad |D_i(\varepsilon)A| = |AD_i(\varepsilon)| = \varepsilon|A|.$$

Доказательство. Как всегда, мы проведем доказательство только для элементарных преобразований строк. Рассмотрим элементарное преобразование первого типа и воспользуемся

линейностью:

$$\begin{vmatrix} \vdots \\ u + \lambda v \\ \vdots \\ v \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ u \\ \vdots \\ v \\ \vdots \end{vmatrix} + \lambda \begin{vmatrix} \vdots \\ v \\ \vdots \\ v \\ \vdots \end{vmatrix}.$$

Заметим, что первое слагаемое результата — это определитель исходной матрицы, а второе слагаемое равно нулю в силу кососимметричности.

Посмотрим на элементарные преобразования второго типа. Для любых строк u, v длины n выполнено

$$0 = \begin{vmatrix} \vdots \\ u + v \\ \vdots \\ u + v \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ u \\ \vdots \\ u \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ u \\ \vdots \\ v \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ v \\ \vdots \\ u \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ v \\ \vdots \\ v \\ \vdots \end{vmatrix} = \begin{vmatrix} \vdots \\ u \\ \vdots \\ v \\ \vdots \end{vmatrix} + \begin{vmatrix} \vdots \\ v \\ \vdots \\ u \\ \vdots \end{vmatrix},$$

откуда

$$\begin{vmatrix} \vdots \\ u \\ \vdots \\ v \\ \vdots \end{vmatrix} = - \begin{vmatrix} \vdots \\ v \\ \vdots \\ u \\ \vdots \end{vmatrix}.$$

Это и означает, что элементарное преобразование второго типа меняет знак определителя. Наконец, для элементарных преобразований третьего типа утверждение теоремы напрямую следует из линейности определителя. □

5.8 Дальнейшие свойства определителя

Литература: [K1], гл. 3, § 2, п. 2; [vdW], гл. 4, § 19.

Теорема 5.8.1 (Определитель блочной верхнетреугольной матрицы). *Пусть матрица* $A \in M(n, k)$ *имеет вид* $A = \begin{pmatrix} B & X \\ 0 & C \end{pmatrix}$, *где* $B \in M(m, k)$, $C \in M(n - m, k)$, $X \in M(m, n, k)$. *Тогда* $|A| = |B| \cdot |C|$.

Доказательство. Мы знаем, что $\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \dots a_{m,\pi(m)} a_{m+1,\pi(m+1)} \dots a_{n,\pi(n)}$. По предположению, $a_{ij} = 0$, если $i > m$ и $j \leqslant m$. Поэтому некоторые слагаемые в этой сумме равны 0. Покажем, что ненулевое слагаемое не может содержать и множителей из блока X, то есть, не может включать в себя множитель a_{ij} для $i \leqslant m$, $j > m$. Действительно, посмотрим на некоторое ненулевое слагаемое $a_{1,\pi(1)} \dots a_{m,\pi(m)} a_{m+1,\pi(m+1)} \dots a_{n,\pi(n)}$, соответствующее перестановке π. Среди чисел $\pi(1), \dots, \pi(n)$ должны встречаться по разу числа $1, \dots, m$. Если

некоторое число $j \leqslant m$ равно $\pi(i)$, то обязательно должно быть $i \leqslant m$, поскольку, по предположению, $a_{ij} = 0$ при $i > m$ и $j \leqslant m$. Значит, все числа $1, \ldots, m$ встречаются среди чисел $\pi(1), \ldots, \pi(m)$. Но тех и других поровну, значит, $\pi(i) \leqslant m$ для любого $i \leqslant m$. Стало быть, $\pi(i) > m$ для любого $i > m$. Мы получили, что наше слагаемое содержит лишь множители вида a_{ij}, где либо $i, j \leqslant m$, либо $i, j > m$. В частности, матричных элементов из блока X среди них не встречается.

Таким образом, на самом деле суммирование в $\det(A)$ производится по тем перестановкам π, которые действуют «отдельно» на наборах $1, \ldots, m$ и $m+1, \ldots, n$, не переставляя числа из разных наборов. Поэтому каждая такая перестановка однозначно определяет две перестановки: на числах $1, \ldots, m$ и на числах $m+1, \ldots, n$. Обозначим первую из них через ρ, а вторую сдвинем на m влево (чтобы получить перестановку чисел $1, \ldots, n-m$, то есть, элемент из S_{n-m}) и обозначим через σ. По перестановке π мы построили пару перестановок $\rho \in S_m$, $\sigma \in S_{n-m}$.

Посмотрим теперь на произведение $\det(B) \cdot \det(C)$. Это

$$\left(\sum_{\rho \in S_m} \operatorname{sgn}(\rho) a_{1,\rho(1)} \ldots a_{m,\rho(m)}\right) \cdot \left(\sum_{\sigma \in S_{n-m}} \operatorname{sgn}(\sigma) a_{m+1,m+\sigma(1)} \ldots a_{n,m+\sigma(n-m)}\right).$$

При раскрытии скобок в этом произведении получим сумму слагаемых вида

$$\operatorname{sgn}(\rho)\operatorname{sgn}(\sigma) a_{1,\rho(1)} \ldots a_{m,\rho(m)} a_{m+1,m+\sigma(1)} \ldots a_{n,m+\sigma(n-m)}$$

для всех пар перестановок $\rho \in S_m$, $\sigma \in S_{n-m}$. По каждой такой паре перестановок построим перестановку $\pi \in S_n$, подействовав перестановкой ρ на числах $1, \ldots, m$ и перестановкой σ (сдвинутой на m вправо) на числах $m+1, \ldots, n$.

Теперь видно, что в формулах для $\det(A)$ и $\det(B) \cdot \det(C)$ происходит суммирование по всем парам перестановок $(\rho, \sigma) \in S_m \times S_{n-m}$ слагаемых одинакового вида. Осталось лишь проверить совпадение знаков: в первой формуле мы видим $\operatorname{sgn}(\pi)$, а во второй — произведение $\operatorname{sgn}(\rho) \cdot \operatorname{sgn}(\sigma)$. Но нетрудно видеть, что число инверсий в перестановке π равно сумме чисел инверсий в соответствующих им перестановках ρ и σ: нет никаких инверсий между числами из набора $1, \ldots, m$ и числами из набора $m+1, \ldots, n$. □

Следствие 5.8.2. *Определитель верхнетреугольной матрицы равен произведению ее диагональных элементов:*

$$\left|\begin{pmatrix} a_1 & * & * & \ldots & * \\ 0 & a_2 & * & \ldots & * \\ 0 & 0 & a_3 & \ldots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & a_n \end{pmatrix}\right| = a_1 a_2 \ldots a_n.$$

В частности, определитель единичной матрицы E_n равен 1.

Доказательство. Это несложно получить из предыдущей теоремы индукцией по размеру матрицы. Можно и напрямую заметить, что в сумме из определения $\det(A)$ для верхнетреугольной матрицы A лишь одно слагаемое отлично от нуля — то, которое отвечает тождественной перестановке. □

Предложение 5.8.3. *Если в матрице присутствует нулевой столбец или нулевая строка, то ее определитель равен нулю.*

Доказательство. Пусть i-ая строка матрицы A равна нулю. В каждое слагаемое из определения $\det(A)$ входит элемент вида $a_{i,\pi(i)}$, равный нулю, поэтому каждое слагаемое равно нулю. Доказательство для нулевого столбца получается транспонированием. □

Предложение 5.8.4. *Определители матриц элементарных преобразований:* $|T_{ij}(\lambda)| = 1$, $|S_{ij}| = -1$, $|D_i(\varepsilon)| = \varepsilon$. *Определитель окаймленной единичной матрицы размера* $n \times n$*:*
$$\begin{vmatrix} E_r & 0 \\ 0 & 0 \end{vmatrix} = \begin{cases} 0, & \text{если } r < n; \\ 1, & \text{если } r = n \end{cases}.$$

Доказательство. Матрица элементарных преобразований приводится к единичной одним элементарным преобразованием, и мы знаем, как при этом меняется ее определитель, поэтому первая часть — тривиальное вычисление. Окаймленная единичная матрица является верхнетреугольной, поэтому вторая часть сразу следует из следствия 5.8.2. □

Теорема 5.8.5 (Мультипликативность определителя). *Определитель произведения матриц равен произведению их определителей:*

$$\det(AB) = \det(A)\det(B) \qquad \text{для любых } A, B \in M(n,k).$$

Доказательство. Заметим, что для любой матрицы $C \in M(n,k)$ выполнены равенства

$$\begin{aligned} \det(T_{ij}(\lambda)C) &= \det(T_{ij}(\lambda))\det(C), \\ \det(S_{ij}C) &= \det(S_{ij})\det(C), \\ \det(D_i(\varepsilon)C) &= \det(D_i(\varepsilon))\det(C), \\ \det(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} C) &= \det(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix})\det(C). \end{aligned}$$

Действительно, первые три равенства следуют из теоремы 5.7.5 и предложения 5.8.4. При $r < n$ матрица $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} C$ имеет нулевую строку, поэтому ее определитель равен нулю (предложение 5.8.3), как и произведение определителей сомножителей (в силу предложения 5.8.4. При $r = n$ указанная матрица является единичной, поэтому результат следует из следствия 5.8.2.

По следствию 5.4.2 мы можем записать

$$A = P_t \dots P_1 \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q_1 \dots Q_s,$$

где $P_1, \dots, P_t, Q_1, \dots, Q_s$ — матрицы элементарных преобразований. Тогда

$$\det(AB) = \det(P_t \dots P_1 \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q_1 \dots Q_s B).$$

Применяя замечание из предыдущего абзаца несколько раз, получаем, что

$$\det(AB) = \det(P_t) \dots \det(P_1) \det(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}) \det(Q_1) \dots \det(Q_s) \det(B).$$

С другой стороны,

$$\det(A) = \det(P_t \dots P_1 \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q_1 \dots Q_s),$$

и, снова применяя замечание выше, получаем

$$\det(A) = \det(P_t) \dots \det(P_1) \det(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}) \det(Q_1) \dots \det(Q_s).$$

Сопоставляя полученные равенства, получаем, что $\det(AB) = \det(A)\det(B)$. □

5.9 Разложение определителя по строке

Литература: [F], гл. IV, § 2, п. 5; [K1], гл. 3, § 2.

Посмотрим на матрицу $A \in M(n, k)$. Вычеркнем из нее строку с номером i и столбец с номером j для некоторых $1 \leqslant i, j \leqslant n$. Обозначим полученную матрицу через $M_{ij} \in M(n-1, k)$. Определитель матрицы M_{ij} (а иногда сама эта матрица) называется **(дополнительным) минором**.

Теперь посмотрим на строку с номером i исходной матрицы A и воспользуемся линейностью определителя:

$$|A| = \begin{vmatrix} \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \end{vmatrix} = \begin{vmatrix} \vdots & \vdots & \ddots & \vdots \\ a_{i1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{vmatrix} + \begin{vmatrix} \vdots & \vdots & \ddots & \vdots \\ 0 & a_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \end{vmatrix} + \begin{vmatrix} \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \end{vmatrix}.$$

Посчитаем отдельно определитель каждого слагаемого в правой части. Слагаемое с номером j имеет вид

$$\begin{vmatrix} \ddots & \vdots & \vdots & \vdots & \ddots \\ \dots & 0 & a_{ij} & 0 & \dots \\ \ddots & \vdots & \vdots & \vdots & \ddots \end{vmatrix}:$$

все элементы в i-ой строчке равны нулю, кроме a_{ij}. Теперь аккуратно переставим строчки и столбцы так, чтобы элемент a_{ij} оказался в левом верхнем углу нашей матрицы; для этого нужно сдвинуть по циклу строки с номерами от 1 до i и столбцы с номерами от 1 до j. То есть, сначала поменяем местами строки i и $i-1$, затем строки $i-1$ и $i-2$, и так далее, пока

не поменяем строки 1 и 2. Нетрудно видеть, что мы совершили ровно $i-1$ элементарное преобразоване второго типа. При этом определитель нашей матрицы умножился на $(-1)^{i-1}$. После этого сделаем то же самое со столбцами, и определитель умножится на $(-1)^{j-1}$. В итоге он умножится на $(-1)^{i-1+j-1} = (-1)^{i+j-2} = (-1)^{i+j}$. После таких операций наша матрица будет иметь следующий блочный вид:

$$\begin{pmatrix} a_{ij} & 0 \\ * & M_{ij} \end{pmatrix}.$$

По теореме 5.8.1 (напомним, что определитель не меняется при транспонировании) ее определитель равен произведению a_{ij} на дополнительный минор $|M_{ij}|$. Значит, j-е слагаемое в разложении $\det(A)$, с которого мы начали, равно $(-1)^{i+j}a_{ij}|M_{ij}|$.

Произведение $(-1)^{i+j}|M_{ij}|$ называется **алгебраическим дополнением** элемента a_{ij} и обозначается через $\widetilde{A}_{ij}$. Мы получили **разложение определителя по строке**: $\det(A) = a_{i1}\widetilde{A}_{i1} + a_{i2}\widetilde{A}_{i2} + \cdots + a_{in}\widetilde{A}_{in}$. Транспонируя полученный результат, мы получаем **разложение определителя по столбцу**: $\det(A) = a_{1i}\widetilde{A}_{1i} + a_{2i}\widetilde{A}_{2i} + \cdots + a_{ni}\widetilde{A}_{ni}$.

Сформулируем чуть более общий результат.

Теорема 5.9.1 (Соотношения ортогональности). *Пусть* $A \in M(n,k)$ *и* $1 \leqslant i \leqslant n$. *Тогда*

$$a_{i1}\widetilde{A}_{j1} + a_{i2}\widetilde{A}_{j2} + \cdots + a_{in}\widetilde{A}_{jn} = \begin{cases} \det(A), & \text{если } i = j; \\ 0, & \text{если } i \neq j. \end{cases}.$$

Доказательство. При $i = j$ это в точности разложение определителя по строке. Если же $i \neq j$, рассмотрим матрицу A', которая совпадает с матрицей A везде, кроме строчки с номером j, а в ее строчке с номером j стоит строчка с номером i матрицы A. Таким образом, строки матрицы A' с номерами i и j совпадают, поэтому ее определитель равен нулю. С другой стороны, раскладывая этот определитель по строке с номером j, мы получим в точности сумму $a_{i1}\widetilde{A}_{j1} + a_{i2}\widetilde{A}_{j2} + \cdots + a_{in}\widetilde{A}_{jn}$, поскольку в строке с номером j стоят элементы $a_{i1}, a_{i2}, \ldots, a_{in}$, а их дополнения совпадают с дополнениями элементов j-ой строки матрицы A, поскольку алгебраические дополнения элементов j-ой строки не зависят от того, что именно стоит в j-ой строке. □

Конечно, несложно сформулировать аналогичные соотношения, исходя из разложения определителя по столбцу.

Эту теорему можно записать в более компактной форме. Для этого рассмотрим матрицу $\operatorname{adj}(A)$, в которой на позиции (i,j) стоит алгебраическое дополнение $\widetilde{A}_{ji}$ (обратите внимание на то, что индексы поменялись местами). Она называется **присоединенной** (или **взаимной**) к матрице A. Соотношения ортогональности (для строк и столбцов) тогда переписываются следующим образом.

Следствие 5.9.2. *Для матрицы* $A \in M(n,k)$ *выполнено*

$$A \cdot \operatorname{adj}(A) = \det(A) \cdot E = \operatorname{adj}(A) \cdot A$$

Теперь нетрудно доказать критерий обратимости квадратной матрицы.

Следствие 5.9.3. *Матрица $A \in M(n,k)$ обратима тогда и только тогда, когда $\det(A) \neq 0$; в этом случае $A^{-1} = (\det(A))^{-1} \operatorname{adj}(A)$.*

Доказательство. Если A обратима, то найдется A^{-1} такая, что $A \cdot A^{-1} = E$; тогда

$$\det(A)\det(A^{-1}) = \det(A \cdot A^{-1}) = \det(E) = 1$$

в силу мультипликативности определителя. Обратно, если $\det(A) \neq 0$, то, разделив соотношение ортогональности на скаляр $\det(A)$, получаем, что

$$A \cdot (\det(A))^{-1} \operatorname{adj}(A) = E = (\det(A))^{-1} \operatorname{adj}(A) \cdot A,$$

что и требовалось. □

В частности, для матрицы 2×2 это следствие означает, что

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$$

(если, конечно, $ad - bc \neq 0$).

Применим теперь полученные результаты к решению системы линейных уравнений с невырожденной матрицей. Рассмотрим систему линейных уравнений $AX = B$ с квадратной матрицей $A = (a_{ij}) \in M(n,k)$, где $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ — столбец неизвестных, $B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in M(n,1,k)$ — столбец правой части. Напомним, что *решить систему* — значит, найти все столбцы $X \in M(n,1,k)$, для которых выполнено $AX = B$. Если матрица A невырождена, то есть, существует обратная матрица A^{-1}, после домножения обеих частей уравнения на A^{-1} получаем $A^{-1}AX = A^{-1}B$, что равносильно равенству $X = A^{-1}B$. Таким образом, система уравнений с невырожденной квадратной матрицей всегда имеет единственное решение.

Более того, для нахождения этого решения нетрудно написать чуть более явные формулы, называемые **формулами Крамера**. Действительно,

$$\begin{aligned} X = A^{-1}B = \frac{1}{\det(A)} \operatorname{adj}(A)B &= \frac{1}{\det(A)} \begin{pmatrix} \widetilde{A}_{11} & \widetilde{A}_{21} & \dots & \widetilde{A}_{n1} \\ \widetilde{A}_{12} & \widetilde{A}_{22} & \dots & \widetilde{A}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{A}_{1n} & \widetilde{A}_{2n} & \dots & \widetilde{A}_{nn} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \\ &= \frac{1}{\det(A)} \begin{pmatrix} b_1\widetilde{A}_{11} + b_2\widetilde{A}_{21} + \dots + b_n\widetilde{A}_{n1} \\ b_1\widetilde{A}_{12} + b_2\widetilde{A}_{22} + \dots + b_n\widetilde{A}_{n2} \\ \vdots \\ b_1\widetilde{A}_{1n} + b_2\widetilde{A}_{2n} + \dots + b_n\widetilde{A}_{nn} \end{pmatrix}. \end{aligned}$$

Итоговые выражения очень похожи на разложения определителя по строке. И действительно, заменим в матрице A столбец под номером i на столбец B. Обозначим полученную матрицу через A'_i. Посчитаем определитель этой матрицы, разложив его по i-ому столбцу: для этого нужно перемножать элементы ее i-го столбца (то есть, элементы столбца B) на их алгебраические дополнения, которые совпадают с соответствующими алгебраическими дополнениями элементов матрицы A. Мы получим в точности $b_1\widetilde{A}_{1i} + b_2\widetilde{A}_{2i} + \cdots + b_n\widetilde{A}_{ni}$ — то, что стоит в столбце X на позиции i (с точностью до множителя $1/\det(A)$. Сформулируем полученный результат в виде теоремы.

Теорема 5.9.4 (Формулы Крамера). *Пусть $A \in M(n,k)$ — невырожденная матрица, $B \in M(n,1,k)$ — некоторый столбец. Обозначим через A'_i матрицу, полученную подстановкой столбца B вместо i-го столбца матрицы A. Тогда решение $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ системы линейных уравнений $AX = B$ единственно и задается формулами*

$$x_i = \frac{\det(A'_i)}{\det(A)}.$$

Посмотрим теперь на множество решений произвольной однородной системы линейных уравнений $AX = 0$ с матрицей $A \in M(m,n,k)$; здесь $X = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ — столбец неизвестных, а в правой части стоит нулевая матрица $0 \in M(m,1,k)$.

Предложение 5.9.5 (Свойства решений однородной системы линейных уравнений). *Если $X, X' \in M(n,1,k)$ — решения системы $AX = 0$, то сумма $X + X'$ также является решением этой системы. Если $X \in M(n,1,k)$ — решение системы $AX = 0$, $\lambda \in k$, то $\lambda X \in M(n,1,k)$ также является решением этой системы.*

Доказательство. Если $AX = 0$ и $AX' = 0$, то $A(X + X') = AX + AX' = 0 + 0 = 0$ и $A(\lambda X) = \lambda(AX) = \lambda \cdot 0 = 0$. □

Теперь посмотрим на произвольную систему линейных уравнений $AX = B$ (мы сохраняем предыдущие обозначения; кроме того, $B \in M(m,1,k)$ — некоторый столбец правой части).

Предложение 5.9.6 (Свойства решений неоднородной системы линейных уравнений). *Пусть X_0 — некоторое фиксированное решение системы $AX = B$ Тогда любое решение этой системы имеет вид $X = X_0 + Y$, где Y — некоторое решение соответствующей однородной системы $AX = 0$. Обратно, для любого решения Y однородной системы $AX = 0$ сумма $X = X_0 + Y$ является решением системы $AX = B$.*

Доказательство. Если $AX_0 = B$ и $AY = 0$, то $A(X_0 + Y) = AX_0 + AY = B + 0 = 0$. Обратно, если $AX_0 = B$ и, кроме того, $AX = B$, то $A(X - X_0) = AX - AX_0 = B - B = 0$, поэтому $X - X_0$ является решением соответствующей однородной системы. □

Поэтому поиск решений произвольной системы линейных уравнений $AX = B$ сводится к нахождению *частного решения* X_0 этой системы (если оно вообще существует), и к нахождению всех решений соответствующей однородной системы $AX = 0$. В главе 6 мы построим общую теорию для изучения свойств решений однородных систем, а в главе 7 сформулируем в рамках этой теории и вопрос о существовании частного решения неоднородной системы.

6 Векторные пространства

6.1 Первые определения

Литература: [F], гл. XII, § 1, п. 1, § 2, пп. 1, 2; [K2], гл. 1, § 1; [KM], ч. 1, § 1; [vdW], гл. 4, § 19.

Неформально говоря, векторное пространство — это множество, элементы которого называются векторами, на котором определены операции сложения векторов и умножения вектора на число, причем выполняются некоторые естественные свойства этих операций. Здесь «число» означает произвольный элемент некоторого основного поля k.

Определение 6.1.1. Пусть k — поле. Множество V вместе с операциями $+\colon V \times V \to V$, $\cdot\colon V \times k \to V$ называется **векторным пространством** (точнее — **правым векторным пространством**), если выполняются следующие свойства (называемые *аксиомами векторного пространства*):

1. $(u + v) + w = u + (v + w)$ для любых $u, v, w \in V$ (*ассоциативность сложения*);
2. существует $0 \in V$ такой, что $0 + v = v + 0 = v$ для всех $v \in V$ (*нейтральный элемент по сложению*);
3. для любого $v \in V$ найдется элемент $-v \in V$ такой, что $v + (-v) = (-v) + v = 0$ (*обратный элемент по сложению=противоположный элемент*);
4. $u + v = v + u$ для любых $u, v \in V$ (*коммутативность сложения*);
5. $(u + v)a = u \cdot a + v \cdot a$ для любых $u, v \in V$, $a \in k$ (*левая дистрибутивность*);
6. $u(a + b) = u \cdot a + u \cdot b$ для любых $u \in V$, $a, b \in k$ (*правая дистрибутивность*);
7. $u \cdot (a \cdot b) = (u \cdot a) \cdot b$ для любых $u \in V$, $a, b \in k$ (*внешняя ассоциативность*);
8. $u \cdot 1 = u$ для любого $u \in U$ (*унитальность*).

При этом элементы пространства V называются **векторами**, а элементы поля k — **скалярами**.

Замечание 6.1.2. Заметим, что первые три аксиомы не включают в себя умножение на скаляр и выражают тот факт, что V с операцией сложения является *группой* (см. определение 5.6.1); четвертая аксиома означает, что эта группа коммутативна.

Замечание 6.1.3. Обратите внимание, что знаки $+$ и $\cdot$ в аксиомах используются в разных смыслах: $+$ может означать сложение как в векторном пространстве V, так и в поле k, а $\cdot$ означает умножение скаляра на вектор и умножение скаляров в поле k. Упражнение: про каждый знак $+$ и $\cdot$ в аксиомах векторного пространства скажите, какую именно операцию он обозначает. Символ «0» также используется в дальнейшем в двух смыслах: он может обозначать как нулевой элемент поля, так и нулевой элемент векторного пространства. При желании мы могли бы как-нибудь различать их (некоторые авторы пишут $\overline{0}$ для нулевого вектора), но не будем этого делать, поскольку из контекста всегда ясно, какой элемент имеется в виду (а если не ясно, читатель получает хорошее упражнение).

Замечание 6.1.4. Мы постараемся всегда при умножении вектора на скаляр записывать вектор слева, а вектор справа, то есть, писать $v \cdot a$ для $v \in V$ и $a \in k$. Вместе с тем, можно было бы везде писать $a \cdot v$ вместо $v \cdot a$. Читателю предлагается переписать определение 6.1.1 в таких терминах и убедиться, что получатся совершенно аналогичные аксиомы (за счет коммутативности умножения в поле!) Более щепетильные авторы различают две конвенции в записи и говорят о *правых векторных пространствах* и *левых векторных пространствах*, соответственно. Отметим, что естественное обобщение понятия векторного пространства на произвольные кольца (не обязательно коммутативные) требует строгого различения этих двух понятий.

Примеры 6.1.5.

1. Для натурального n рассмотрим множество всех столбцов высоты n, состоящих из элементов поля k: $k^n = \{\begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \mid a_i \in k\}$. Введем на k^n естественные операции [покомпонентного] сложения и [покомпонентного] умножения на скаляры. Тогда k^n превратится в векторное пространство над полем k: справедливость всех аксиом немедленно следует из свойств операций над матрицами, поскольку можно рассматривать такие столбцы как матрицы $n \times 1$: $k^n = M(n, 1, k)$.

2. Аналогично, множество всех строк длины n над k с покомпонентными операциями сложения и умножения на скаляры образует векторное пространство над k; мы будем обозначать его через ${}^n k$. Альтернативно, ${}^n k = M(1, n, k)$.

3. Обобщая предыдущие примеры, можно заметить, что множество $M(m, n, k)$ всех матриц фиксированного размера $m \times n$ с обычными операциями сложения матриц и умножения на скаляры образует векторное пространство над k.

4. Аналогично первым двум примерам, можно рассмотреть множества столбцов *бесконечной высоты* и строк *бесконечной ширины*, состоящих из элементов поля k. И то, и другое — это просто множество бесконечных последовательностей $a_1, a_2, \ldots$, где все a_i лежат в k. Различие между множеством столбцов и множеством строк лишь в форме записи. Множество таких последовательностей, воспринимаемых как столбцы, мы будем обозначать через k^∞, а множество последовательностей, воспринимаемых как строки — через ${}^\infty k$. На каждом из этих множеств определены операции [покомпонентного] сложения и [покомпонентного] умножения на элементы поля k. Несложно проверить выполнение для них всех свойств из определения 6.1.1, поэтому k^∞ и ${}^\infty k$ являются векторными пространствами над полем k.

5. Пусть E — множество [свободных] векторов на стандартной эвклидовой плоскости. Из школьного курса известно, что сложение векторов и умножение векторов на вещественные числа обладает всеми свойствами из определения векторного пространства. Поэтому E можно рассматривать как векторное пространство над $\mathbb{R}$. Аналогично, множество векторов в трехмерном пространстве является векторным пространством над $\mathbb{R}$.

6. Пусть $k \subseteq L$ — поля. Элементы L можно складывать между собой и умножать на элементы поля k (на самом деле, их можно перемножать и между собой, но мы забудем про эту операцию). Все свойства из определения векторного пространства немедленно следуют из свойств операций в поле. Поэтому L естественным образом является векторным пространством над k. Например, $\mathbb{R}$ — векторное пространство над $\mathbb{Q}$, а $\mathbb{C}$ — векторное пространство над $\mathbb{Q}$ и над $\mathbb{R}$. Кроме того, любое поле является (не очень интересным) векторным пространством над самим собой.

7. Многочлены от одной переменной над полем k можно складывать между собой и умножать на скаляры из k; поэтому $k[x]$ (с естественными операциями) является векторным пространством над k (необходимые аксиомы немедленно следуют из свойств операций в $k[x]$).

Предложение 6.1.6. *Пусть V — векторное пространство над k. Тогда*

1. *$v \cdot 0 = 0$ для любого вектора $v \in V$, где $0 \in k$;*

2. *$0 \cdot a = 0$ для любого скаляра $a \in k$, где 0 — нулевой вектор;*

3. *$v \cdot (-1) = -v$ для любого вектора $v \in V$.*

Доказательство. 1. Заметим, что $v \cdot 0 = v \cdot (0 + 0) = v \cdot 0 + v \cdot 0$. Прибавим к обеим частям $-(v \cdot 0)$; получим $(-v \cdot 0) + v \cdot 0 = (-v \cdot 0) + v \cdot 0 + v \cdot 0$, откуда $0 = 0 + v \cdot 0 = v \cdot 0$, что и требовалось.

2. Заметим, что $0 \cdot a = (0 + 0) \cdot a = 0 \cdot a + 0 \cdot a$. Прибавим к обеим частям $-(0 \cdot a)$; получим $-(0 \cdot a) + 0 \cdot a = -(0 \cdot a) + 0 \cdot a + 0 \cdot a$, откуда $0 = 0 + 0 \cdot a = 0 \cdot a$, что и требовалось.

3. Воспользуемся первой частью: $0 = v \cdot 0 = v \cdot (1 + (-1)) = v \cdot 1 + v \cdot (-1) = v + v \cdot (-1)$. Прибавим к обеим частям $(-v)$; получим $-v = (-v) + v + v \cdot (-1) = 0 + v \cdot (-1) = v \cdot (-1)$. □

6.2 Подпространства

Определение 6.2.1. Пусть V — векторное пространство над полем k. Подмножество $U \subseteq V$ называется **подпространством**, если выполнены следующие условия:

1. $0 \in U$;

2. если $u, v \in U$, то и $u + v \in U$;

3. если $u \in U$, $a \in k$, то $u \cdot a \in U$.

Тот факт, что U является подпространством V, мы будем обозначать так: $U \leqslant V$.

Замечание 6.2.2. Если $U \leqslant V$, то $-u \in U$ для любого $u \in U$. Действительно, для любого $u \in U$ выполнено $-u = u \cdot (-1) \in U$.

Примеры 6.2.3. 1. В любом пространстве V есть «тривиальные» подпространства $0 \leqslant V$ и $V \leqslant V$.

2. Пусть $V = k[x]$, $U = \{f \in k[x] \mid f(1) = 0\}$. Тогда $U \leqslant V$.

3. Пусть $k[x]_{\leqslant n}$ — множество многочленов степени не выше n: $k[x]_{\leqslant n} = \{f \in k[x] \mid \deg(f) \leqslant n\}$. Нетрудно проверить, что $k[x]_{\leqslant n} \leqslant k[x]$.

4. Множество векторов, параллельных некоторой плоскости, является подпространством трехмерного пространства векторов.

Лемма 6.2.4. *Пересечение произвольного набора подпространств пространства V является подпространством в V.*

Доказательство. Пусть $\{U_\alpha\}_{\alpha\in A}$ — подпространства в V. Пусть $u, v \in \bigcap_{\alpha\in A} U_\alpha$. По определению пересечения выполнено $u, v \in U_\alpha$ для всех α. Так как $U_\alpha \leqslant V$, то для каждого α выполнено $u + v \in U_\alpha$, откуда $u + v \in \bigcap_{\alpha\in A} U_\alpha$. Кроме того, если $a \in k$, то для каждого α выполнено $ua \in U_\alpha$, откуда $ua \in \bigcap_{\alpha\in A} U_\alpha$. □

Определение 6.2.5. Пусть $U_1, \ldots, U_m$ — подпространства в V. **Суммой** подпространств $U_1, \ldots, U_m$ называется множество всевозможных сумм элементов $U_1, \ldots, U_m$. Обозначение: $U_1 + \cdots + U_m$. Более точно,

$$U_1 + \cdots + U_m = \{u_1 + \cdots + u_m \mid u_1 \in U_1, \ldots, u_m \in U_m\}.$$

Несложно проверить (упражнение!), что для любых подпространств $U_1, \ldots, U_m$ в V их сумма $U_1 + \cdots + U_m$ также является подпространством в V.

Лемма 6.2.6. *Пусть $U_1, \ldots, U_m$ — подпространства векторного пространства V. Тогда их сумма $U_1 + \cdots + U_m$ — это наименьшее (по включение) векторное подпространство в V, содержащее каждое из подпространств $U_1, \ldots, U_m$.*

Доказательство. Очевидно, что каждое из подпространств $U_1, \ldots, U_m$ содержится в сумме $U_1 + \cdots + U_m$ (достаточно рассмотреть суммы вида $u_1 + \cdots + u_m$, в которых все элементы, кроме одного, равны нулю). С другой стороны, если некоторое подпространство пространства V содержит $U_1, \ldots, U_m$, то оно обязано содержать и все элементы вида $u_1 + \cdots + u_m$ ($u_i \in U_i$), поэтому обязано содержать $U_1 + \cdots + U_m$. □

Итак, любой элемент $u \in U_1 + \cdots + U_m$ можно представить в виде $u = u_1 + \cdots + u_m$ для некоторых $u_i \in U_i$. Нас интересует случай, когда такое представление *единственно*.

Определение 6.2.7. Пусть $U_1, \ldots, U_m$ — подпространства векторного пространства V. Будем говорить, что V является **прямой суммой** подпространств $U_1, \ldots, U_m$, если каждый элемент $v \in V$ можно единственным образом представить в виде суммы $v = u_1 + \cdots + u_m$, где все $u_i \in U_i$. Обозначение: $V = U_1 \oplus \cdots \oplus U_m$ или $V = \bigoplus_{i=1}^m U_i$.

Примеры 6.2.8. 1. Пусть $V = k^3$ — пространство столбцов высоты 3 над полем k, $U = \{\begin{pmatrix} * \\ * \\ 0 \end{pmatrix}\}$ — подпространство столбцов, третья координата которых равна нулю, $W = \{\begin{pmatrix} 0 \\ 0 \\ * \end{pmatrix}\}$ — подпространство столбцов, первые две координаты которых равны нулю. Тогда V является прямой суммой U и W: $V = U \oplus W$.

2. Пусть $V = k^n$ — пространство столбцов высоты n над полем k. Обозначим через U_i подпространство столбцов в V, в которых на всех местах кроме, возможно, i-го, стоит нуль:

$$U_i = \{\begin{pmatrix} 0 \\ \vdots \\ 0 \\ * \\ 0 \\ \vdots \\ 0 \end{pmatrix}\}.$$

Тогда $V = U_1 \oplus \cdots \oplus U_n$.

3. Пусть теперь снова $V = k^3$, U_1 — множество столбцов вида $\begin{pmatrix} a \\ a \\ 0 \end{pmatrix}$, где $a \in k$; U_2 — множество столбцов вида $\begin{pmatrix} b \\ 0 \\ 0 \end{pmatrix}$, где $b \in k$; U_3 — множество столбцов вида $\begin{pmatrix} 0 \\ c \\ d \end{pmatrix}$, где $c, d \in k$. Тогда V *не является* прямой суммой подпространств U_1, U_2, U_3. Дело в том, что столбец вида $\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$ можно разными способами представить в виде суммы трех векторов $u_1 \in U_1$, $u_2 \in U_2$, $u_3 \in U_3$. Действительно, во-первых,

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} + \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -1 \\ 0 \end{pmatrix},$$

а во-вторых, разумеется,

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}.$$

В последнем примере мы показали, что пространство *не является* прямой суммой данных подпространств, предъявив два различных разложения для *нулевого* вектора. Предположим теперь, что у нас есть набор подпространств в V, сумма которых равна V. Следующее предложение показывает, что для доказательства того, что эта сумма прямая, достаточно доказать, что 0 единственным образом представляется в виде суммы векторов из этих подпространств.

Предложение 6.2.9. *Пусть $U_1, \ldots, U_n$ — подпространства в V. Пространство V является прямой суммой этих подпространств тогда и только тогда, когда выполняются два следующих условия:*

1. $V = U_1 + \cdots + U_n$;
2. *если* $0 = u_1 + \cdots + u_n$ *для некоторых* $u_i \in U_i$, *то* $u_1 = \cdots = u_n = 0$.

Доказательство. Предположим сначала, что $V = U_1 \oplus \cdots \oplus V_n$. Тогда по определению $V = U_1 + \cdots + U_n$. Предположим, что $0 = u_1 + \cdots + u_n$, где $u_1 \in U_1, \ldots, u_n \in U_n$. Заметим, что также $0 = 0 + \cdots + 0$, где $0 \in U_1, \ldots, 0 \in U_n$. Из определения прямой суммы теперь следует, что $u_1 = 0, \ldots, u_n = 0$.

Обратно, пусть выполняются два условия выше, и пусть $v \in V$. Из первого условия следует, что мы можем записать $v = u_1 + \cdots + u_n$ для некоторых $u_1 \in U_1, \ldots, u_n \in U_n$. Осталось доказать, что такое представление единственно. Если $v = u_1' + \cdots + u_n'$ для $u_1' \in U_1, \ldots, u_n' \in U_n$, то $0 = v - v = (u_1 - u_1') + \cdots + (u_n - u_n')$, где каждая разность $u_i - u_i'$ лежит в U_i. Из второго условия теперь следует, что $u_i - u_i' = 0$ для всех i, то есть, что два данных разложения на самом деле совпадают. □

Приведем еще один полезный критерий разложения пространства в прямую сумму *двух* подпространств.

Предложение 6.2.10. *Пусть $U, W \leqslant V$. Пространство V является прямой суммой U и W тогда и только тогда, когда $V = U + W$ и $U \cap W = \{0\}$.*

Доказательство. Предположим, что $V = U \oplus W$. Тогда $V = U + W$ по определению прямой суммы. Если $v \in U \cap W$, то можно записать $0 = v + (-v)$, где $v \in U$, $(-v) \in W$. Из единственности представления 0 в виде суммы векторов из U и W теперь следует, что $v = 0$. Поэтому $U \cap W = \{0\}$.

Для доказательства обратного утверждения предположим, что $V = U + W$ и $U \cap W = \{0\}$. Пусть $0 = u + w$, где $u \in U$, $w \in W$. По предложению 6.2.9 нам достаточно доказать, что $u = w = 0$. Но из $0 = u + w$ следует, что $u = -w \in W$, в то время $u \in U$. Значит, $u \in U \cap W$, и потому $u = 0$ и $w = -u = 0$, что и требовалось. □

Замечание 6.2.11. Представьте три прямые U_1, U_2, U_3, проходящие через 0 на эвклидовой плоскости V. Очевидно, что $V = U_1 + U_2 + U_3$ и $U_1 \cap U_2 = U_2 \cap U_3 = U_3 \cap U_1 = \{0\}$. Это значит, что *наивное* обобщение предложения 6.2.10 неверно.

6.3 Линейная зависимость и независимость

Литература: [F], гл. XII, § 1, п. 2; [K2], гл. 1, § 1, п. 2, § 2, п. 1; [KM], ч. 1, § 2; [vdW], гл. 4, § 19.

Определение 6.3.1. Пусть V — векторное пространство над k, $v_1, \ldots, v_n \in V$ и $a_1, \ldots, a_n \in k$. Выражение вида $v_1a_1 + \cdots + v_na_n$ называется **линейной комбинацией** элементов $v_1, \ldots, v_n$. Отметим, что иногда линейной комбинацией называется сама формальная сумма $v_1a_1 + \cdots + v_na_n$, а иногда — ее значение (то есть, элемент V). Множество всех линейных комбинаций векторов $v_1, \ldots, v_m$ называется их **линейной оболочкой** и обозначается через $\langle v_1, \ldots, v_m \rangle$. Полезно определить линейную оболочку и для бесконечного множества векторов: пусть $S \subseteq V$ — произвольное подмножество векторного пространства V. Его линейной оболочкой называется множество всех линейных комбинаций вида $v_1a_1 + \cdots + v_na_n$, где $v_1, \ldots, v_n \in S$. Обозначение: $\langle S \rangle$.

Замечание 6.3.2. Нетрудно проверить, что линейная оболочка произвольного подмножества в V является векторным подпространством в V. Заметим также, что линейная оболочка пустого подмножества $\varnothing \subset V$ равна тривиальному подпространству $\{0\}$.

Определение 6.3.3. Пусть V — векторное пространство, $v_1, \ldots, v_m \in V$. Будем говорить, что $v_1, \ldots, v_m$ — **система образующих** пространства V (или что векторы $v_1, \ldots, v_m$ **порождают** пространство V, или что пространство V **порождается** векторами $v_1, \ldots, v_m$), если их линейная оболочка совпадает с V: $\langle v_1, \ldots, v_m \rangle = V$. Пространство называется **конечномерным**, если оно порождается некоторым конечным набором векторов. Можно определить систему образующих и в случае бесконечного набора векторов: подмножество $S \subseteq V$ называется **системой образующих** пространства V, если его линейная оболочка совпадает с V.

Примеры 6.3.4.
1. Пространство столбцов k^n конечномерно. Действительно, обозначим через $e_i \in k^n$ столбец, у которого в i-ой позиции стоит 1, а в остальных — 0. Нетрудно проверить, что векторы $e_1, \ldots, e_n$ порождают k^n.
2. Пространство многочленов $k[x]$ над полем k не является конечномерным. Действительно, предположим, что оно порождается некоторым конечным набором многочленов. Пусть m — наибольшая из степеней этих многочленов. Тогда все линейные комбинации элементов нашего набора являются многочленами степени не выше m, и поэтому их множество не совпадает со всем пространством $k[x]$.

Определение 6.3.5. Пространство, не являющееся конечномерным, называется **бесконечномерным**. По определению это означает, что *никакой* конечный набор элементов этого пространства не порождает его.

Пусть $v_1, \ldots, v_n \in V$, и пусть $v \in \langle v_1, \ldots, v_n \rangle$. По определению это означает, что существуют коэффициенты $a_1, \ldots, a_n \in k$ такие, что $v = v_1a_1 + \cdots + v_na_n$. Зададимся вопросом: единственен ли такой набор коэффициентов? Пусть $b_1, \ldots, b_n \in k$ — еще один набор скаляров, для которого $v = v_1b_1 + \cdots + v_nb_n$. Вычитая одно равенство из другого, получаем $0 = v_1(b_1 - a_1) + \cdots + v_n(b_n - a_n)$. Мы записали 0 как линейную комбинацию векторов $v_1, \ldots, v_m$. Если единственный способ сделать это тривиален (положить все коэффициенты равными 0), то $b_i = a_i$ для всех i, и поэтому наш набор коэффициентов $a_1, \ldots, a_n$ единственен.

Определение 6.3.6. Набор векторов $v_1, \dots, v_n \in V$ называется **линейно независимым**, если из равенства $v_1 a_1 + \dots + v_n a_n = 0$ следует, что $a_1 = \dots = a_n$. Назовем выражение вида $v_1 a_1 + \dots + v_n a_n$ **тривиальной линейной комбинацией**, если все ее коэффициенты равны нулю: $a_1 = \dots = a_n$. Тогда векторы $v_1, \dots, v_n \in V$ линейно независимым если и только если никакая их нетривиальная линейная комбинация не равна нулю. В таком виде определение удобно обобщить на произвольное (не обязательно конечное) множество векторов: подмножество $S \subseteq V$ назовем **линейно независимым**, если из того, что некоторая линейная комбинация векторов S равна нулю, следует, что все ее коэффициенты равны нулю.

Определение 6.3.7. Набор векторов $S \subseteq V$, который *не является* линейно независимым, называется **линейно зависимым**. По определению это означает, что *существует* некоторая нетривиальная линейная комбинация векторов из S, которая равна нулю. Таким образом, набор $v_1, \dots, v_n \in V$ **линейно зависим**, если существуют коэффициенты $a_1, \dots, a_n \in k$, не все из которых равны нулю, такие, что $v_1 a_1 + \dots + v_n a_n = 0$

Замечание 6.3.8. Еще одна полезная переформулировка: набор векторов линейно зависим тогда и только тогда, когда некоторый вектор из него выражается через остальные (то есть, лежит в линейной оболочке остальных). Действительно, если набор S линейно зависим, то существует нетривиальная линейная зависимость вида $v_1 a_1 + \dots + v_n a_n = 0$. Нетривиальность означает, что некоторый ее коэффициент отличен от нуля; без ограничения общности можно считать, что $a_1 \neq 0$. Но тогда $v_1 = -\frac{a_2}{a_1} v_2 - \dots - \frac{a_n}{a_1} v_n$. Обратное следствие очевидно. Упражнение: проверьте, что наша переформулировка работает и для «вырожденных» случаев наборов из одного вектора.

Замечание 6.3.9. Рассуждение перед определением 6.3.6 показывает, что набор $v_1, \dots, v_n$ линейно независим тогда и только тогда, когда у каждого вектора из линейной оболочки $\langle v_1, \dots, v_n \rangle$ есть только одно представление в виде линейной комбинации векторов $v_1, \dots, v_n$. Аналогично, линейная независимость произвольного подмножества $S \subseteq V$ означает, что у каждого вектора из линейной оболочки $\langle S \rangle$ есть только одно представление в виде линейной комбинации векторов из S.

Примеры 6.3.10.

1. Набор из трех векторов $\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \in k^4$ линейно независим. Действительно, их линейная комбинация с коэффициентами a_1, a_2, a_3 равна $\begin{pmatrix} a_1 \\ 0 \\ a_2 \\ a_3 \end{pmatrix}$, и из равенства нулю этого вектора следует, что $a_1 = a_2 = a_3$.
2. Пусть n — произвольное натуральное число. Тогда набор $1, x, x^2, \dots, x^n$ линейно независим в пространстве многочленов $k[x]$ (упражнение!). Более того, бесконечное множество $\{1, x, x^2, \dots, x^n, \dots\}$ линейно независимо в $k[x]$.

3. Любое множество векторов, содержащее нулевой вектор, линейно зависимо.

4. Набор из одного вектора $v \in V$ линейно независим тогда и только тогда, когда $v \neq 0$.

5. Набор из двух векторов $u, v \in V$ линейно независим тогда и только тогда, когда ни один из них не получается из другого умножением на скаляр (почему?).

Лемма 6.3.11. *Пусть V — векторное пространство, $X \subseteq Y \subseteq V$. Если Y линейно независимо, то и X линейно независимо. Если X линейно зависимо, то и Y линейно зависимо.*

Доказательство. Очевидно. □

Следующая лемма окажется чрезвычайно полезной. Она утверждает, что если имеется линейно зависимый набор векторов, в котором первый вектор отличен от нуля, то один из векторов набора выражается через предыдущие; тогда его можно выбросить, не изменив линейную оболочку набора.

Лемма 6.3.12 (о линейной зависимости). *Пусть набор $(v_1, \dots, v_n)$ векторов пространства V линейно зависим, и $v_1 \neq 0$. Тогда существует индекс $j \in \{2, \dots, n\}$ такой, что*

- $v_j \in \langle v_1, \dots, v_{j-1} \rangle$;
- $\langle v_1, \dots, v_n \rangle = \langle v_1, \dots, \widehat{v_j}, \dots, v_n \rangle$.

Доказательство. По условию найдутся $a_1, \dots, a_n \in k$ такие, что $v_1 a_1 + \dots + v_n a_n = 0$. Пусть j — наибольший индекс, для которого $a_j \neq 0$. Тогда

$$v_j = -\frac{a_1}{a_j} v_1 - \dots - \frac{a_{j-1}}{a_j} v_{j-1},$$

и первый пункт доказан. Очевидно, что $\langle v_1, \dots, \widehat{v_j}, \dots, v_n \rangle \subseteq \langle v_1, \dots, v_n \rangle$. Покажем обратное включение. Пусть $u \in \langle v_1, \dots, v_n \rangle$. Это означает, что $u = v_1 c_1 + \dots + v_n c_n$ для некоторых $c_1, \dots, c_n \in k$. Заменим в правой части вектор v_j на его выражение через $v_1, \dots, v_{j-1}$; получим, что u есть линейная комбинация векторов $v_1, \dots, \widehat{v_j}, \dots, v_n$, что и требовалось. □

Следствие 6.3.13. *Пусть набор векторов $v_1, \dots, v_n$ линейно независим, и $v \in V$. Набор $v_1, \dots, v_n, v$ линейно зависим тогда и только тогда, когда v лежит в $\langle v_1, \dots, v_n \rangle$.*

Доказательство. Если набор $v_1, \dots, v_n, v$ линейно зависим, то (по лемме 6.3.12) некоторый вектор в нем выражается через предыдущие. Это не может быть один из $v_1, \dots, v_n$ в силу линейной независимости $v_1, \dots, v_n$ □

Следующая теорема играет ключевую роль в изучении линейно независимых и порождающих систем.

Теорема 6.3.14. *В конечномерном векторном пространстве количество элементов в любом линейно независимом множестве не превосходит количества элементов в любом порождающем множестве. Иными словами, если $u_1, \dots, u_m$ линейно независимые векторы пространства V, и $\langle v_1, \dots, v_n \rangle = V$, то $m \leqslant n$.*

Доказательство. Опишем процесс, на каждом шаге которого мы заменяем один вектор из $\{v_i\}$ на один вектор из $\{u_j\}$. Заметим сначала, что при добавлении к $v_1, \ldots, v_n$ любого вектора мы получим линейно зависимую систему. В частности, набор $u_1, v_1, \ldots, v_n$ линейно зависим. По лемме 6.3.12 мы можем выкинуть из этого набора один из векторов $v_1, \ldots, v_n$ (скажем, v_j) так, что оставшиеся векторы все еще будут порождать V. Мы получили набор вида $u_1, v_1, \ldots, \widehat{v_j}, \ldots, v_n$, порождающий V. Снова заметим, что при добавлении к нему любого вектора мы получим линейно зависимую систему. В частности, система $u_1, u_2, v_1, \ldots, \widehat{v_j}, \ldots, v_n$ линейно зависима. По лемме 6.3.12 какой-то вектор в ней выражается через предыдущие. Понятно, что это не u_2: это бы означало, что u_1, u_2 линейно зависимы. Значит, это один из v_i. Лемма 6.3.12 утверждает, что его можно выбросить, и оставшиеся векторы все еще будут порождать V.

Теперь ясно, что мы можем продолжать этот процесс: на i-ом шаге у нас есть порождающий набор $u_1, \ldots, u_{i-1}, v_{j_1}, \ldots$ длины n. Добавим к нему вектор u_i, поместив его после u_{i-1}, и получим линейно зависимый набор $u_1, \ldots, u_i, v_{j_1}, \ldots$. По лемме 6.3.12 некоторый вектор из этого набора выражается через предыдущие. Это не может быть один из векторов $u_1, \ldots, u_i$ в силу линейной независимости набора $u_1, \ldots, u_m$. Поэтому это один из v_i; его можно выбросить и линейная оболочка набора не изменится.

Заметим теперь, что на каждом шаге мы заменяем один вектор из v_i на один вектор из u_j. Если же $m > n$, это означает, что после n-го шага мы получили порождающий набор вида $u_1, \ldots, u_n$. Добавляя вектор u_{n+1} мы должны получить линейно зависимый набор, который в то же время является подмножеством линейно независимого набора $u_1, \ldots, u_m$, чего не может быть. □

Предложение 6.3.15. *Любое подпространство конечномерного векторного пространства конечномерно.*

Доказательство. Пусть V — конечномерное пространство, $U \leqslant V$. Построим цепочку векторов $v_1, v_2, \ldots$ следующим образом. Заметим для начала, что если $U = \{0\}$, то U конечномерно и доказывать нечего. Если же $U \neq \{0\}$, выберем ненулевой вектор $v_1 \in U$. Очевидно, что $\langle v_1 \rangle \subseteq U$. Если на самом деле $\langle v_1 \rangle = U$, то доказательство окончено. Иначе можно выбрать $v_2 \in U$ так, что $v_2 \notin \langle v_1 \rangle$. Теперь мы получили набор v_1, v_2, и $\langle v_1, v_2 \rangle \subseteq U$. Продолжим процесс: на i-ом шаге у нас есть набор $v_1, \ldots, v_{i-1}$ такой, что $\langle v_1, \ldots, v_{i-1} \rangle \subseteq U$. Если на самом деле имеет место равенство, то U конечномерно, что и требовалось. Если нет — выберем $v_i \in U$ так, что $v_i \notin \langle v_1, \ldots, v_{i-1}$. Заметим, что на каждом шаге мы получаем линейно независимый набор. Действительно, если векторы $v_1, \ldots, v_i$ линейно зависимы, то по лемме 6.3.12 какой-то из них выражается через предыдущие, что невозможно в силу выбора каждого вектора. Но по теореме 6.3.14 длина этого линейно независимого набора векторов пространства V не превосходит количества элементов в некотором (конечном) порождающем множестве (которое существует по предположению теоремы). Поэтому описанный процесс не может продолжаться бесконечно. □

6.4 Базис

Литература: [F], гл. XII, § 1, п. 2; [K2], гл. 1, § 2, п. 1–2; [KM], ч. 1, § 2; [vdW], гл. 4, § 20.

Определение 6.4.1. Пусть V — векторное пространство над полем k. Набор векторов называется **базисом** пространства V, если он одновременно линейно независим и порождает V.

Неформально говоря, линейно независимые наборы векторов очень «маленькие», а системы образующих — «большие». На стыке этих двух плохо совместимых свойств возникает понятие базиса. Сейчас мы сформулируем и докажем несколько эквивалентных переформулировок понятия базиса.

Теорема 6.4.2. *Подмножество $\mathcal{B} \subseteq V$ является базисом тогда и только тогда, когда любой вектор V представляется в виде линейной комбинации элементов из $\mathcal{B}$, причем единственным образом.*

Доказательство. Если $\mathcal{B}$ — базис, то по определению системы образующих любой вектор из V представляется в виде линейной комбинации элементов из $\mathcal{B}$. Если таких представления у вектора $v \in V$ два, например, $u_1a_1 + \cdots + u_na_n = v = u_1b_1 + \cdots + u_nb_n$ для некоторых $u_i \in \mathcal{B}$, $a_i, b_i \in k$, то $u_1(a_1 - b_1) + \cdots + u_n(a_n - b_n) = 0$, и из линейной независимости $\mathcal{B}$ следует, что все коэффициенты в этой линейной комбинации равны 0, откуда $a_i = b_i$ для всех i, и на самом деле два представления вектора v совпадают.

Обратно, если любой вектор V представляется в виде линейной комбинации элементов из $\mathcal{B}$ единственным образом, то $\mathcal{B}$ является системой образующих, и если она линейно зависима, то имеется нетривиальная линейная комбинация $v_1a_1 + \cdots + v_na_n = 0 = v_1 \cdot 0 + \cdots + v_n \cdot 0$. Мы получили два различных представления одного вектора $0 \in V$ (они различны, поскольку не все a_i равны нулю) — противоречие. □

Теорема 6.4.3. *Из любой конечной системы образующих пространства V можно выбрать базис.*

Доказательство. Пусть $v_1, \ldots, v_n$ — система образующих пространства V. Сейчас мы выбросим из нее некоторые векторы так, чтобы она стала базисом V. А именно, последовательно для $j = 1, 2, \ldots, n$, мы выбросим v_j, если $v_j \in \langle v_1, \ldots, v_{j-1}\rangle$. Заметим, что при каждом выбрасывании линейная оболочка векторов не меняется, поскольку мы выбрасываем только такие векторы, которые выражаются через предыдущие. Покажем, что полученный в итоге набор векторов линейно независим. Если он линейно зависим, то по лемме 6.3.12 там найдется вектор, лежащий в линейной оболочке предыдущих; но такой вектор был бы выкинут в процессе. Заметим, что лемму 6.3.12 можно применить, поскольку первый вектор в нашем наборе обязан быть ненулевым: линейная оболочка пустого набора равна $\{0\}$. □

Следствие 6.4.4. *В любом конечномерном пространстве есть базис.*

Доказательство. По определению, в конечномерном пространстве есть конечная система образующих. По теореме 6.4.3 из нее можно выбрать базис. □

Замечание 6.4.5. На самом деле, базис есть в любом пространстве, даже бесконечномерном. Доказательство этого факта, однако, требует тонкого рассуждения с использованием *аксиомы выбора* (см. замечание 1.3.7 в недрах доказательства теоремы 1.3.6), поэтому мы воздержимся от него. В нашем курсе речь будет вестись только о конечномерных пространствах; формулировки для бесконечномерных пространств мы приводим только тогда, когда они в точности повторяют формулировки в конечномерном случае.

Следующая теорема в некотором смысле двойственна теореме 6.4.3.

Теорема 6.4.6. *Любой линейно независимый набор векторов в конечномерном пространстве можно дополнить до базиса.*

Доказательство. Пусть $u_1, \dots, u_m$ — линейно независимая система векторов пространства V, и пусть $v_1, \dots, v_n$ — произвольная порождающая система пространства V (она существует по определению конечномерности). Положим для начала $\mathcal{B} = \{u_1, \dots, u_m\}$ и проделаем следующую процедуру последовательно для $j = 1, \dots, n$: если вектор v_j не лежит в линейной оболочке $\langle \mathcal{B} \rangle$ множества $\mathcal{B}$, то добавим его к $\mathcal{B}$; а если лежит — пропустим. Заметим, что после каждого такого шага множество $\mathcal{B}$ все еще линейно независимо (следствие 6.3.13). После n-го шага мы получим, что *каждый* из векторов $v_1, \dots, v_n$ лежит в $\langle \mathcal{B} \rangle$. Но тогда и любой вектор, выражающийся через $v_1, \dots, v_n$, лежит в $\langle \mathcal{B} \rangle$. Поэтому $\langle \mathcal{B} \rangle = V$. □

В качестве применения теоремы 6.4.6 приведем следующий полезный результат.

Предложение 6.4.7. *Пусть V — конечномерное пространство, $U \leqslant V$. Тогда существует подпространство $W \leqslant V$ такое, что $U \oplus W = V$.*

Доказательство. По предложению 6.3.15 пространство U конечномерно. По следствию 6.4.4 в нем есть базис, скажем, $u_1, \dots, u_m$. Система векторов $u_1, \dots, u_m$ в пространстве V линейно независима; по теореме 6.4.6 ее можно дополнить до базиса. Этот базис имеет вид $u_1, \dots, u_m, w_1, \dots, w_n$ для некоторых векторов $w_1, \dots, w_n \in V$. Пусть $W = \langle w_1, \dots, w_n \rangle$. Покажем, что $U \oplus W = V$. По предложению 6.2.10 для этого достаточно проверить, что $U + W = V$ и $U \cap W = \{0\}$.

Покажем сначала, что $U + W = V$. Пусть $v \in V$; поскольку $u_1, \dots, u_m, w_1, \dots, w_n$ — базис V, можно записать $v = u_1 a_1 + \dots + u_m a_m + w_1 b_1 + \dots + w_n b_n$ для некоторых скаляров $a_i, b_j \in k$. Обозначим $u = u_1 a_1 + \dots + u_m a_m$, $w = w_1 b_1 + \dots + w_n b_n$; тогда $v = u + w$, причем $u \in U$, $w \in W$.

Пусть теперь $v \in U \cap W$. Тогда существуют скаляры $a_i, b_j \in k$ такие, что $v = u_1 a_1 + \dots + u_m a_m = w_1 b_1 + \dots + w_n b_n$. Но тогда $u_1 a_1 + \dots + u_m a_m - w_1 b_1 - \dots - w_n b_n = 0$ — линейная комбинация, равная нулю. Из линейной независимости нашего набора следует, что все ее коэффициенты равны нулю, а потому и $v = 0$. □

6.5 Размерность

Литература: [F], гл. XII, § 1, п. 2; [K2], гл. 1, § 2, п. 1–2; [KM], ч. 1, § 2; [vdW], гл. 4, § 19.

Мы говорили о *конечномерных* пространствах, не зная, что такое *размерность*. Как же определить размерность векторного пространства? Интуитивно понятно, что размерность пространства столбцов k^n должна равняться n. Заметим, что столбцы

$$\begin{pmatrix}1\\0\\\vdots\\0\end{pmatrix}, \begin{pmatrix}0\\1\\\vdots\\0\end{pmatrix}, \ldots, \begin{pmatrix}0\\0\\\vdots\\1\end{pmatrix}$$

образуют базис в k^n. Поэтому хочется определить размерность пространства V как количество элементов в базисе V. Но возникает проблема: в *каком* базисе? Конечномерное пространство V может иметь много различных базисов, и могло бы оказаться, что у него есть базисы разной длины. Следующая теорема утверждает, что этого не происходит.

Теорема 6.5.1. *Пусть V — конечномерное векторное пространство. В любых двух базисах V поровну элементов.*

Доказательство. Пусть $\mathcal{B}_1$, $\mathcal{B}_2$ — два [конечных] базиса V. Тогда $\mathcal{B}_1$ — линейно независимая система, а $\mathcal{B}_2$ — порождающая система; по теореме 6.3.14 количество элементов в $\mathcal{B}_1$ не больше, чем в $\mathcal{B}_2$. С другой стороны, $\mathcal{B}_2$ — линейно независимая система, а $\mathcal{B}_1$ — порождающая, поэтому количество элементов в $\mathcal{B}_2$ не больше, чем в $\mathcal{B}_1$. Поэтому в них поровну элементов. □

Определение 6.5.2. Пусть V — конечномерное векторное пространство над полем k. Количество элементов в любом его базисе называется **размерностью** пространства V и обозначается через $\dim_k V$ или просто через $\dim V$. Если же в V нет конечной системы образующих, то любой базис V содержит бесконечное число элементов; в этом случае мы пишем $\dim_k V = \infty$ и говорим, что пространство V **бесконечномерно**.

Предложение 6.5.3. *Пусть V — конечномерное векторное пространство над k и $U < V$. Тогда $\dim_k U \leqslant \dim_k V$. Более того, $\dim_k U = \dim_k V$ тогда и только тогда, когда $U = V$.*

Доказательство. Пусть $n = \dim_k V$ и $\mathcal{B}$ — некоторый базис U. Заметим, что $\mathcal{B}$ — линейно независимая система векторов в пространстве V. По теореме 6.4.6 ее можно дополнить до базиса V. Значит, $|\mathcal{B}| = \dim_k U$ не превосходит размерности V.

Если при этом $\dim_k U = \dim_k V$, то это дополнение должно быть того же размера, что и само множество $\mathcal{B}$. Это означает, что $\mathcal{B}$ является базисом всего пространства V, значит, $U = \langle\mathcal{B}\rangle = V$. Обратное очевидно: если $U = V$, то $\dim_k U = \dim_k V$. □

Представим, что перед нами [конечный] набор векторов пространства V. Как показать, что он образует базис? Можно действовать по определению и проверить два факта:

- этот набор линейно независим;
- этот набор порождает V.

Оказывается, из теорем 6.4.3 и 6.4.6 (вместе с теоремой 6.5.1) следует, что проверку любого одного из этих пунктов можно опустить, если мы уже знаем, что в нашем наборе нужное количество элементов: столько, какова размерность пространства V. Разумеется, для этого мы должны заранее знать эту размерность.

Предложение 6.5.4. *Пусть V — конечномерное векторное пространство. Любая система образующих V длины $\dim(V)$ является базисом V. Любая линейно независимая система длины $\dim(V)$ является базисом V.*

Доказательство. По теореме 6.4.3 из системы образующих можно выбрать базис. Поскольку этот базис должен иметь длину $\dim(V)$, как и исходная система, то она сама является базисом. Аналогично, по теореме 6.4.6 любую линейно независимую систему можно дополнить до базиса. Поскольку в ней уже столько же элементов, сколько в любом базисе, это дополнение должно быть пустым. Значит, она сама является базисом. □

Следующая теорема выражает размерность суммы подпространств через размерности самих подпространств и их пересечения.

Теорема 6.5.5 (Грассмана). *Пусть $U_1, U_2 \leqslant V$. Тогда*

$$\dim(U_1 + U_2) = \dim(U_1) + \dim(U_2) - \dim(U_1 \cap U_2).$$

Доказательство. Пусть $\{u_1, \dots, u_m\}$ — произвольный базис пространства $U_1 \cap U_2$ (и, таким образом, $m = \dim(U_1 \cap U_2)$. Система $\{u_1, \dots, u_m\}$ линейно независима как набор векторов в U_1, и поэтому ее можно дополнить до базиса: пусть $\{u_1, \dots, u_m, v_1 \,\mathrm{dots}, v_l\}$ — базис U_1. Аналогично, система $\{u_1, \dots, u_m\}$ линейно независима как набор векторов в U_2, и поэтому ее можно дополнить до базиса пространства U_2: пусть $\{u_1, \dots, u_m, w_1, \dots, w_n\}$ — этот базис.

Покажем, что набор $\mathcal{B} = \{u_1, \dots, u_m, v_1, \dots, v_l, w_1, \dots, w_n\}$ является базисом пространства $U_1 + U_2$. Это система образующих: действительно, любой вектор в $U_1 + U_2$ по определению есть сумма вектора из U_1 и вектора из U_2, и каждый из этих двух векторов есть линейная комбинация векторов из $\mathcal{B}$. Поэтому $\langle\mathcal{B}\rangle$ содержит $U_1 + U_2$; с другой стороны, все векторы из $\mathcal{B}$ лежат в $U_1 + U_2$, поэтому на самом деле $\langle\mathcal{B}\rangle = U_1 + U_2$.

Осталось проверить, что множество $\mathcal{B}$ линейно независимо. Предположим, что $u_1a_1 + \dots + u_ma_m + v_1b_1 + \dots + v_lb_l + w_1c_1 + \dots + w_nc_n = 0$. Перепишем это равенство:

$$w_1c_1 + \dots + w_nc_n = -u_1a_1 - \dots - u_ma_m - v_1b_1 - \dots - v_lb_l.$$

Заметим, что левая часть лежит в U_2, а правая лежит в U_1. Поэтому $w_1c_1 + \dots + w_nc_n \in U_1 \cap U_2$. Мы знаем базис в $U_1 \cap U_2$ — это $\{u_1, \dots, u_m\}$. Поэтому

$$w_1c_1 + \dots + w_nc_n = u_1d_1 + \dots + u_md_m.$$

Но набор векторов $\{u_1, \dots, u_m, w_1, \dots, w_n\}$ линейно независим; поэтому из последнего равенства следует, что все коэффициенты в нем равны 0. В частности, $c_1 = \dots = c_n = 0$. Поэтому наша исходная линейная зависимость имеет вид

$$u_1a_1 + \dots + u_ma_m + v_1b_1 + \dots + v_lb_l = 0.$$

Но набор $\{u_1, \ldots, u_m, v_1, \ldots, v_l\}$ также линейно независим, и потому $a_1 = \cdots = a_m = v_1 = \cdots = v_l = 0$; значит, исходная линейная комбинация тривиальна, что и требовалось. □

Следствие 6.5.6. *Если* $V = U_1 \oplus U_2$, *то* $\dim(V) = \dim(U_1) + \dim(U_2)$.

Доказательство. Очевидно. □

Предложение 6.5.7. *Пусть пространство* V *конечномерно, и* $U_1, \ldots, U_m$ *— его подпространства такие, что* $V = U_1 + \cdots + U_m$ *и* $\dim(V) = \dim(U_1) + \cdots + \dim(U_m)$. *Тогда* $V = U_1 \oplus \cdots \oplus U_m$.

Доказательство. Выберем базис в каждом подпространстве U_i. Объединение этих базисов является порождающей системой векторов в V (поскольку V является суммой U_i), а их количество равно размерности V. По предложению 6.5.4 он является базисом в V. Обозначим этот базис через $\mathcal{B}$. По определению прямой суммы нам нужно доказать, что если $0 = u_1 + \cdots + u_m$ для некоторых $u_i \in U_i$, то $u_1 = \cdots = u_m = 0$. Разложим каждый вектор u_i по выбранному базису пространства U_i — мы получим некоторую линейную комбинацию элементов базиса $\mathcal{B}$. Из ее равенства нулю следует, что все ее коэффициенты равны нулю, а потому и все u_i равны нулю, что и требовалось. □

7 Линейные отображения

7.1 Первые определения

Литература: [F], гл. XII, § 4, п. 1.; [K2], гл. 2, § 1, п. 1; [KM], ч. 1, § 3, пп. 1, 2; [vdW], гл. IV, § 23.

Определение 7.1.1. Пусть V, W — векторные пространства над полем k. Отображение $T\colon V \to W$ называется **линейным**, если

- $T(u + v) = T(u) + T(v)$;
- $T(va) = T(v)a$ для всех $a \in k$, $v \in V$.

Иногда вместо $T(v)$ мы будем писать Tv. Множество всех линейных отображений из V в W мы будем обозначать через $\mathrm{Hom}(V, W)$. Линейное отображение часто называется **гомоморфизмом** векторных пространств; оно называется **эндоморфизмом**, если $U = V$.

Пример 7.1.2. Обозначим через 0 отображение, которое любой вектор $v \in V$ переводит в $0 \in W$; то есть, $0(v) = 0$ для всех $v \in V$. Нетрудно видеть, что оно линейно, то есть, $0 \in \mathrm{Hom}(V, W)$. Обратите внимание, что мы используем тот же символ 0, что и для обозначения нулевого элемента поля k и нулевых элементов в векторных пространствах V и W.

Пример 7.1.3. Для каждого векторного пространства V можно рассмотреть тождественное отображение $\mathrm{id}_V\colon V \to V$. Нетрудно проверить, что он линейно; таким образом, $\mathrm{id}_V \in \mathrm{Hom}(V, W)$.

Пример 7.1.4. Для пространства многочленов $k[x]$ можно рассмотреть отображение *дифференцирования* $T\colon k[x] \to k[x]$, сопоставляющее каждому многочлену $f \in k[x]$ его производную f'. Это отображение линейно, поскольку $(f + g)' = f' + g'$ и $(fa)' = f'a$ для всех $f, g \in k[x]$ и $a \in k$ (см. предложение 4.5.6).

Пример 7.1.5. Отображение $k[x] \to k[x]$, умножающее каждый многочлен на x, является линейным.

Пример 7.1.6. Снова рассмотрим пространство многочленов $k[x]$, и пусть $c \in k$ — фиксированный элемент основного поля. Рассмотрим отображение $\mathrm{ev}_c\colon k[x] \to k$, сопоставляющее каждому многочлену $f \in k[x]$ его значение в точке c. Иными словами, $\mathrm{ev}_c(f) = f(c)$. Это отображение линейно (см. предложение 4.3.3); оно называется **эвалюацией в точке** c.

Пример 7.1.7. Пусть $k = \mathbb{R}$; рассмотрим отображение $T\colon \mathbb{R}[x] \to \mathbb{R}$, сопоставляющее многочлену $f \in \mathbb{R}[x]$ значение интеграла

$$T(f) = \int_0^1 f(x)\,dx.$$

Из простейших свойств определенного интеграла следует, что отображение T линейно.

Пример 7.1.8. Рассмотрим пространство бесконечных последовательностей ${}^\infty k$. Отображение $T\colon {}^\infty k \to {}^\infty k$, сопоставляющее последовательности $(x_1, x_2, \dots)$ последовательность $(x_2, x_3, \dots)$ (*сдвиг влево*) является линейным.

Пусть $T\colon V \to W$ — линейное отображение, и пусть $v_1, \ldots, v_n$ — базис пространства V. Если $v \in V$, то можно записать $v = v_1 a_1 + \cdots + v_n a_n$ для некоторых $a_1, \ldots, a_n \in k$. Тогда из определения линейности следует, что $T(v) = T(v_1)a_1 + \cdots + T(v_n)a_n$. Это означает, что значение T на любом векторе v полностью определяется своими значениями на базисе. Обратно, можно задать значения $T(v_1), \ldots, T(v_n) \in W$ *произвольным* образом, и по этим данным однозначно восстанавливается единственное линейное отображение из V в W.

Теорема 7.1.9 (Универсальное свойство базиса). *Пусть V, W — конечномерные векторные пространства, $v_1, \ldots, v_n$ — базис V, и пусть заданы произвольные векторы $w_1, \ldots, w_n \in W$. Существует единственное линейное отображение $T\colon V \to W$ такое, что $T(v_i) = w_i$ для всех $i = 1, \ldots, n$.*

Доказательство. Возьмем вектор $v \in V$ и разложим его базису $v_1, \ldots, v_n$: $v = v_1 a_1 + \cdots + v_n a_n$. Если $T(v_i) = w_i$ для $i = 1, \ldots, n$, то

$$\begin{aligned} T(v) &= T(v_1 a_1 + \cdots + v_n a_n) \\ &= T(v_1)a_1 + \cdots + T(v_n)a_n \\ &= w_1 a_1 + \cdots + w_n a_n. \end{aligned}$$

Таким образом, значение T на v однозначно определено (поскольку коэффициенты $a_1, \ldots, a_n$ однозначно определяются вектором v, см. теорему 6.4.2). Это рассуждение работает для произвольного вектора $v \in V$, поэтому линейное отображение T, удовлетворяющее условиям $T(v_i) = w_i$, единственно.

Обратно, если нам дан базис $\{v_i\}$ в V и векторы $\{w_i\}$, то для произвольного вектора $v = v_1 a_1 + \cdots + v_n a_n$ положим $T(v) = w_1 a_1 + \cdots + w_n a_n$ (это выражение определено однозначно по теореме 6.4.2). Мы получили отображение $T\colon V \to W$; осталось доказать, что оно линейно. Действительно, пусть $u, v \in V$, причем $v = v_1 a_1 + \cdots + v_n a_n$ и $u = v_1 b_1 + \cdots + v_n b_n$. Тогда по нашему определению $T(v) = w_1 a_1 + \cdots + w_n a_n$, $T(u) = w_1 b_1 + \cdots + w_n b_n$. Сложение выражений для u и v показывает, что $u + v = v_1(a_1 + b_1) + \cdots + v_n(a_n + b_n)$, и по определению T тогда $T(u + v) = w_1(a_1 + b_1) + \cdots + w_n(a_n + b_n)$. Нетрудно видеть теперь, что $T(u + v) = T(u) + T(v)$. Если, кроме того, $a \in k$, то $va = v_1 a_1 a + \cdots + v_n a_n a$, и потому $T(va) = w_1 a_1 a + \cdots + w_n a_n a$. Легко проверить, что $T(va) = T(v)a$. □

7.2 Операции над линейными отображениями

Литература: [F], гл. XII, § 4, пп. 4–6; [K2], гл. 2, § 1, п. 1; § 2, пп. 1–2; [KM], ч. 1, § 3; [vdW], гл. IV, § 23.

Пусть V, W — векторные пространства над k. Оказывается, множество $\mathrm{Hom}(V, W)$ всех линейных отображений из V в W естественным образом снабжается структурой векторного пространства над k. Чтобы продемонстрировать это, мы должны определить на нем две операции: сложение и умножение на скаляр. Пусть $S, T\colon V \to W$ — линейные отображения. Определим новое отображение $S + T\colon V \to W$ формулой $(S + T)(v) = S(v) + T(v)$ для всех $v \in V$. Нетрудно проверить, что отображение $S + T$ линейно. Поэтому для $S, T \in \mathrm{Hom}(V, W)$ мы

построили их сумму $S + T \in \mathrm{Hom}(V, W)$. Если же $S\colon V \to W$ — линейное отображение, и $a \in k$, можно определить отображение $Sa\colon V \to W$ формулой $(Sa)(v) = S(v)a$. Это отображение также линейно, то есть, $Sa \in \mathrm{Hom}(V, W)$.

Теперь можно проверить, что введенные операции действительно превращают $\mathrm{Hom}(V, W)$ в векторное пространство. Роль нулевого элемента в нем играет нулевое отображение $0\colon \mathrm{Hom}(V, W)$. Для примера проверим одно условие из определения векторного пространства: пусть $S, T \in \mathrm{Hom}(V, W)$, $a \in k$. Тогда для всех $v \in V$ выполнены равенства

$$\begin{aligned}((S + T)a)(v) &= ((S + T)(v)) \cdot a \\ &= (S(v) + T(v))a \\ &= (S(v)a) + (T(v)a) \\ &= (Sa)(v) + (Ta)(v) \\ &= (Sa + Ta)(v)\end{aligned}$$

Поэтому отображения $(S + T)a$, $Sa + Ta$ из V в W совпадают.

Более того, некоторые линейные отображения можно «перемножать». Пусть U, V, W — векторные пространства над k. Возьмем линейные отображения $T \in \mathrm{Hom}(U, V)$ и $S \in \mathrm{Hom}(V, W)$. Тогда имеет смысл рассматривать их композицию $S \circ T\colon U \to W$. Оказывается, отображение $S \circ T$ также является линейным. Действительно, напомним, что $(S \circ T)(u) = S(T(u))$ для всех $u \in U$ по определению композиции. Поэтому

$$\begin{aligned}(S \circ T)(u_1 + u_2) &= S(T(u_1 + u_2)) \\ &= S(T(u_1) + T(u_2)) \\ &= S(T(u_1)) + S(T(u_2)) \\ &= (S \circ T)(u_1) + (S \circ T)(u_2)\end{aligned}$$

для всех $u_1, u_2 \in U$. Если же $u \in U$, $a \in k$, то

$$(S \circ T)(ua) = S(T(ua)) = S(T(u)a) = S(T(u))a = (S \circ T)(u)a.$$

Значит, $S \circ T \in \mathrm{Hom}(U, W)$. Вместо $S \circ T$ мы будем часто писать ST и воспринимать ST как *произведение* линейных отображений S и T.

Заметим, что композиция линейных отображений автоматически ассоциативна (по теореме 1.3.1), то есть, $R(ST) = (RS)T$ для трех линейных отображений таких, что указанные композиции имеют смысл. Тождественные отображение линейны и играют роль нейтральных элементов: $T\,\mathrm{id}_V = \mathrm{id}_W T$ для $T \in \mathrm{Hom}(V, W)$. Наконец, несложно проверить (упражнение!), что умножение и сложение линейных отображений обладают свойством дистрибутивности: если $T, T_1, T_2 \in \mathrm{Hom}(U, V)$ и $S, S_1, S_2 \in \mathrm{Hom}(V, W)$ то $(S_1 + S_2)T = S_1T + S_2T$ и $S(T_1 + T_2) = ST_1 + ST_2$.

Конечно, произведение линейных отображений некоммутативно: равенство $ST = TS$ не обязано выполняться, даже если обе его части имеют смысл. Например, если $T \in \mathrm{Hom}(k[x], k[x])$ — отображение дифференцирования многочленов (см. пример 7.1.4), а $S \in \mathrm{Hom}(k[x], k[x])$ — умножение на x (см. пример 7.1.5), то $((ST)(f))(x) = xf'(x)$, а $((TS)(f))(x) = (xf(x))' = xf'(x) + f(x)$. Таким образом, $ST - TS = \mathrm{id}_{k[x]}$.

7.3 Ядро и образ

Литература: [F], гл. XII, § 4, п. 1; [K2], гл. 2, § 1, пп. 1, 3; [KM], ч. 1, § 3.

Определение 7.3.1. Пусть $T \in \mathrm{Hom}(V, W)$ — линейное отображение. Его **ядром** называется множество векторов, переходящих в 0 под действием T:

$$\mathrm{Ker}(T) = \{v \in V \mid T(v) = 0\}.$$

Пример 7.3.2. Если $T \in \mathrm{Hom}(k[x], k[x])$ — дифференцирование (см. пример 7.1.4), то $\mathrm{Ker}(T) = \{f \in k[x] \mid f' = 0\}$. Если поле k имеет характеристику 0, то $\mathrm{Ker}(T)$ состоит только из констант, то есть, $\mathrm{Ker}(T) = k \subseteq k[x]$ — одномерное подпространство в $k[x]$. Если же $\mathrm{char}\, k = p$, то существуют и неконстантные многочлены $f \in k[x]$ такие, что $f' = 0$. Например, таков многочлен x^p, а потому и любой многочлен от x^p: действительно, обозначим $g(x) = x^p$, тогда $(f(g(x)))' = f'(g(x)) \cdot g'(x) = 0$. Можно показать (упражнение!), что $\mathrm{Ker}(T)$ в этом случае в точности состоит из многочленов от x^p, то есть, от многочленов вида $\sum_{j=0}^{n} a_j x^{jp}$. Таким образом, $\mathrm{Ker}(T) = k[x^p]$ в этом случае бесконечномерно.

Пример 7.3.3. Пусть $T \in \mathrm{Hom}(k[x], k[x])$ — умножение на x (см. пример 7.1.5). Тогда $\mathrm{Ker}(T) = 0$.

Предложение 7.3.4. *Если* $T \in \mathrm{Hom}(V, W)$, *то* $\mathrm{Ker}(T)$ *является подпространством в* V.

Доказательство. Заметим, что $T(0) = T(0 + 0) = T(0) + T(0)$, откуда $T(0) = 0$. Значит, $0 \in \mathrm{Ker}(T)$. Если $u, v \in \mathrm{Ker}(T)$, то по определению $T(u) = T(v) = 0$. Тогда и $T(u + v) = T(u) + T(v) = 0 + 0 = 0$, то есть, $u + v \in \mathrm{Ker}(T)$. Наконец, если $u \in \mathrm{Ker}(T)$ и $a \in k$, то $T(u) = 0$ и $T(ua) = T(u)a = 0 \cdot a = 0$, откуда $ua \in \mathrm{Ker}(T)$. Вышесказанное означает, что $\mathrm{Ker}(T) \leqslant V$. □

Предложение 7.3.5. *Пусть* $T \in \mathrm{Hom}(V, W)$. *Отображение* T *инъективно тогда и только тогда, когда* $\mathrm{Ker}(T) = 0$.

Доказательство. Предположим, что T инъективно. Множество $\mathrm{Ker}(T)$ состоит из тех векторов v, для которых $T(v) = 0$. Мы знаем, что $T(0) = 0$ и из инъективности следует, что других таких векторов нет; поэтому $\mathrm{Ker}(T) = \{0\}$.

Обратно, предположим, что $\mathrm{Ker}(T) = 0$. Для проверки инъективности возьмем $v_1, v_2 \in V$ такие, что $T(v_1) = T(v_2)$ и покажем, что $v_1 = v_2$. Действительно, тогда $T(v_1 - v_2) = T(v_1) - T(v_2) = 0$, и потому $v_1 - v_2 \in \mathrm{Ker}(T) = \{0\}$, откуда $v_1 - v_2 = 0$, что и требовалось. □

Определение 7.3.6. Пусть $T \in \mathrm{Hom}(V, W)$. Его **образом** называется его образ как обычного отображения, то есть, множество

$$\mathrm{Im}(T) = \{T(v) \mid v \in V\}.$$

Предложение 7.3.7. *Если* $T \in \mathrm{Hom}(V, W)$, *то* $\mathrm{Im}(T)$ *является подпространством в* W.

Доказательство. Из равенства $T(0) = 0$ следует, что $0 \in \mathrm{Im}(T)$. Если $w_1, w_2 \in \mathrm{Im}(T)$, то найдутся $v_1, v_2 \in V$ такие, что $T(v_1) = w_1$ и $T(v_2) = w_2$. Но тогда $T(v_1 + v_2) = T(v_1) + T(v_2) = w_1 + w_2$, и потому $w_1 + w_2 \in \mathrm{Im}(T)$. Если $w \in \mathrm{Im}(T)$, то $T(v) = w$ для некоторого $v \in V$. Пусть $a \in k$; тогда $T(va) = T(v)a = wa$, и потому $wa \in \mathrm{Im}(T)$. По определению тогда $\mathrm{Im}(T) \leqslant W$. □

Теорема 7.3.8 (О гомоморфизме). *Пусть* V *— конечномерное пространство,* $T \in \mathrm{Hom}(V, W)$ *— линейное отображение. Тогда* $\mathrm{Im}(T)$ *является конечномерным подпространством в* W *и, кроме того,*

$$\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)).$$

Доказательство. Пусть $u_1, \ldots, u_m$ — базис $\mathrm{Ker}(T)$. Этот линейно независимый набор векторов можно продолжить до базиса $(u_1, \ldots, u_m, v_1, \ldots, v_n)$ всего пространства V по теореме 6.4.6. Таким образом, $\dim(\mathrm{Ker}(T)) = m$ и $\dim(V) = m + n$; нам остается лишь доказать, что $\dim(\mathrm{Im}(T)) = n$. Для этого рассмотрим векторы $T(v_1), \ldots, T(v_n)$ и покажем, что они образуют базис подпространства $\mathrm{Im}(T)$. Очевидно, что они лежат в $\mathrm{Im}(T)$, и потому $\langle T(v_1), \ldots, T(v_n)\rangle \subseteq \mathrm{Im}(T)$. Обратно, если $w \in \mathrm{Im}(T)$, то $w = T(v)$ для некоторого $v \in V$. Разложим v по нашем базису пространства V:

$$v = u_1 a_1 + \cdots + u_m a_m + v_1 b_1 + \cdots + v_n b_n$$

и применим к этому разложению отображение T:

$$w = T(v) = T(u_1 a_1 + \cdots + u_m a_m + v_1 b_1 + \cdots + v_n b_n) = T(v_1) b_1 + \cdots + T(v_n) b_n.$$

Поэтому $w \in \langle T(v_1), \ldots, T(v_n)$. Осталось показать, что векторы $T(v_1), \ldots, T(v_n)$ линейно независимы. Пусть $T(v_1)c_1 + T(v_n)c_n = 0$ — некоторая линейная комбинация. Тогда $0 = T(v_1 c_1 + \cdots + v_n c_n)$. Это означает, что вектор $v_1 c_1 + \cdots + v_n c_n$ лежит в $\mathrm{Ker}(T)$. Мы знаем базис $\mathrm{Ker}(T)$,потому $v_1 c_1 + \cdots + v_n c_n = u_1 d_1 + \cdots + u_m d_m$ для некоторых $d_i \in k$. Но набор векторов $u_1, \ldots, u_m, v_1, \ldots, v_n$ лниейно независим. Значит, все коэффициенты c_i, d_j равны нулю, и исходная линейная комбинация векторов $T(v_1), \ldots, T(v_n)$ тривиальна. □

Приведем пару полезных следствий этой теоремы; оказывается, уже тривиальные соображения неотрицательности размерности имеют серьезные последствия.

Следствие 7.3.9. *Пусть* V, W *— векторные пространства над* k*, и* $\dim V < \dim W$*. Не существует сюръективных линейных отображений* $V \to W$*.*

Доказательство. Предположим, что линейное отображение $T\colon V \to W$ сюръективно. Тогда $\mathrm{Im}(T) = W$, и по теореме 7.3.8 $\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)) = \dim(\mathrm{Ker}(T)) + \dim(W)$. Но $\dim(\mathrm{Ker}(T)) \geqslant 0$, и поэтому $\dim(V) \geqslant \dim(W)$ — противоречие с условием. □

Следствие 7.3.10. *Пусть* V, W *— векторные пространства над* k*, и* $\dim V > \dim W$*. Не существует инъективных линейных отображений* $V \to W$*.*

Доказательство. Предположим, что линейное отображение $T\colon V \to W$ инъективно. По предложению 7.3.5 ядро T тривиально. По теореме 7.3.8 $\dim(V) = \dim(\mathrm{Ker}(T)) + \dim(\mathrm{Im}(T)) = \dim(\mathrm{Im}(T)) \leqslant \dim(W)$ (последнее неравенство выполнено по предложению 6.5.3) — противоречие с условием. □

7.4 Матрица линейного отображения

Литература: [F], гл. XII, § 4, пп. 1–3; [K2], гл. 2, § 1, п. 2; § 2, п. 3; [KM], ч. 1, § 4; [vdW], гл. IV, § 23.

Пусть V, W — два конечномерных пространства, и пусть $\mathcal{B} = (v_1, \ldots, v_n)$ — упорядоченный базис V, а $\mathcal{B}' = (w_1, \ldots, w_m)$ — упорядоченный базис W. Универсальное свойства базиса (теорема 7.1.9) означает, что для задания линейного отображение $T\colon V \to W$ достаточно задать векторы $T(v_1), \ldots, T(v_n) \in W$. Каждый вектор $T(v_j)$, в свою очередь, можно разложить по базису $\mathcal{B}'$. Задание $T(v_j)$, таким образом, равносильно заданию коэффициентов в этом разложении. Мы получили, что линейное отображение $T\colon V \to W$ в итоге задается конечным набором скаляров — при условии, что в пространствах V и W выбраны базисы. Этот набор скаляров удобно записывать в виде матрицы.

Определение 7.4.1. Пусть $T\colon V \to W$ — линейное отображение между конечномерными пространствами, и пусть выбраны упорядоченные базисы $\mathcal{B} = (v_1, \ldots, v_n)$ в V и $\mathcal{B}' = (w_1, \ldots, w_m)$ в W. Разложим каждый вектор $T(v_j)$ по базису $\mathcal{B}'$ и запишем

$$T(v_j) = w_1 a_{1j} + w_2 a_{2j} + \cdots + w_m a_{mj}.$$

Набор коэффициентов $(a_{ij})_{\substack{1 \leqslant i \leqslant m \\ 1 \leqslant j \leqslant n}}$ мы воспринимаем как матрицу размера $m \times n$; она называется **матрицей линейного отображения T в базисах $\mathcal{B}$, $\mathcal{B}'$** и обозначается через $[T]_{\mathcal{B},\mathcal{B}'}$.

Как мы увидим ниже (см. теорему 7.5.6), линейное отображение полностью определяется своей матрицей (в выбранных базисах). Известные нам операции над линейными отображениями (сложение, умножение на скаляр, композиция) при этом превращаются в известные нам операции над матрицами (сложение, умножение на скаляр, произведение). Ниже мы введем понятие координат вектора, и тогда рассуждения с абстрактными векторными пространствами и линейными отображениями можно будет сводить к конкретным матричным вычислениям. Иными словами, матрицы полезны, когда вам нужно «засучить рукава» и вычислить что-нибудь конкретное. В то же время, всегда нужно помнить, что для перехода к матрицам нужно зафиксировать базисы в рассматриваемых пространствах, что может привести к утрате симметрии и некоторой неуклюжести.

Пусть $T, S\colon V \to W$ — линейные отображения, и в пространствах V, W выбраны базисы, как в определении 7.4.1. Покажем, что матрица суммы $T + S$ этих отображений является суммой матрицы отображения T и матрицы отображения S. Иными словами, $[T + S]_{\mathcal{B},\mathcal{B}'} = [T]_{\mathcal{B},\mathcal{B}'} + [S]_{\mathcal{B},\mathcal{B}'}$. Пусть $[T]_{\mathcal{B},\mathcal{B}'} = (a_{ij})$, $[S]_{\mathcal{B},\mathcal{B}'} = (b_{ij})$. По определению это означает, что $T(v_j) = \sum_{i=1}^{m} w_i a_{ij}$, $S(v_j) = \sum_{i=1}^{m} w_i b_{ij}$. Но тогда $(T + S)(v_j) = T(v_j) + S(v_j) = \sum_{i=1}^{m} w_i (a_{ij} + b_{ij})$. Значит, в разложении вектора $(T+S)(v_j)$ по базису $\mathcal{B}'$ коэффициент при w_i равен $a_{ij} + b_{ij}$. Это означает, что в матрице $[T + S]_{\mathcal{B},\mathcal{B}'}$ в позиции (i, j) стоит $a_{ij} + b_{ij}$. Но это и есть определение суммы матриц $[T]_{\mathcal{B},\mathcal{B}'}$ и $[S]_{\mathcal{B},\mathcal{B}'}$.

Совершенно аналогичное рассуждение показывает, что $[Ta]_{\mathcal{B},\mathcal{B}'} = [T]_{\mathcal{B},\mathcal{B}'} \cdot a$ для любого скаляра $a \in k$. Доказанные факты можно сформулировать следующим образом.

Теорема 7.4.2. *Пусть* V, W *— конечномерные векторные пространства над полем* k*, и* $\mathcal{B}, \mathcal{B}'$ *— базисы в* V, W *соответственно. Обозначим* $n = \dim(V)$*,* $m = \dim(W)$*. Отображение* $\varphi\colon \mathrm{Hom}(V, W) \to M(m, n, k)$*, сопоставляющее линейному отображению* $T \in \mathrm{Hom}(V, W)$ *его матрицу* $[T]_{\mathcal{B},\mathcal{B}'}$ *в базисах* $\mathcal{B}, \mathcal{B}'$*, является линейным.*

Доказательство. Для проверки линейности φ по определению нужно показать, что $[T + S]_{\mathcal{B},\mathcal{B}'} = [T]_{\mathcal{B},\mathcal{B}'} + [S]_{\mathcal{B},\mathcal{B}'}$ и $[Ta]_{\mathcal{B},\mathcal{B}'} = [T]_{\mathcal{B},\mathcal{B}'}a$ для всех $T, S \in \mathrm{Hom}(V, W)$, $a \in k$, что и было доказано выше. □

Гораздо интереснее посмотреть, что происходит при композиции линейных отображений.

Теорема 7.4.3. *Пусть* U, V, W *— три векторных пространства с базисами* $\mathcal{B} = (u_1, \ldots, u_l)$*,* $\mathcal{B}' = (v_1, \ldots, v_m)$*,* $\mathcal{B}'' = (w_1, \ldots, w_n)$*, соответственно, и пусть* $S\colon U \to V$*,* $T\colon V \to W$ *— линейные отображения. Тогда* $[T \circ S]_{\mathcal{B},\mathcal{B}''} = [T]_{\mathcal{B}',\mathcal{B}''} \cdot [S]_{\mathcal{B},\mathcal{B}'}$*.*

Читатель может проверить, что написанное выражение имеет смысл: в правой части стоят матрицы таких размеров, что их можно перемножить, и в результате получается матрица того же размера, что и в левой части.

Доказательство этого факта нужно воспринимать как (слегка запоздалое) объяснение определения умножения матриц. В самом деле, единственная причина, по которой умножение матриц выглядит так, как оно выглядит — это взаимно однозначное соответствие между матрицами и линейными отображениями, которое превращает композицию линейных отображений в умножение матриц. Каждый, кто задумается, что происходит при композиции линейных отображений (подстановке одних линейных выражений в другие), неизбежно обязан открыть умножение матриц.

Итак, пусть $[T]_{\mathcal{B}',\mathcal{B}''} = (a_{ij}) \in M(n, m, k)$, $[S]_{\mathcal{B},\mathcal{B}'} = (b_{ij}) \in M(m, l, k)$. Как найти матрицу отображения $T \circ S$? По определению мы должны разложить каждый вектор вида $(T \circ S)(u_k)$ по базису $w_1, \ldots, w_n$. Заметим, что $(T \circ S)(u_k) = T(S(u_k))$, а $S(u_k)$ мы умеем раскладывать по базису пространства V. А именно,

$$S(u_k) = \sum_{j=1}^{m} v_j b_{jk}.$$

Получаем, что

$$\begin{aligned}(T \circ S)(u_k) &= T\left(\sum_{j=1}^{m} v_j b_{jk}\right) \\ &= \sum_{j=1}^{m} T(v_j) b_{jk},\end{aligned}$$

где в последнем равенстве мы воспользовались линейностью отображения T. Теперь можно подставить в полученное выражение разложение для каждого вектора вида $T(v_j) = \sum_{i=1}^{n} w_i a_{ij}$.

После несложных преобразований сумм получаем

$$\begin{aligned}(T \circ S)(u_k) &= \sum_{j=1}^{m} T(v_j) b_{ji} \\ &= \sum_{j=1}^{m} \sum_{i=1}^{n} w_i a_{ij} b_{jk} \\ &= \sum_{i=1}^{n} w_i \left(\sum_{j=1}^{m} a_{ij} b_{jk} \right).\end{aligned}$$

Коэффициент при w_i в полученном разложении и равен коэффициенту, стоящему в позиции (i, k) матрицы $[T \circ S]_{\mathcal{B},\mathcal{B}''}$. Он оказался равен $\sum_{j=1}^{m} a_{ij} b_{jk}$, и потому матрица $[T \circ S]_{\mathcal{B},\mathcal{B}''}$ равна произведению матриц $[T]_{\mathcal{B}',\mathcal{B}''} \cdot [S]_{\mathcal{B},\mathcal{B}'}$.

Мы узнали, как понятие матрицы линейного отображение ведет себя при сложении отображений, умножении на скаляры, композиции. Есть еще одна операция над линейными отображениями, самая простая: мы можем в линейное отображение $T\colon V \to W$ подставить вектор из V и получить вектор из W. Отображению T мы сопоставили матрицу; сейчас мы сопоставим векторам из V и W некоторые столбцы (матрицы ширину 1) таким образом, что вычисление результата действия линейного отображения на векторе сведется к умножению матрицы на столбец.

А именно, пусть $\mathcal{B} = (v_1, \ldots, v_n)$ — базис векторного пространства V. Любой вектор $v \in V$ можно разложить по этому базису, то есть, записать его в виде линейной комбинации элементов $\mathcal{B}$:

$$v = v_1 a_1 + \cdots + v_n a_n, \quad a_i \in k.$$

Запишем полученные скаляры $a_1, \ldots, a_n$ в столбец. Полученный элемент пространства k^n называется **столбцом координат** (или **координатным столбцом**) **вектора** v **в базисе** $\mathcal{B}$ и обозначается так:

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}.$$

Коэффициенты $a_1, \ldots, a_n$ называются **координатами вектора** v **в базисе** $\mathcal{B}$. Обратите внимание на сходство этой записи с обозначением для матрицы линейного оператора в выбранных базисах.

Таким образом, как только мы выбрали базис $\mathcal{B}$ в пространстве V, каждому вектору из V сопоставляется столбец $[v]_{\mathcal{B}} \in k^n$. Более того, указанное сопоставление хорошо согласовано с операциями в пространстве V: если сложить два вектора, то соответствующие им координатные столбцы сложатся, а если вектор умножить на скаляр, то его координатный столбец умножится на этот же скаляр. Есть более короткий способ выразить указанные свойства: сопоставление вектору $v \in V$ его координатного столбца *линейно*. Сформулируем это в виде теоремы.

Теорема 7.4.4. *Пусть V — конечномерное векторное пространство над полем k; $\mathcal{B} = \{v_1, \ldots, v_n\}$ — его базис. Отображение*

$$V \to k^n,$$
$$v \mapsto [v]_{\mathcal{B}}$$

линейно.

Доказательство. Фактически, нам нужно показать, что если $v, v' \in V$, $a \in k$, то $[v + v']_{\mathcal{B}} = [v]_{\mathcal{B}} + [v']_{\mathcal{B}}$ и $[va]_{\mathcal{B}} = [v]_{\mathcal{B}} \cdot a$. Пусть

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix}, \quad [v']_{\mathcal{B}} = \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix}.$$

По определению это означает, что

$$v = v_1 a_1 + \cdots + v_n a_n,$$
$$v' = v_1 b_1 + \cdots + v_n b_n.$$

Сложим эти два равенства:

$$v + v' = v_1(a_1 + b_1) + \cdots + v_m(a_n + b_n).$$

Но тогда

$$[v + v']_{\mathcal{B}} = \begin{pmatrix} a_1 + b_1 \\ \vdots \\ a_n + b_n \end{pmatrix} = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ \vdots \\ b_n \end{pmatrix} = [v]_{\mathcal{B}} + [v']_{\mathcal{B}},$$

что и требовалось. Доказательство для умножения на скаляр совершенно аналогично и оставляется читателю в качестве упражнения. □

Теперь мы готовы сделать последний шаг в установлении соответствия между действиями с векторными пространствами с одной стороны, и вычислениями с матрицами с другой стороны.

Теорема 7.4.5. *Пусть $T\colon V \to W$ — линейное отображение между конечномерными пространствами V и W, и пусть $\mathcal{B} = (v_1, \ldots, v_n)$ — базис V, а $\mathcal{B}' = (w_1, \ldots, v_m)$ — базис W. Тогда*

$$[Tv]_{\mathcal{B}'} = [T]_{\mathcal{B},\mathcal{B}'} \cdot [v]_{\mathcal{B}}$$

для любого вектора $v \in V$.

Доказательство. Пусть $v = v_1 c_1 + \cdots + v_n c_n$, то есть,

$$[v]_{\mathcal{B}} = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix},$$

и пусть $[T]_{\mathcal{B},\mathcal{B}'} = (a_{ij})$ — матрица отображения T. Тогда

$$T(v) = T(\sum_{j=1}^{n} v_j c_j) = \sum_{j=1}^{n} T(v_j) c_j = \sum_{j=1}^{n} \left(\sum_{i=1}^{m} w_i a_{ij} \right) c_j = \sum_{i=1}^{m} w_i \left(\sum_{j=1}^{n} a_{ij} c_j \right).$$

Значит, i-я координата вектора $T(v)$ в базисе $\mathcal{B}'$ равна $\sum_{j=1}^{n} a_{ij} c_j$. Но это и означает, что столбец $[T(v)]_{\mathcal{B}'}$ равен произведению матрицы $(a_{ij}) = [T]_{\mathcal{B},\mathcal{B}'}$ на столбец $[v]_{\mathcal{B}}$. □

7.5 Изоморфизм

Определение 7.5.1. Линейное отображение $T\colon V \to W$ называется **обратимым**, если существует линейное отображение $S\colon W \to V$ такое, что $S \circ T = \mathrm{id}_V$ и $T \circ S = \mathrm{id}_W$. Такое S называется **обратным** к T.

Предложение 7.5.2. *Линейное отображение* $T\colon V \to W$ *обратимо тогда и только тогда, когда оно биективно.*

Доказательство. Если T обратимо, то обратное к нему является обратным отображением в теоретико-множественном смысле (определение 1.3.4), и потому биективно по теореме 1.3.6.

Если же отображение T биективно, то (снова по теореме 1.3.6) существует отображение множеств $S\colon W \to V$ такое, что $S \circ T = \mathrm{id}_V$ и $T \circ S = \mathrm{id}_W$. Можно и явно построить это S: для каждого $w \in W$ заметим, что (по определению биективности) существует единственное $v \in V$ такое, что $T(v) = w$; тогда положим $S(w) = v$. Осталось проверить, что это отображение линейно. Действительно, возьмем $w_1, w_2 \in W$ и пусть $S(w_1) = v_1$, $S(w_2) = v_2$. Это означает, что $T(v_1) = w_1$, $T(v_2) = w_2$. Но тогда $T(v_1 + v_2) = w_1 + w_2$, и потому $S(w_1 + w_2) = v_1 + v_2 = S(w_1) + S(w_2)$. Кроме того, если $w \in W$ и $a \in k$, пусть $S(w) = v$. Это означает, что $T(v) = w$, откуда $T(va) = wa$, и, стало быть, $S(wa) = va = S(w)a$. □

Определение 7.5.3. Обратимое линейное отображение иногда называется **изоморфизмом**. Если между пространствами V и W существует изоморфизм $T\colon V \to W$, они называются **изоморфными**. Обозначение: $V \cong W$.

Теорема 7.5.4. *Два конечномерных векторных пространства над* k *изоморфны тогда и только тогда, когда их размерности равны.*

Доказательство. Пусть $V \cong W$, то есть, существует обратимое линейное отображение $T\colon V \to W$. По предложению 7.5.2 T биективно. В частности, T инъективно, и потому $\mathrm{Ker}(T) = 0$ (теорема 7.3.5); кроме того, T сюръективно, и потому $\mathrm{Im}(T) = W$. Воспользуемся теоремой о гомоморфизме 7.3.8:

$$\dim \mathrm{Ker}(T) + \dim \mathrm{Im}(T) = \dim(V).$$

В нашем случае $\dim \mathrm{Ker}(T) = 0$ и $\dim \mathrm{Im}(T) = \dim W$; поэтому $\dim V = \dim W$, что и требовалось.

Обратно, пусть $\dim V = \dim W = n$. Выберем базис $v_1, \ldots, v_n$ в V и базис $w_1, \ldots, w_n$ в W. По теореме 7.1.9 для задания линейного отображения $T\colon V \to W$ достаточно задать

$T(v_i)$ для всех i. Положим $T(v_i) = w_i$ и покажем, что полученное отображение T является изоморфизмом. Для этого (по предложению 7.5.2) достаточно проверить, что оно инъективно и сюръективно.

Для инъективности (по предложению 7.3.5) нужно показать, что $\mathrm{Ker}(T) = 0$. Возьмем $v \in \mathrm{Ker}(T)$. Разложим v по базису пространства V: $v = v_1 a_1 + \cdots + v_n a_n$. Тогда $0 = T(v) = T(v_1)a_1 + \cdots + T(v_n)a_n = w_1 a_1 + \cdots + w_n a_n$. Но элементы $w_1, \ldots, w_n \in W$ образуют базис, и потому линейно независимы. Их линейная комбинация оказалась равна нулю — поэтому все ее коэффициенты равны нулю: $a_1 = \cdots = a_n = 0$. Но тогда и $v = 0$.

Осталось проверить, что T сюръективно. Но любой вектор W есть линейная комбинация векторов $w_1, \ldots, w_n$, поэтому является образом соответствующей линейной комбинации векторов $v_1, \ldots, v_n$. □

Следствие 7.5.5. *Любое конечномерное векторное пространство V изоморфно пространству k^n, где $n = \dim(V)$. Более того, если $\mathcal{B}$ — некоторый базис пространства V, то отображение $\varphi\colon v \mapsto [v]_{\mathcal{B}}$ устанавливает изоморфизм между V и k^n.*

Доказательство. Пусть $\dim(V) = n$; тогда $\dim(k^n) = n = \dim(V)$, и по теореме 7.5.4 пространства V и k^n изоморфны.

Для доказательства второго утверждения обозначим элементы базиса $\mathcal{B}$ через $v_1, \ldots, v_n$. Мы уже знаем, что отображение $v \mapsto [v]_{\mathcal{B}}$ линейно (теорема 7.4.4); проверим, что это изоморфизм. Для этого нужно проверить, что его ядро тривиально, а образ совпадает с k^n. Возьмем $v \in \mathrm{Ker}(\varphi)$; это означает, что столбец координат вектора v нулевой. Но тогда по определению координат $v = v_1 0 + \cdots + v_n 0 = 0$. Значит, $\mathrm{Ker}(\varphi) = 0$. Пусть теперь $w \in k^n$ — некоторый столбец, состоящий из скаляров $a_1, \ldots, a_n$. Рассмотрим вектор $v = v_1 a_1 + \cdots + v_n a_n \in V$. Легко видеть, что $[v]_{\mathcal{B}} = w$, что доказывает сюръективность отображения φ. □

Таким образом, любое конечномерное пространство изоморфно пространству столбцов. Подчеркнем, что этот изоморфизм зависит от выбора базиса (в таком случае говорят, что этот изоморфизм *не является каноническим*): в разных базисах один и тот же вектор, как правило, имеет разные наборы координат.

Теорема 7.5.6. *Пусть V, W — конечномерные векторные пространства над полем k. Пространство $\mathrm{Hom}(V, W)$ линейных отображений из V в W изоморфно векторному пространству $M(m, n, k)$ матриц размера $m \times n$ над k, где $m = \dim W$, $n = \dim V$. Более того, если $\mathcal{B}, \mathcal{B}'$ — базисы в V, W соответственно, то отображение $\varphi\colon T \mapsto [T]_{\mathcal{B},\mathcal{B}'}$ устанавливает изоморфизм между $\mathrm{Hom}(V, W)$ и $M(m, n, k)$.*

Доказательство. Мы сразу докажем второе утверждение. Обозначим элементы $\mathcal{B}$ через $v_1, \ldots, v_n$, а элементы $\mathcal{B}'$ через $w_1, \ldots, w_m$. По теореме 7.4.2 отображение φ линейно. Проверим, что его ядро тривиально, а образ совпадает с $M(m, n, k)$. Пусть $T \in \mathrm{Ker}(\varphi)$. Это значит, что у линейного отображения T матрица нулевая. По определению матрицы это значит, что все координаты вектора $T(v_j)$ в базисе $\mathcal{B}'$ равны нулю, а потому $T(v_j) = 0$ для всех j. Но мы знаем одно такое линейное отображение: это $0 \in \mathrm{Hom}(V, W)$. По единственности в универсальном

свойстве базиса (теорема 7.1.9) $T = 0$. Наконец, пусть $A = (a_{ij}) \in M(m, n, k)$ — некоторая матрица. Мы утверждаем, что существует линейное отображение $T\colon U \to V$, матрица которого в базисах $\mathcal{B}, \mathcal{B}'$ совпадает с A. Действительно, положим $T(v_j) = w_1 a_1 + \cdots + w_m a_m$. По теореме 7.1.9 это однозначно определяет линейное отображение T, и очевидно, что $[T]_{\mathcal{B},\mathcal{B}'} = A$. □

Следствие 7.5.7. *Если пространства V, W конечномерны, то* $\dim \operatorname{Hom}(V, W) = \dim V \cdot \dim W$.

Доказательство. Очевидно, что размерность пространства матриц $M(m, n, k)$ равна mn; осталось применить теорему 7.5.6 и теорему 7.5.4. □

Важный частный случай понятия линейного отображения — *линейный оператор*.

Определение 7.5.8. Линейное отображение $T\colon V \to V$ называется **линейным оператором** на пространстве V, или **эндоморфизмом** пространства V.

Предложение 7.5.9. *Пусть $T\colon V \to V$ — линейный оператор на конечномерном пространстве V. Следующие утверждения равносильны.*

1. *Отображение T биективно.*
2. *Отображение T инъективно.*
3. *Отображение T сюръективно.*

Доказательство. Очевидно, что из (1) следуют (2) и (3). Покажем, что из (2) следует (1). Если T инъективно, то $\operatorname{Ker} T = 0$ (предложение 7.3.5). По теореме о гомоморфизме (теорема 7.3.8) $\dim \operatorname{Ker} T + \dim \operatorname{Im} T = \dim V$. Первое слагаемое равно нулю, поэтому $\dim \operatorname{Im} T = \dim V$. В то же время, $\operatorname{Im} T$ — подпространство в V, и по предложению 6.5.3 из совпадения размерностей следует, что $\operatorname{Im} T = V$, что означает сюръективность, а потому и биективность отображения T.

Осталось показать, что из (3) следует (1). Снова воспользуемся теоремой о гомоморфизме: $\dim \operatorname{Ker} T + \dim \operatorname{Im} T = \dim V$. Теперь по предположению $\operatorname{Im} T = \dim V$, и, стало быть, $\dim \operatorname{Ker} T = 0$. Значит, подпространство $\operatorname{Ker} T$ тривиально, и потому T инъективно и, следовательно, биективно. □

Теорема 7.5.10. *Пусть V — векторное пространство. Множество $\operatorname{Hom}(V, V)$ всех линейных операторов на V образует ассоциативное кольцо с единицей относительно сложения и композиции.*

Доказательство. Мы уже знаем, что сложение линейных отображений ассоциативно, коммутативно, обладает нейтральным элементом 0 и обратными элементами. Кроме того, композиция (которая играет роль умножения) ассоциативна и обладает нейтральным элементом id_V. Осталось проверить левую и правую дистрибутивность. Ограничимся проверкой одной из них. Пусть $S, T, U \in \operatorname{Hom}(V, V)$. Для каждого $v \in V$ выполнено

$$(S\circ(T+U))(v) = S((T+U)(v)) = S(T(v)+U(v)) = S(T(v))+S(U(v)) = (S\circ T)(v)+(S\circ U)(v) = (S\circ T+S\circ U)(v),$$

а потому отображения $S \circ (T + U)$ и $S \circ T + S \circ U$ совпадают. □

Отметим, что в конечномерном случае кольцо операторов на V *изоморфно* кольцу квадратных матриц порядка $n = \dim V$ (см. замечание 5.3.5). Поясним, что означает слово «изоморфизм» в этом контексте (пока мы обсуждали только изоморфизм векторных пространств, но не колец). Пусть $\mathcal{B}$ — базис пространства V, и $\dim V = n$. Из теоремы 7.5.6 следует, что отображение $T \mapsto [T]_{\mathcal{B}}$ является биекцией между $\mathrm{Hom}(V, V)$ и $M(n, n, k)$, переводящей сложение в сложение. Кроме того, по теореме 7.4.3 она переводит композицию операторов в умножение. Наконец, тождественный оператор переходит при этом отображении в единичную матрицу. Мы получили биекцию между кольцами, которая сохраняет все операции (включая «взятие единичного элемента»). Такая биекция и называется «изоморфизмом колец»; ее существование означает, что указанные кольца «ведут себя одинаково».

7.6 Ранг матрицы

Литература: [F], гл. IV, § 3, пп. 4–6; [K1], гл. 2, § 2, п. 1–2; [vdW], гл. IV, §§ 22, 23.

Первым приложением теории векторных пространств для нас станет определение ранга матрицы, которые мы неформально обсуждали после доказательства теоремы 5.4.1. Напомним, что любую матрицу $A \in M(m, n, k)$ можно представить в виде $A = P \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q$, где P, Q — некоторые обратимые матрицы. Мы покажем, что на самом деле натуральное число r не зависит от выбора такого представления, и поэтому имеет право называться *рангом* матрицы A. Для этого мы введем еще несколько понятий ранга, и покажем, что все они совпадают друг с другом.

Определение 7.6.1. Пусть $A = (a_{ij}) \in M(m, n, k)$. Линейная оболочка столбцов матрицы A называется **пространством столбцов матрицы** A; по определению оно является подпространством в k^m. Иными словами, это пространство

$$\left\langle \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\rangle \leqslant k^m.$$

Линейная оболочка строк матрицы A называется **пространством строк матрицы** A; по определению оно является подпространством в ${}^n k$. Иными словами, это пространство

$$\left\langle \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \end{pmatrix}, \dots, \begin{pmatrix} a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix} \right\rangle \leqslant {}^n k.$$

Таким образом, пространство столбцов состоит из всевозможных линейных комбинаций столбцов матрицы A; аналогично и со строками.

Определение 7.6.2. Столбцовым рангом матрицы A называется размерность ее пространства столбцов; **строчным рангом** A называется размерность ее пространства строк.

Очевидно, что столбцовый ранг матрицы $A \in M(m, n, k)$ не превосходит n, а ее строчный ранг не превосходит m. Для определения следующего понятия — *тензорного ранга* — необходимо сначала определить матрицы ранга 1.

Определение 7.6.3. Матрица $A \in M(m, n, k)$ называется **матрицей ранга** 1, если $A \neq 0$ и A можно представить в виде $A = uv$, где $u \in k^m$, $v \in {}^n k$. **Тензорным рангом** матрицы A называется наименьшее натуральное число r такое, что A можно представить в виде суммы r матриц ранга 1. Иными словами, тензорный ранг A — это наименьшее r, при котором существуют столбцы $u_1, \ldots, u_r \in k^m$ и строки $v_1, \ldots v_r \in {}^n k$ такие, что $A = u_1 v_1 + \cdots + u_r v_r$.

Заметим, что тензорный ранг матрицы $A \in M(m, n, k)$ определен: он не превосходит mn. Действительно, несложно представить матрицу $A = (a_{ij})$ в виде суммы mn матриц ранга 1: мы видели, что $A = \sum_{i,j} a_{ij} e_{ij}$, а матрица $a_{ij} e_{ij}$ имеет ранг 1:

$$a_{ij} e_{ij} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ a_{ij} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \cdot \begin{pmatrix} 0 & \ldots & 0 & 1 & 0 & \ldots & 0 \end{pmatrix}.$$

Здесь в столбце высоты m элемент a_{ij} стоит в позиции i, и в строке длины n элемент 1 стоит в позиции j.

Теорема 7.6.4. *Тензорный ранг матрицы не изменяется при домножении ее слева или справа на обратимую матрицу. В частности, тензорный ранг матрицы сохраняется при элементарных преобразованиях ее строк и столбцов.*

Доказательство. Пусть $A \in M(m, n, k)$ — матрица тензорного ранга r. Тогда мы можем записать $A = u_1 v_1 + \cdots + u_r v_r$ для некоторых столбцов $u_1, \ldots, u_r \in k^m$ и строк $v_1, \ldots, v_r \in {}^n k$. Если матрица $B \in M(m, k)$ обратима, то $BA = B(u_1 v_1 + \cdots + u_r v_r) = (Bu_1)v_1 + \cdots + (Bu_r)v_r$ — сумма r матриц ранга 1, поэтому тензорный ранг BA не превосходит r. С другой стороны, если тензорный ранг BA меньше r, то можно записать $BA = u'_1 v'_1 + \cdots + u'_p v'_p$ для $p < r$ и после домножения на B^{-1} слева мы получили бы, что A является суммой p матриц ранга 1 — противоречие. Доказательство для домножения на обратимую матрицу справа совершенно аналогично. □

Теорема 7.6.5. *Тензорный ранг матрицы равен ее строчному рангу и столбцовому рангу.*

Доказательство. Пусть размерность пространства строк матрицы $A \in M(m, n, k)$ равна d. Это значит, что каждая строка матрицы A является некоторой линейной комбинацией строк $v_1, \ldots, v_d \in {}^n k$. Запишем эту линейную комбинацию: $a_{i*} = \lambda_{i1} v_1 + \cdots + \lambda_{id} v_d$. Заметим, что

$A = e_1a_{1*} + e_2a_{2*} + \cdots + e_ma_{m*}$, где $e_i = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ — стандартный базисный столбец в k^m. Таким образом,

$$A = e_1(\lambda_{11}v_1 + \cdots + \lambda_{1d}v_d) + \cdots + e_m(\lambda_{21}v_1 + \cdots + \lambda_{md}v_d).$$

Раскрывая скобки, получаем, что $A = u_1v_1 + \cdots + u_dv_d$ для некоторых столбцов $u_1, \ldots, u_d \in k^m$. Поэтому тензорный ранг A не превосходит d.

Обратно, если r — тензорный ранг матрицы A, то $u_1v_1 + \cdots + u_rv_r$, поэтому каждая строка матрицы A является линейной комбинацией строк $v_1, \ldots, v_r$. Это означает, что $v_1, \ldots, v_r$ — система образующих пространства строк матрицы A. В силу следствия 6.3.14 получаем, что $d \leqslant r$.

Доказательство для столбцового ранга совершенно аналогично (или можно заметить, что тензорный ранг не меняется при транспонировании). □

Определение 7.6.6. Общее значение тензорного, строчного и столбцового рангов матрицы A называется ее **рангом** и обозначается через $\mathrm{rk}(A)$.

Теперь мы можем связать понятие тензорного ранга с понятием ранга, введенным после доказательства следствия 5.4.2.

Следствие 7.6.7. *Пусть матрица $A \in M(m, n, k)$ представлена в виде $A = PDQ$, где $P \in M(m, k)$, $Q \in M(n, k)$ — обратимые матрицы, а $D = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$ — окаймленная единичная матрица. Тогда r равно тензорному рангу матрицы A.*

Доказательство. По теореме 7.6.5 тензорный ранг матрицы A равен тензорному рангу матрицы $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$; с другой стороны, очевидно, что строчный ранг этой матрицы равен r. □

Следствие 7.6.8. *Матрица $A \in M(n, k)$ обратима тогда и только тогда, когда ее ранг равен n.*

Доказательство. Простая комбинация следствия 5.4.3 и следствия 7.6.7. □

Теорема 7.6.9 (Кронекера–Капелли). *Система линейных уравнений имеет решение (**совместна**) тогда и только тогда, когда ранг матрицы этой системы равен рангу ее расширенной матрицы. Если, кроме того, этот ранг равен количеству неизвестных, то система имеет единственное решение.*

Доказательство. Рассмотрим систему линейных уравнений $AX = B$. Пусть $u_1, \dots, u_n$ — столбцы матрицы A. Система $AX = B$ имеет решение тогда и только тогда, когда существуют $x_1, \dots, x_n \in k$ такие, что $u_1x_1 + \dots + u_nx_n = B$. Это, в свою очередь равносильно тому, что B лежит в линейной оболочке векторов $u_1, \dots, u_n$, то есть, тому, что $\langle u_1, \dots, u_n \rangle = \langle u_1, \dots, u_n, B \rangle$. Это равенство и означает совпадение [столбцовых] рангов матриц A и $(A|B)$.

Если же ранг равен количеству неизвестных n, то пространство $\langle u_1, \dots, u_n \rangle$ имеет размерность n. При этом $\langle u_1, \dots, u_n \rangle$ — его система образующих, и из нее можно выбрать базис, в котором должно быть n элементов. Значит, $u_1, \dots, u_n$ образуют базис пространства столбцов матрицы A. Поэтому вектор B имеет единственное представление в виде $B = u_1x_1 + \dots + u_nx_n$, что и означает единственность решения системы. □

7.7 Фактор-пространство

Литература: [F], гл. XII, § 2, п. 5; [K2], гл. 1, § 2, п. 6; [KM], ч. 1, § 6.

Определение 7.7.1. Пусть V — векторное пространство над полем k, $U \leqslant V$. Будем говорить, что элементы $v_1, v_2 \in V$ **сравнимы по модулю** U, если $v_1 - v_2 \in U$. Обозначения: $v_1 \sim_U v_2$, $v_1 \sim v_2$ (если понятно, по модулю какого подпространства рассматривается сравнение).

Пользуясь определением подпространства, несложно проверить, что сравнение по модулю подпространства $U \leqslant V$ является отношением эквивалентности на V. Действительно, это отношение рефлексивно: $v \sim v$, поскольку $v - v = 0 \in U$. Оно симметрично: если $v_1 \sim v_2$, то $v_1 - v_2 \in U$; тогда и $v_2 - v_1 = (v_1 - v_2) \cdot (-1) \in U$. Наконец, если $v_1 \sim v_2$ и $v_2 \sim v_3$, то $v_1 - v_2 \in U$ и $v_2 - v_3 \in U$; отсюда $v_1 - v_3 = (v_1 - v_2) + (v_2 - v_3) \in U$, поэтому $v_1 \sim v_3$.

Раз мы получили отношение эквивалентности, то по теореме 1.5.4 сразу получаем разбиение на классы эквивалентности. Мы будем обозначать класс эквивалентности элемента $v \in V$ по отношению $\sim_U$ через $\overline{v}$ или через $v + U$. Последнее обозначение имеет также следующий смысл: для любых подмножеств $S, T \subseteq V$ можно определить их сумму $S + T = \{s + t \mid s \in S, t \in T\}$ и результат умножения на скаляр $\lambda \in k$: $S\lambda = \{s\lambda \mid s \in S\}$. В этих обозначениях класс эквивалентности $v + U$ — это в точности $\{v\} + U = \{v + u \mid u \in U\}$.

Фактор-множество множества V по отношению эквивалентности $\sim_U$ мы будем обозначать через V/U. Наша ближайшая цель — ввести на нем структуру векторного пространства. Для этого необходимо определить сумму классов и результат умножения класса на скаляр из k. Это, как и в случае построения кольца классов вычетов (см. п. 2.7), осуществляется с помощью операций над представителями классов: чтобы сложить два элемента фактор-пространства, посмотрим, в каком классе лежит сумма двух [любых] представителей этих элементов; чтобы умножить элемент на скаляр, умножим любой его представитель на этот скаляр и посмотрим на класс результата. Точнее, положим $(v_1 + U) + (v_2 + U) = (v_1 + v_2) + U$ и $(v + U)a = va + U$ для любых $v, v_1, v_2 \in V$ и $a \in k$. В других обозначениях, $\overline{v_1} + \overline{v_2} = \overline{v_1 + v_2}$ и $\overline{v} \cdot a = \overline{v \cdot a}$. Как всегда, необходимо проверить *корректность* данного определения, то есть, тот факт, что результат операций не зависит от выбора представителей. Это делается совершенно прямолинейно, поэтому мы оставляем проверку читателю в качестве упражнения. Наконец, проверим, что полученные операции превращают V/U в векторное пространство над k.

Предложение 7.7.2. *Пусть* V *— векторное пространство над полем* k, $U \leqslant V$. *Фактор-множество* V/U *вместе с введенными выше операциями является векторным пространством над* k.

Доказательство. Все проверки тривиальны; приведем выкладки с минимальными комментариями.

1. $(\overline{v_1} + \overline{v_2}) + \overline{v_3} = \overline{v_1 + v_2} + \overline{v_3} = \overline{(v_1 + v_2) + v_3} = \overline{v_1 + (v_2 + v_3)} = \overline{v_1} + \overline{v_2 + v_3} = \overline{v_1} + (\overline{v_2} + \overline{v_3})$.
2. $\overline{v} + \overline{0} = \overline{v + 0} = \overline{v}$, поэтому $\overline{0} \in V/U$ играет роль нейтрального элемента по сложению.
3. $\overline{v} + \overline{-v} = \overline{v + (-v)} = \overline{0}$, поэтому $\overline{-v}$ — обратный по сложению к $\overline{v}$.
4. $\overline{v_1} + \overline{v_2} = \overline{v_1 + v_2} = \overline{v_2 + v_1} = \overline{v_2} + \overline{v_1}$.
5. $(\overline{v_1} + \overline{v_2}) \cdot a = \overline{v_1 + v_2} \cdot a = \overline{(v_1 + v_2) \cdot a} = \overline{v_1 a + v_2 a} = \overline{v_1 a} + \overline{v_2 a} = \overline{v_1} \cdot a + \overline{v_2} \cdot a$.
6. $\overline{v}(a + b) = \overline{v(a + b)} = \overline{va + vb} = \overline{va} + \overline{vb} = \overline{v} \cdot + \overline{v} \cdot b$.
7. $\overline{v}(ab) = \overline{v(ab)} = \overline{(va)b} = \overline{va} \cdot b = (\overline{v} \cdot a) \cdot b$.
8. $\overline{v} \cdot 1 = \overline{v \cdot 1} = \overline{v}$.

□

С каждым отношением эквивалентности связана каноническая проекция исходного множества на фактор-множество. В нашем случае она является отображением $V \to V/U$, сопоставляющим вектору $v \in V$ его класс $\overline{v} = v + U$. Нетрудно видеть, что это отображение является линейным: действительно, $\overline{v_1 + v_2} = \overline{v_1} + \overline{v_2}$ и $\overline{v\lambda} = (\overline{v})\lambda$ просто по определению операций в фактор-пространстве.

Теорема 7.7.3 (Теорема о гомоморфизме). *Пусть* $\varphi\colon U \to V$ *— линейное отображение. Тогда* $U/\operatorname{Ker}(\varphi) \cong \operatorname{Im}(\varphi)$.

Доказательство. Построим отображение $f\colon U/\operatorname{Ker}(\varphi) \to \operatorname{Im}(\varphi)$: отправим класс $u + \operatorname{Ker}(\varphi)$ в $\varphi(u) \in \operatorname{Im}(\varphi)$. Проверим, что f корректно определено, то есть, не зависит от выбора представителя класса из $U/\operatorname{Ker}(\varphi)$. Действительно, если $u + \operatorname{Ker}(\varphi) = u' + \operatorname{Ker}(\varphi)$, то $u' - u \in \operatorname{Ker}(\varphi)$, откуда $0 = \varphi(u' - u) = \varphi(u') - \varphi(u)$. Значит, $\varphi(u') = \varphi(u)$, что и требовалось.

Отображение f является линейным. Действительно, если $u_1, u_2 \in U$, то $f(\overline{u_1}) = \varphi(u_1)$ и $f(\overline{u_2}) = \varphi(u_2)$, поэтому $f(\overline{u_1}) + f(\overline{u_2}) = \varphi(u_1) + \varphi(u_2)$. С другой стороны, $f(\overline{u_1} + \overline{u_2}) = f(\overline{u_1 + u_2}) = \varphi(u_1 + u_2) = \varphi(u_1) + \varphi(u_2)$ — то же самое. Наконец, если $u \in U$ и $a \in k$, то $f(\overline{u})a = \varphi(u)a$ и $f(\overline{u} \cdot a) = f(\overline{ua}) = \varphi(ua) = \varphi(u)a$.

Проверим, что f биективно. Заметим, что из $\varphi(u) = 0$ следует, что $u \in \operatorname{Ker}(\varphi)$, то есть, что $\overline{u} = \overline{0} \in U/\operatorname{Ker}(\varphi)$; поэтому f инъективно. С другой стороны, для каждого $v \in \operatorname{Im}(\varphi)$ существует $u \in U$ такое, что $v = \varphi(u)$. Тогда $f(\overline{u}) = \varphi(u) = v$, поэтому f сюръективно. □

7.8 Относительный базис

Литература: [F], гл. XII, § 2, пп. 4–6; [K2], гл. 1, § 2, пп. 4, 5.

Пусть V — векторное пространство над полем k, $U \leqslant V$.

Определение 7.8.1. Набор векторов $v_1, \ldots, v_n \in V$ называется **линейно независимым над** U, если из $v_1a_1 + \ldots v_na_n \in U$ следует, что $a_1 = \cdots = a_n = 0$. Набор векторов $v_1, \ldots, v_n \in V$ называется **порождающей системой над** U (или **системой образующих** V **над** U), если любой вектор из V можно представить в виде $v_1a_1 + \cdots + v_na_n + u$ для некоторых $a_1, \ldots, a_n \in k$ и $u \in U$. Наконец, набор $v_1, \ldots, v_n \in V$ называется **относительным базисом** V **над** U, если он линейно независим над U и является порождающей системой над U. Нетрудно видеть, что это равносильно тому, что любой вектор V представляется в виде $v_1a_1 + \cdots + v_na_n + u$ для некоторого $u \in U$ *единственным образом.*

Теорема 7.8.2. *Следующие условия равносильны:*

1. $v_1, \ldots, v_n$ — *относительный базис* V *над* U*;*
2. $v_1 + U, \ldots, v_n + U$ — *базис фактор-пространства* V/U*;*
3. $v_1, \ldots, v_n$ *вместе с некоторым базисом пространства* U *в совокупности образуют базис пространства* V*;*
4. $v_1, \ldots, v_n$ — *базис некоторого дополнения* U *в* V*.*

Доказательство. $1 \Rightarrow 2$ Пусть $v_1, \ldots, v_n$ — относительный базис V над U. Проверим, что система $v_1+U, \ldots, v_n+U$ линейно независима. Действительно, если $(v_1+U)a_1+\cdots+(v_n+U)a_n = 0 \in V/U$, то $(v_1a_1+\cdots+v_na_n)+U = 0 \in V/U$. Это означает, что $v_1a_1+\cdots+v_na_n \in U$, откуда по определению линейной независимости над U следует $a_1 = \cdots = a_n = 0$. Кроме того, любой вектор $v \in V$ можно представить в виде $v = v_1a_1 + \cdots + v_na_n + u$ для некоторых $a_1, \ldots, a_n \in k$ и $u \in U$. Тогда $\overline{v} = \overline{v_1}a_1 + \cdots + \overline{v_n}a_n$, поскольку $\overline{u} = 0$. Значит, $\overline{v_1}, \ldots, \overline{v_n}$ — система образующих V/U.

$2 \Rightarrow 3$ Пусть $v_1 + U, \ldots, v_n + U$ — базис V/U, $u_1, \ldots, u_k$ — некоторый базис U. Тогда для любого вектора $v \in V$ класс $v + U \in V/U$ можно представить в виде $v + U = (v_1 + U)a_1 + \cdots + (v_n + U)a_n = (v_1a_1 + \cdots + v_na_n) + U$. Поэтому $v \sim_U v_1a_1 + \cdots + v_na_n$ и $v - (v_1a_1 + \cdots + v_na_n) = u \in U$. Разложим вектор u по базису $u_1, \ldots, u_k$: $u = u_1b_1 + \cdots + u_kb_k$. Получаем, что $v = v_1a_1 + \cdots + v_na_n + u_1b_1 + \cdots + u_kb_k$. Это доказывает, что $v_1, \ldots, v_n, u_1, \ldots, u_k$ — базис V. Наконец, если $v_1a_1 + \cdots + v_na_n + u_1b_1 + \cdots + u_kb_k = 0$, то $v_1a_1 + \cdots + v_na_n = -u_1b_1 - \cdots - u_kb_k \in U$, поэтому $\overline{v_1a_1 + \cdots + v_na_n} = \overline{0}$, и в силу линейной независимости $\overline{v_1}, \ldots, \overline{v_n}$ в V/U из этого следует, что $a_1 = \cdots = a_n = 0$.

$3 \Rightarrow 4$ Пусть $u_1, \ldots, u_k$ — базис U такой, что $v_1, \ldots, v_n, u_1, \ldots, u_k$ — базис V. Тогда $\langle v_1, \ldots, v_n\rangle + \langle u_1, \ldots, u_k\rangle = V$, откуда $\langle v_1, \ldots, v_n\rangle$ — дополнение к U в V.

$4 \Rightarrow 1$ Пусть $\langle v_1, \ldots, v_n \rangle = U'$; по предположению, $V = U \oplus U'$. Если $v = v_1 a_1 + \cdots + v_n a_n \in U$, то $v \in U \cap U'$, откуда $v = 0$, и в силу линейной независимости v_i, получаем $a_1 = \cdots = a_n = 0$. Наконец, любой вектор $v \in V$ можно представить в виде $v = u + u'$ для некоторых $u \in U$, $u' \in U'$. Запишем $u' = v_1 a_1 + \cdots + v_n a_n$; получаем, что $v = v_1 a_1 + \cdots + v_n a_n + u$.

□

Следствие 7.8.3. *Пусть* $U \leqslant V$ — *векторные пространства. Тогда* $\dim(V/U) = \dim(V) - \dim(U)$.

Доказательство. Выберем базис $u_1, \ldots, u_k$ в U и базис $\overline{v_1}, \ldots, \overline{v_n}$ в V/U. По части 3 теоремы 7.8.2 набор $u_1, \ldots, u_k, v_1, \ldots, v_n$ является базисом в V, состоящим из $k + n$ элементов. □

7.9 Матрица перехода

Литература: [F], гл. XII, § 1, п. 4; [K2], гл. I, § 2, п. 3; [KM], ч. 1, § 4, п. 7.

Напомним, что выбор базиса $\mathcal{B}$ в конечномерном пространстве V, $\dim(V) = n$, задает изоморфизм между V и пространством столбцов k^n: у каждого вектора v появляется координатный столбец $[v]_{\mathcal{B}}$, состоящий из n координат вектора v в базисе $\mathcal{B}$.

Пусть теперь $\mathcal{B}'$ — еще один базис пространства V. Возникает естественный вопрос: как связаны между собой координаты вектора v в базисах $\mathcal{B}$ и $\mathcal{B}'$? Ответ на этот вопрос формулируется с помощью *матрицы перехода* между базисами.

Определение 7.9.1. Пусть $\mathcal{B} = \{u_1, \ldots, u_n\}$, $\mathcal{B}' = \{v_1, \ldots, v_n\}$ — базисы конечномерного пространства V. В частности, векторы v_j можно разложить по базису $\mathcal{B}$:

$$v_j = \sum_{i=1}^{n} u_i c_{ij}.$$

Матрица $C = (c_{ij})_{i,j=1}^{n}$, составленная из коэффициентов этих разложений, называется **матрицей перехода** от базиса $\mathcal{B}$ к базису $\mathcal{B}'$ и обозначается через $(\mathcal{B} \rightsquigarrow \mathcal{B}')$. Иными словами, матрица $(\mathcal{B} \rightsquigarrow \mathcal{B}')$ составлена из координатных столбцов векторов $v_1, \ldots, v_n$ в базисе $\mathcal{B}$:

$$(\mathcal{B} \rightsquigarrow \mathcal{B}') = \begin{pmatrix} [v_1]_{\mathcal{B}} & [v_2]_{\mathcal{B}} & \ldots & [v_n]_{\mathcal{B}} \end{pmatrix}.$$

В этой ситуации $\mathcal{B}$ называется **старым базисом**, $\mathcal{B}'$ — **новым базисом**, а $(\mathcal{B} \rightsquigarrow \mathcal{B}')$ — **матрицей перехода от старого базиса к новому**.

Символически мы можем записать

$$\begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} = \begin{pmatrix} u_1 & u_2 & \ldots & u_n \end{pmatrix} \cdot (\mathcal{B} \rightsquigarrow \mathcal{B}').$$

В такой записи слева стоит строчка, составленная из *векторов* пространства V, а справа — произведение такой строчки на матрицу над k. Перемножая строчку векторов на столбцы матрицы над k мы будем получать линейные комбинации этих векторов, поэтому в правой части после перемножения окажется строчка, состоящая из n линейных комбинаций векторов $u_1, \ldots, u_n$. Равенство теперь означает, что вектор v_i равен i-й их этих линейных комбинаций.

Предложение 7.9.2 (Свойства матрицы перехода). *Пусть* $\mathcal{B} = \{u_1, \dots, u_n\}$, $\mathcal{B}' = \{v_1, \dots, v_n\}$, $\mathcal{B}'' = \{w_1, \dots, w_n\}$ — *базисы конечномерного пространства* V. *Тогда*

1. $(\mathcal{B} \rightsquigarrow \mathcal{B}) = E$;
2. $(\mathcal{B} \rightsquigarrow \mathcal{B}'') = (\mathcal{B} \rightsquigarrow \mathcal{B}') \cdot (\mathcal{B}' \rightsquigarrow \mathcal{B}'')$;
3. *матрица* $(\mathcal{B} \rightsquigarrow \mathcal{B}')$ *обратима и* $(\mathcal{B} \rightsquigarrow \mathcal{B}')^{-1} = (\mathcal{B}' \rightsquigarrow \mathcal{B})$.

Доказательство. 1. Очевидно: столбец координат вектора u_i в базисе $\{u_1, \dots, u_n\}$ равен e_i, то есть, равен i-му столбцу единичной матрицы.

2. Мы знаем, что
$$(w_1, \dots, w_n) = (u_1, \dots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}'').$$
С другой стороны, $(w_1, \dots, w_n) = (v_1, \dots, v_n)(\mathcal{B}' \rightsquigarrow \mathcal{B}'') = (u_1, \dots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}')(\mathcal{B}' \rightsquigarrow \mathcal{B}'')$. Поэтому
$$(u_1, \dots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}'') = (u_1, \dots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}')(\mathcal{B}' \rightsquigarrow \mathcal{B}'').$$
Поскольку $(u_1, \dots, u_n)$ является базисом, из равенства линейных комбинаций векторов $u_1, \dots, u_n$ следует равенство всех их коэффициентов, поэтому
$$(\mathcal{B} \rightsquigarrow \mathcal{B}'') = (\mathcal{B} \rightsquigarrow \mathcal{B}')(\mathcal{B}' \rightsquigarrow \mathcal{B}''),$$
что и требовалось.

3. Из первых двух пунктов следует, что $(\mathcal{B} \rightsquigarrow \mathcal{B}') \cdot (\mathcal{B}' \rightsquigarrow \mathcal{B}) = (\mathcal{B} \rightsquigarrow \mathcal{B}) = E$; аналогично, $(\mathcal{B}' \rightsquigarrow \mathcal{B}) \cdot (\mathcal{B} \rightsquigarrow \mathcal{B}') = (\mathcal{B}' \rightsquigarrow \mathcal{B}') = E$.

□

Теперь мы можем связать координаты одного и того же вектора в разных базисах.

Теорема 7.9.3. *Пусть* V — *конечномерное векторное пространство,* $\mathcal{B}$, $\mathcal{B}'$ — *базисы* V. *Тогда для любого вектора* $v \in V$ *выполнено*
$$[v]_{\mathcal{B}'} = (\mathcal{B}' \rightsquigarrow \mathcal{B}) \cdot [v]_{\mathcal{B}}.$$

Замечание 7.9.4. Это означает, что координаты вектора в базисе преобразуются *контравариантно* при замене базиса: координаты в новом базисе получается из координат в старом базисе домножением на матрицу перехода *из нового базиса в старый*.

Доказательство. Пусть $\mathcal{B} = \{u_1, \ldots, u_n\}$, $\mathcal{B}' = \{v_1, \ldots, v_n\}$. Запишем $[v]_{\mathcal{B}} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$ и $[v]_{\mathcal{B}'} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$. По определению это означает, что $v = u_1x_1 + \cdots + u_nx_n = v_1y_1 + \cdots + v_2y_2$, то есть,

$$v = \begin{pmatrix} u_1 & \ldots & u_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix} \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}.$$

По определению матрицы перехода имеем $\begin{pmatrix} v_1 & \ldots & v_n \end{pmatrix} = \begin{pmatrix} u_1 & \ldots & u_n \end{pmatrix} \cdot (\mathcal{B} \rightsquigarrow \mathcal{B}')$. Подставляя это в полученное равенство, получаем

$$v = \begin{pmatrix} u_1 & \ldots & u_n \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} == \begin{pmatrix} u_1 & \ldots & u_n \end{pmatrix} (\mathcal{B} \rightsquigarrow \mathcal{B}') \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

Но $(u_1, \ldots, u_n)$ является базисом, поэтому из равенства линейных комбинаций этих векторов следует равенство их коэффициентов. Значит,

$$\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = (\mathcal{B} \rightsquigarrow \mathcal{B}') \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix},$$

что и требовалось доказать. □

Еще один естественный вопрос — что происходит с матрицей отображения при замене базисов в пространствах? Пусть в пространстве U заданы базисы $\mathcal{B}$ и $\mathcal{C}$, а в пространстве V — базисы $\mathcal{B}'$ и $\mathcal{C}'$. У каждого линейного отображения $\varphi\colon U \to V$ имеется матрица $[\varphi]_{\mathcal{B},\mathcal{B}'}$ в базисах $\mathcal{B}, \mathcal{B}'$ и матрица $[\varphi]_{\mathcal{C},\mathcal{C}'}$ в базисах $\mathcal{C}, \mathcal{C}'$.

Теорема 7.9.5. *Пусть U, V — векторные пространства над полем k, $\varphi\colon U \to V$ — линейное отображение, $\mathcal{B}$, $\mathcal{C}$ — базисы в U, $\mathcal{B}'$, $\mathcal{C}'$ — базисы в V. Тогда*

$$[\varphi]_{\mathcal{C},\mathcal{C}'} = (\mathcal{B}' \rightsquigarrow \mathcal{C}')^{-1} [\varphi]_{\mathcal{B},\mathcal{B}'} (\mathcal{B} \rightsquigarrow \mathcal{C})$$

Доказательство. Пусть $u \in U$; тогда $[\varphi(u)]_{\mathcal{B}'} = [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}}$ и $[\varphi(u)]_{\mathcal{C}'} = [\varphi]_{\mathcal{C},\mathcal{C}'} \cdot [u]_{\mathcal{C}}$. Кроме того, $[u]_{\mathcal{B}} = (\mathcal{B} \rightsquigarrow \mathcal{C})[u]_{\mathcal{C}}$ и $[\varphi(u)]_{\mathcal{C}'} = (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi(u)]_{\mathcal{B}'}$. Поэтому

$$\begin{aligned} [\varphi]_{\mathcal{C},\mathcal{C}'} \cdot [u]_{\mathcal{C}} = [\varphi(u)]_{\mathcal{C}'} &= (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi(u)]_{\mathcal{B}'} \\ &= (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}} \\ &= (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi]_{\mathcal{B},\mathcal{B}'} \cdot (\mathcal{B} \rightsquigarrow \mathcal{C})[u]_{\mathcal{C}} \end{aligned}$$

для всех векторов $u \in U$. По предложению 5.3.8 из этого следует нужное равенство матриц. □

Итак, при замене базисов в пространствах U и V матрица отображения $\varphi\colon U \to V$ домножается справа на матрицу замены базиса в U и слева — на обратную матрицу замены базиса в V. Это можно использовать следующим образом: для фиксированного отображения φ попробуем подобрать базисы в U и V так, чтобы матрица φ в этих базисах выглядела наиболее простым образом.

Теорема 7.9.6 (Каноническая форма матрицы линейного отображения). *Пусть $\varphi\colon U \to V$ — гомоморфизм векторных пространства. Тогда найдутся базис $\mathcal{B}$ в U и базис $\mathcal{B}'$ в V такие, что матрица $[\varphi]_{\mathcal{B},\mathcal{B}'}$ является окаймленной единичной: $[\varphi]_{\mathcal{B},\mathcal{B}'} = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$. При этом $r = \dim(\operatorname{Im}(\varphi))$.*

Доказательство. По теореме о гомоморфизме (7.7.3) имеется изоморфизм $\widetilde{\varphi}\colon U/\operatorname{Ker}(\varphi) \cong \operatorname{Im}(\varphi)$. Выберем какой-нибудь базис в $\operatorname{Ker}(\varphi)$ и базис в $U/\operatorname{Ker}(\varphi)$; по теореме 7.8.2 мы получим базис в U; пусть это $e_1, \dots, e_n$, причем $e_1, \dots, e_r$ — относительный базис U над $\operatorname{Ker}(\varphi)$, а $e_{r+1}, \dots, e_n$ — базис $\operatorname{Ker}(\varphi)$. Базису $\overline{e_1}, \dots, \overline{e_r}$ в $U/\operatorname{Ker}(\varphi)$ в силу изоморфизма $\widetilde{\varphi}$ соответствует базис $f_1, \dots, f_r$ в $\operatorname{Im}(\varphi)$; при этом $\varphi(e_i) = f_i$ для $i = 1, \dots, r$, и видно, что $r = \dim(\operatorname{Im}(\varphi))$. Наконец, поскольку $\operatorname{Im}(\varphi) \leqslant V$, можно дополнить систему $f_1, \dots, f_r$ до базиса V векторами $f_{r+1}, \dots, f_m$. Поскольку $\varphi(e_i) = f_i$ для $i = 1, \dots, r$ и $\varphi(e_i) = 0$ для $i \geqslant r+1$, матрица φ в базисах $(e_1, \dots, e_n)$, $(f_1, \dots, f_m)$ имеет нужный вид. □

Фактически мы получили еще одно доказательство следствия 5.4.2.

Замечание 7.9.7. Размерность образа отображения φ называется **рангом** φ; по теореме 7.9.6 ранг линейного отображения равен рангу его матрицы (в любой паре базисов, поскольку при домножении на обратимые матрицы ранг не меняется).

Замечание 7.9.8. Приведем еще одну характеризацию ранга: *размерность образа линейного отображения равна рангу его матрицы.* Действительно, по теореме 7.9.6 можно выбрать базис так, что матрица нашего отображения станет окаймленной единичной. Для окаймленной единичной матрицы ранга r очевидно, что образ соответствующего линейного отображения имеет размерность r — этот образ есть просто линейная оболочка первых r базисных векторов. Осталось вспомнить, что при замене базиса происходит домножение матрицы линейного отображения на обратимые матрицы слева и справа, что, как мы знаем, не меняет ранга матрицы.

Предложение 7.9.9. *Размерность пространства решений однородной системы линейных уравнений равна числу неизвестных минус ранг матрицы этой системы.*

Доказательство. Пусть речь идет о системе $AX = 0$, где $A \in M(m, n, k)$, и $X \in k^n$ — столбец неизвестных. Рассмотрим линейный оператор $T\colon k^n \to k^m$, $X \mapsto AX$. Нетрудно понять, что его матрица относительно стандартных базисов k^n, k^m равна A. Пространство решений системы $AX = 0$ — это в точности ядро оператора T. Ранг матрицы A, как мы заметили выше — это размерность образа оператора T. Число неизвестных здесь равно n. Осталось применить теорему о гомоморфизме 7.3.8. □

Следствие 7.9.10. *Пусть* $A \in M(m, n, k)$. *Однородная линейная система уравнений* $AX = 0$ *имеет нетривиальное (то есть, ненулевое) решение тогда и только тогда, когда* $\mathrm{rk}(A) < n$. *В частности, если* $m < n$, *то эта система всегда имеет нетривиальное решение; если же* $m = n$, *то она имеет нетривиальное решение тогда и только тогда, когда матрица* A *необратима.*

Доказательство. Нетривиальное решение системы $AX = 0$ существует тогда и только тогда, когда размерность пространства решение строго больше 0, что по предыдущей теореме равносильно неравенству $\mathrm{rk}(A) < n$. Если $m < n$, то ранг матрицы A, будучи равен строчному рангу, не превосходит m: $\mathrm{rk}(A) \leqslant m < n$, поэтому нетривиальное решение имеется. Если же $m = n$, то неравенство $\mathrm{rk}(A) < n$ по следствию 7.6.8 равносильно необратимости A. □

Докажем еще раз теорему Кронекера–Капелли.

Теорема 7.9.11 (Кронекера–Капелли). *Система линейных уравнений* $AX = B$ *имеет решение тогда и только тогда, когда ранг матрицы* A *равен рангу расширенной матрицы* $(A|B)$. *При этом решение единственно тогда и только тогда, когда, дополнительно, этот ранг равен числу неизвестных* n.

Доказательство. Рассмотрим соответствующее линейное отображение $T\colon k^n \to k^m$, $X \mapsto AX$. Образ T — это подпространство, порожденное векторами $T(e_1), \ldots, T(e_n)$, то есть, пространство столбцов матрицы A. Значит, B лежит в $\mathrm{Im}(T)$ тогда и только тогда, когда столбец B является линейной комбинацией столбцов матрицы A. По предложению 5.9.6 имеется биекция между множеством решений системы $AX = B$ и множеством решений однородной системы $AX = 0$; это множество состоит из одной точки тогда и только тогда, когда $\mathrm{Ker}(T) = 0$, то есть, когда $\mathrm{rk}(A) = \dim(\mathrm{Im}(T)) = n$. □

8 Жорданова нормальная форма

Пусть U, V — конечномерные пространства над k. В прошлой главе мы выяснили, что для линейного отображения $T\colon U \to V$ можно выбрать базисы в U и в V так, что матрица φ в этих базисах будет окаймленной единичной. Пусть теперь $T\colon V \to V$ — линейное отображение из пространства в себя. Мы будем называть его **линейным оператором** (или просто **оператором**) на V. Не очень-то удобно выбирать два разных базиса в одном и том же пространстве V для записи матрицы линейного оператора. Пусть $\mathcal{B}$ — базис пространства V. **Матрицей оператора** $T\colon V \to V$ в базисе $\mathcal{B}$ называется матрица отображения T в базисах $\mathcal{B}$, $\mathcal{B}$. Мы будем обозначать ее через $[T]_{\mathcal{B}}$ вместо $[T]_{\mathcal{B},\mathcal{B}}$. Цель настоящей главы — выяснить, к какому наиболее простому виду можно привести матрицу оператора T с помощью выбора базиса в V. По теореме 7.9.5 при замене базиса $\mathcal{B}$ на $\mathcal{B}'$ матрица оператора T домножается справа на матрицу замены базиса и слева на обратную к ней. Таким образом, если $A = [T]_{\mathcal{B}}$, $A' = [T]_{\mathcal{B}'}$, C — матрица перехода от $\mathcal{B}$ к $\mathcal{B}'$, то $A' = C^{-1}AC$. Эта процедура называется **сопряжением**: говорят, что $C^{-1}AC$ — матрица, **сопряженная** к матрице A при помощи C.

В этой главе нас будет интересовать вопрос: к какому хорошему виду можно привести матрицу произвольного линейного оператора? В отличие от случая линейного отображения, рассчитывать на окаймленный единичный вид уже не приходится. Тем не менее, мы получим достаточно разумный ответ на этот вопрос. Можно сформулировать эту задачу на матричном языке: в прошлой главе мы видели, что с помощью домножения слева и справа на обратимые матрицы любую матрицу можно привести к окаймленной единичной форме; а к какому виду можно привести квадратную матрицу с помощью сопряжения?

Мы будем предполагать в этой главе, что все встречающиеся нам векторные пространства конечномерны.

8.1 Инвариантные подпространства и собственные числа

Литература: [F], гл. XII, § 6, п. 1; гл. IV, § 6, п. 1; [K2], гл. 2, § 3, п. 3; [KM], ч. 1, § 8; [vdW], гл. XII, § 88.

Первая идея для изучения операторов на пространстве состоит в следующем: можно попытаться посмотреть на то, что происходит в собственном подпространстве U оператора V, решить вопрос для него (что проще, поскольку размерность U меньше размерности V), а потом попробовать «подняться» в пространство V. Пусть $T\colon V \to V$ — линейный оператор, $U \leqslant V$ — некоторое подпространство. Проблема состоит в том, что ограничение $T|_U$ действует из U в V и уже не является линейным оператором! Опишем подпространства, для которых такого не происходит.

Определение 8.1.1. Пусть $T\colon V \to V$ — линейный оператор на пространстве V. Подпространство $U \leqslant V$ называется **инвариантным** относительно оператора T (или T-**инвариантным**), если $T(U) \subseteq U$. Иными словами: для любого $u \in U$ образ $T(u)$ также лежит в U.

Пример 8.1.2. Можно привести тривиальные примеры: подпространства $0 \leqslant V$ и $V \leqslant V$ инвариантны относительно любого линейного оператора на V.

Самый простой пример инвариантного подпространства возникают, когда это подпространство одномерно. Тогда U порождается одним ненулевым вектором $u \in U$, и для T-инвариантности U достаточно потребовать, чтобы образ $T(u)$ лежал в U, то есть, имел вид $u\lambda$ для некоторого $\lambda \in k$

Определение 8.1.3. Пусть $T\colon V \to V$ — линейный оператор. Скаляр $\lambda \in k$ называется **собственным числом** оператора T, если существует ненулевой вектор $u \in V$ такой, что $T(u) = u\lambda$. В этом случае u называется **собственным вектором** оператора T (соответствующим собственному числу λ).

Полезны следующие эквивалентные переформулировки понятия собственного числа.

Предложение 8.1.4. *Пусть $T\colon V \to V$ — линейный оператор, $\lambda \in k$. Следующие утверждения равносильны:*

1. *λ — собственное число оператора T;*
2. *оператор $T - \lambda\,\mathrm{id}_V$ неинъективен;*
3. *оператор $T - \lambda\,\mathrm{id}_V$ несюръективен;*
4. *оператор $T - \lambda\,\mathrm{id}_V$ необратим.*

Доказательство. Если λ — собственное число T, то $(T - \mathrm{id}_V\lambda)(u) = 0$ для некоторого ненулевого $u \in V$, и потому $T - \mathrm{id}_V\lambda$ неинъективен. Обратно, неинъективность $T - \mathrm{id}_V\lambda$ означает, что $\mathrm{Ker}(T - \mathrm{id}_V\lambda) \neq 0$, и если u — ненулевой вектор из этого ядра, то $T(u) = u\lambda$, что и означает, что λ — собственное число T. Равносильность утверждений (2), (3), (4) сразу следует из предложения 7.5.9. □

Таким образом, собственные числа оператора T — это в точности те скаляры λ, для которых оператор $T - \mathrm{id}_V\lambda$ имеет нетривиальное ядро, а соответствующие собственные векторы — это в точности ненулевые элементы этого ядра.

Теорема 8.1.5. *Пусть $T\colon V \to V$ — линейный оператор, $v_1, \ldots, v_n \in V$ — собственные векторы, соответствующие попарно различным собственным числам $\lambda_1, \ldots, \lambda_n \in k$. Тогда векторы $v_1, \ldots, v_n$ линейно независимы.*

Доказательство. Будем доказывать от противного: пусть $v_1, \ldots, v_n$ линейно зависиым. По лемме 6.3.12 найдется индекс j такой, что v_j выражается через $v_1, \ldots, v_{j-1}$. Выберем наименьший из таких индексов j и запишем полученную линейную зависимость:

$$v_j = v_1a_1 + \cdots + v_{j-1}a_{j-1}.$$

Применим оператор T к обеим частям этого равенства:

$$T(v_j) = T(v_1)a_1 + \cdots + T(v_{j-1})a_{j-1}.$$

Мы знаем, что $T(v_i) = v_i\lambda_i$ для всех $i = 1, \ldots, n$, потому

$$v_j\lambda_j = v_1\lambda_1 a_1 + \cdots + v_{j-1}\lambda_{j-1}a_{j-1}.$$

С другой стороны, мы можем умножить исходную линейную зависимость на λ_j:

$$v_j\lambda_j = v_1\lambda_j a_1 + \cdots + v_{j-1}\lambda_j a_{j-1}.$$

Вычтем два последних равенства:

$$0 = v_1(\lambda_1 - \lambda_j)a_1 + \cdots + v_{j-1}(\lambda_{j-1} - \lambda_j)a_{j-1}.$$

В силу нашего выбора j векторы $v_1, \ldots, v_{j-1}$ линейно независимы. Поэтому в полученном выражении все коэффициенты $(\lambda_i - \lambda_j)a_i$ должны быть нулевыми. Но скаляры λ_i попарно различны, потому $\lambda_j - \lambda_j \neq 0$ при всех $i = 1, \ldots, j-1$. Значит, $a_i = 0$ для $i = 1, \ldots, j-1$. Подставляя в исходную линейную комбинацию, получаем, что $v_j = 0$, что противоречит определению собственного вектора. □

Следствие 8.1.6. *Количество различных собственных чисел оператора на пространстве V не превосходит* $\dim(V)$.

Доказательство. Если нашлось больше, чем $\dim(V)$, различных собственных чисел, то соответствующие им собственные векторы линейно независимы по теореме 8.1.5, а это противоречит теореме 6.3.14. □

Возвращаясь к общему понятию инвариантного подпространства, мы теперь можем уточнить, в каком смысле наличие инвариантных подпространств помогает свести изучение оператора на пространстве к изучению операторов на меньших пространствах.

Определение 8.1.7. Пусть $T\colon V \to V$ — линейный оператор, $U \leqslant V$ — T-инвариантное подпространство. Отображение $T|_U\colon U \to U$, заданное формулой $(T|_U)(u) = T(u)$, называется **ограничением линейного оператора** на инвариантное подпространство U. Отображение $T_{V/U}\colon V/U \to V/U$, заданное формулой $T_{V/U}(v + U) = T(v) + U$, называется **индуцированным оператором** на факторпространстве V/U.

Предложение 8.1.8. *Ограничение на инвариантное подпространство и индуцированный оператор на фактор-пространстве корректно определены и являются линейными операторами.*

Доказательство. В силу инвариантности U элемент $T(u)$ лежит в U для всех $u \in U$, поэтому формула $(T|_U)(u) = T(u)$ задает отображение $T|_U\colon U \to U$. Его линейность очевидным образом следует из линейности T.

Для индуцированного отображения на фактор-пространстве сначала нужно проверить его корректность, то есть, то, что правило $v + U \mapsto T(v) + U$ не зависит от выбора представителей. Пусть v' — другой представитель класса $v + U$, то есть, $v' = v + u$ для некоторого $u \in U$.

Тогда $T(v') = T(v) + T(u)$. В силу T-инвариантности подпространства U вектор $T(u)$ лежит в U. Значит, $T(v')$ и $T(v)$ отличаются на элемент из U, а потому лежат в одном классе по модулю U.

После этого линейность отображения $T_{V/U}$ также напрямую следует из линейности оператора T. □

8.2 Собственные числа оператора над алгебраически замкнутым полем

Напомним, что линейные операторы на пространстве V образуют кольцо относительно сложения и композиции (а композицию мы часто записываем как умножение; в кольце матриц она буквально соответствует умножению матриц). Поэтому не очень удивительно, что мы можем рассматривать многочлены от оператора T на V. А именно, пусть $T\colon V \to V$ — линейный оператор на векторном пространстве V над k, и пусть $f \in k[x]$ — некоторый многочлен с коэффициентами в том же поле k. Запишем $f = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n$. Определим **результат подстановки оператора T в многочлен f** следующим образом:

$$f(T) = \mathrm{id}_V a_0 + Ta_1 + T^2a_2 + \cdots + T^na_n.$$

Здесь $T^n = \underbrace{T \circ \cdots \circ T}_{n}$ — результат n-кратной композиции T с собой. Нетрудно проверить, что это «возведение в степень» определено для всех натуральных n и обладает обычными свойствами, например, что $T^{m+n} = T^m \circ T^n$.

Итак, мы получили новый линейный оператор $f(T)$ по каждому многочлену $f \in k[x]$ и оператору T на V. Эта операция напоминает «подстановку скаляра в многочлен» (оно же «вычисление значение многочлена в точке», см. определение 4.3.1), и обладает похожими свойствами (см. предложение 4.3.3): если $f, g \in k[x]$, $\lambda \in k$, T — оператор на V, то $(f+g)(T) = f(T)+g(T)$, $(fg)(T) = f(T)g(T)$, $(f\lambda)(T) = f(T)\lambda$. Эти свойства проверяются простым раскрытием скобок. Действительно, пусть $f = a_0 + a_1x + \cdots + a_mx^m$, $g = b_0 + b_1x + \cdots + b_nx^n$. Тогда $fg = \sum_k \left(\sum_{i+j=k} a_ib_j\right) x^k$. Подставляя оператор T, получаем $f(T) = \mathrm{id}_V a_0 + Ta_1 + \cdots + T^ma_m$, $g(T) = \mathrm{id}_V b_0 + Tb_1 + \cdots + T^nb_n$, и потому $f(T)g(T) = \sum_k \left(\sum_{i+j=k} T^ia_iT^jb_j\right) = \sum_k T_i \left(\sum_{i+j=k} a_ib_j\right) = (fg)(T)$. Остальные свойства проверяются аналогично.

В частности, $f(T)g(T) = g(T)f(T)$: *многочлены от одного оператора коммутируют между собой* (обратите внимание, что композиция операторов, вообще говоря, некоммутативна: $ST \neq TS$).

Предложение 8.2.1. *Пусть поле k алгебраически замкнуто, $V \neq 0$ — векторное пространство над k, $T\colon V \to V$ — линейный оператор на V. Тогда у T есть собственное число.*

Доказательство. Выберем произвольный ненулевой вектор $v \in V$. Пусть $\dim V = n$. Рассмотрим векторы $v, T(v), T^2(v), \ldots, T^n(v)$. Это $n+1$ вектор в n-мерном векторном пространстве, и потому они линейно зависимы. По лемме 6.3.12 найдется индекс $j > 0$ такой, что $T^j(v)$ выражается через векторы вида $T^i(v)$ для $i < j$. Запишем это выражение: $va_0 + T(v)a_1 + \cdots + T^{j-1}(v)a_{j-1} = T^j(v)$. Перенесем все в одну часть и вынесем v:

$$(T^j - T^{j-1}a_{j-1} - \cdots - Ta_1 - \mathrm{id}_V a_0)(v) = 0.$$

В скобках стоит многочлен от оператора T, а именно, $f(T)$, где $f(x) = x^j - a_{j-1}x^{j-1} - \dots - a_1 x - a_0$. Наше поле алгебраически замкнуто, а степень f больше нуля, потому f раскладывается на линейные множители: $f(x) = (x-\lambda_1)\dots(x-\lambda_j)$, и, стало быть, $f(T) = (T - \mathrm{id}_V \lambda_1)\dots(T - \mathrm{id}_V \lambda_j)$.

Итак, мы получили, что $f(T)(v) = 0$, то есть, что $(T - \mathrm{id}_V \lambda_1)\dots(T - \mathrm{id}_V \lambda_j)(v) = 0$. Происходит следующее: на ненулевой вектор v действуют по очереди операторы вида $T - \mathrm{id}_V \lambda_i$, и получается 0. Из этого следует, что хотя бы один из них неинъективен — иначе из ненулевого вектора на каждом шаге получался бы ненулевой. Но неинъективность оператора $T - \mathrm{id}_V \lambda_i$ как раз и означает, что λ_i является собственным числом T (предложение 8.1.4). □

Итак, в случае алгебраически замкнутого поля, у каждого оператора T есть хотя бы одно собственное число λ, и, разумеется, есть соответствующий этому числу [ненулевой] собственный вектор v. Дополним этот вектор до некоторого базиса $\mathcal{B} = \{v, v_2, \dots, v_n\}$. Матрица оператора T в этом базисе выглядит следующим образом:

$$\begin{pmatrix} \lambda & * & \dots & * \\ 0 & *\dots & * & \\ \vdots & \vdots & \ddots & \vdots \\ 0 & * & \dots & * \end{pmatrix}.$$

Мы совершили небольшое продвижение к нашей цели: мы нашли базис, в котором матрица нашего оператора выглядит чуть-чуть лучше, чем наугад взятая матрица, а именно, в ней появилось несколько нулей. Оказывается, мы можем продолжить этот процесс по индукции, и найти базис, в котором матрица нашего оператора верхнетреугольна. Для этого нам понадобится следующее описание верхнетреугольных матриц.

Предложение 8.2.2. *Пусть $T\colon V \to V$ — линейный оператор, $\mathcal{B} = \{v_1, \dots, v_n\}$ — некоторый базис пространства V. Следующие утверждения равносильны:*

1. *матрица $[T]_{\mathcal{B}}$ верхнетреугольна;*
2. *для всех $j = 1, \dots, n$ вектор $T(v_j)$ лежит в $\langle v_1, \dots, v_j \rangle$;*
3. *для всех $j = 1, \dots, n$ подпространство $\langle v_1, \dots, v_j \rangle$ является T-инвариантным.*

Доказательство. Предположим, что матрица $[T]_{\mathcal{B}}$ верхнетреугольна. Посмотрим на ее j-й столбец: в нем стоит разложение вектора $T(v_j)$ по базису $\mathcal{B}$. То, что ниже диагонали там стоят нули, означает, что $T(v_j)$ на самом деле выражается только через векторы $v_1, \dots, v_j$. Обратно, если $T(v_j)$ выражается только через $v_1, \dots, v_j$, то в j-м столбце ниже диагонального элемента должны стоять нули. Поэтому первые два условия равносильны.

Очевидно, что из третьего условия следует второе. Осталось лишь показать, что из второго следует третье. Итак, пусть выполняется (2). Тогда

$$\begin{aligned} T(v_1) &\in \langle v_1 \rangle \subseteq \langle v_1, \dots, v_j \rangle, \\ T(v_2) &\in \langle v_1, v_2 \rangle \subseteq \langle v_1, \dots, v_j \rangle, \\ &\vdots \\ T(v_j) &\in \langle v_1, \dots, v_j \rangle. \end{aligned}$$

Если v — любая линейная комбинация векторов $v_1, \dots, v_j$, то $T(v)$ является линейной комбинацией векторов $T(v_1), \dots, T(v_j)$, и потому лежит в $\langle v_1, \dots, v_j \rangle$. Это означает, что подпространство $\langle v_1, \dots, v_j \rangle$ является T-инвариантным. □

Теорема 8.2.3. *Пусть* k *— алгебраически замкнутое поле,* $T\colon V \to V$ *— линейный оператор на конечномерном векторном пространстве* V *над полем* k*. Тогда существует базис* $v_1, \dots, v_n$ *пространства* V*, в котором матрица оператора* T *имеет верхнетреугольный вид.*

Доказательство. Пусть $\dim(V) = n$; будем доказывать теорему индукцией по n. Случай $n = 1$ очевиден; пусть теперь $n > 1$. По предложению 8.2.1 у T есть собственное число λ. Обозначим $U = \operatorname{Im}(T - \mathrm{id}_V \lambda) \leqslant V$. По предложению 8.1.4 оператор $T - \mathrm{id}_V \lambda$ не сюръективен, и потому $U \neq V$. Покажем, что подпространство U является T-инвариантным. Действительно, для любого $u \in U$ выполнено $T(u) = (T - \mathrm{id}_V \lambda)(u) + u\lambda$, и очевидно, что оба слагаемых лежат в U.

Теперь мы можем рассмотреть ограничение $T|_U$ оператора T на подпространство U. Мы знаем, что $\dim(U) < \dim(V)$, и потому к U можно применить предположение индукции и заключить, что существует базис $u_1, \dots, u_m$ пространства U, в котором матрица оператора $T|_U$ верхнетреугольна. По предложению 8.2.2 из этого следует, что $T(u_j) = (T|_U)(u_j) \in \langle u_1, \dots, u_j \rangle$ для всех $j = 1, \dots, m$.

Дополним $u_1, \dots, u_m$ до базиса $u_1, \dots, u_m, v_1, \dots, v_s$ пространства V. Тогда $T(v_k) = (T - \mathrm{id}_V \lambda)v_k + v_k\lambda$ для всех $k = 1, \dots, s$. По определению $(T - \mathrm{id}_V \lambda)v_k \in U$, и потому $T(v_k) \in \langle u_1, \dots, u_m, v_1, \dots, v_k \rangle$. По предложению 8.2.2 из этого следует, что матрица оператора T в базисе $u_1, \dots, u_m, v_1, \dots, v_s$ верхнетреугольна. □

Зная базис, в котором матрица оператора верхнетреугольна, легко определить, когда этот оператор обратим.

Предложение 8.2.4. *Пусть матрица оператора* $T\colon V \to V$ *в некотором базисе верхнетреугольна. Оператора* T *обратим тогда и только тогда, когда все диагональные элементы этой матрицы отличны от нуля.*

Доказательство. Пусть $\mathcal{B} = (v_1, \dots, v_n)$ — базис, в котором матрица оператора T верхнетреугольна, и пусть

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & * & \dots & * \\ 0 & \lambda_2 & \dots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Предположим, что оператор T обратим. Тогда $\lambda_1 \neq 0$ (иначе $T(v_1) = v_1\lambda_1 = 0$). Предположим, что $\lambda_j = 0$ для некоторого $j > 1$. Глядя на матрицу T, мы видим, что T отображает подпространство $\langle v_1, \dots, v_j \rangle$ в подпространство $\langle v_1, \dots, v_{j-1} \rangle$. При этом размерность первого подпространства равна j, а второго — $j - 1$. По следствию 7.3.10 не существует инъективных

линейных отображений из j-мерного пространства в $(j-1)$-мерное. Значит, ограничение оператора T на подпространство $\langle v_1, \ldots, v_j \rangle$ неинъективно. Это означает, что найдется ненулевой вектор $v \in \langle v_1, \ldots, v_j \rangle$, для которого $T(v) = 0$. Поэтому T неинъективен, что противоречит предположению об обратимости T.

Обратно, предположим теперь, что все $\lambda_1, \ldots, \lambda_n$ отличны от нуля. Глядя на первый столбец матрицы оператора T, мы видим, что $T(v_1) = v_1\lambda_1$, и потому $T(v_1\lambda_1^{-1}) = v_1$. Значит, $v_1 \in \operatorname{Im}(T)$. Далее, судя по второму столбцу матрицы оператора T, $T(v_2\lambda_2^{-1}) = v_1 a + v_2$ для некоторого $a \in k$. При этом $T(v_2\lambda_2^{-1})$ и $v_1 a$ лежат в $\operatorname{Im}(T)$. Поэтому и $v_2 \in \operatorname{Im}(T)$. Аналогично, $T(v_3\lambda_3^{-1}) = v_1 b + v_2 c + v_3$ для некоторых $b, c \in k$. Мы уже знаем, что все члены этого равенства, кроме v_3, лежат в $\operatorname{Im}(T)$, потому и $v_3 \in \operatorname{Im}(T)$.

Продолжая аналогичным образом, мы получаем, что $v_1, \ldots, v_n \in \operatorname{Im}(T)$. Тогда и $\langle v_1, \ldots, v_n \rangle \subseteq \operatorname{Im}(T)$. Но $v_1, \ldots, v_n$ — базис пространства V, и потому $\operatorname{Im}(T) = V$. Значит, оператор T сюръективен, что по предложению 7.5.9 влечет его обратимость. □

Теперь несложно показать, что если мы смогли привести матрицу оператора к верхнетреугольному виду, то на диагонали в точности стоят собственные числа этого оператора.

Предложение 8.2.5. *Пусть матрица оператора T относительно некоторого базиса верхнетреугольна. Тогда собственные числа оператора T — это в точности диагональные элементы этой матрицы.*

Доказательство. Пусть

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & * & \ldots & * \\ 0 & \lambda_2 & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

Для $\lambda \in k$ рассмотрим оператор $\lambda - \mathrm{id}_V \lambda$. Его матрица в том же базисе имеет вид

$$[T - \mathrm{id}_V \lambda]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 - \lambda & * & \ldots & * \\ 0 & \lambda_2 - \lambda & \ldots & * \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n - \lambda \end{pmatrix}.$$

По предложению 8.2.4 обратимость оператора $T - \mathrm{id}_V \lambda$ равносильна тому, что $\lambda_j - \lambda = 0$ для некоторого j, то есть, что λ стоит на диагонали. С другой стороны, по предложению 8.1.4 обратимость оператора $T - \mathrm{id}_V \lambda$ равносильна тому, что λ — собственное число оператора T. □

Определение 8.2.6. Пусть $T\colon V \to V$ — линейный оператор на векторном пространстве V, $\lambda \in k$. Подпространство $V_\lambda(T) = \operatorname{Ker}(T - \mathrm{id}_V \lambda)$ в V называется **собственным подпространством** оператора T, соответствующим числу λ. Часто, если понятно, о каком операторе идет речь, мы опускаем T в обозначении и пишем V_λ вместо $V_\lambda(T)$.

Нетрудно видеть, что V_λ — это в точности множество всех собственных векторов оператора T, соответствующих λ, вместе с 0. Скаляр λ является собственным числом оператора T тогда и только тогда, когда подпространство V_λ отлично от нулевого.

Предложение 8.2.7. *Пусть V — конечномерное пространство над полем k, $T\colon V \to V$ — линейный оператор. Пусть $\lambda_1, \ldots, \lambda_m$ — различные собственные числа оператора T. Тогда сумма $V_{\lambda_1} + \cdots + V_{\lambda_m}$ прямая. Кроме того, $\dim V_{\lambda_1} + \cdots + \dim V_{\lambda_m} \leqslant \dim V$.*

Доказательство. Пусть $u_1 + \cdots + u_m = 0$, где $u_j \in V_{\lambda_j}$ Из линейной независимости собственных векторов (теорема 8.1.5) следует, что $u_1 = \cdots = u_m = 0$. Поэтому сумма $V_{\lambda_1} + \cdots + V_{\lambda_m}$ прямая. Утверждение про размерность теперь напрямую следует из того, что размерность прямой суммы подпространств равна сумме их размерностей (следствие 6.5.6). □

8.3 Диагонализуемые операторы

Литература: [К2], гл. 2, § 3, п. 4; [КМ], ч. 1, § 8.

Определение 8.3.1. Оператор $T\colon V \to V$ называется **диагонализуемым**, если его матрица относительно некоторого базиса пространства V диагональна.

Диагонализуемые операторы составляют важный класс операторов, для которых задача приведения к «наиболее удобной форме» решается просто (нет ничего удобнее диагональной матрицы). Поэтому полезно уметь распознавать их.

Теорема 8.3.2. *Пусть V — конечномерное векторное пространство, $T\colon V \to V$ — линейный оператор. Пусть $\lambda_1, \ldots, \lambda_m$ — все различные собственные числа оператора T. Следующие условия эквивалентны:*

1. *оператор T диагонализуем;*
2. *у пространства V есть базис, состоящий из собственных векторов оператора T;*
3. *найдутся одномерные подпространства $U_1, \ldots, U_n$ в V, каждое из которых T-инвариантно, такие, что $V = U_1 \oplus \cdots \oplus U_n$;*
4. *$V = V_{\lambda_1}(T) \oplus \cdots \oplus V_{\lambda_m}(T)$;*
5. *$\dim V = \dim V_{\lambda_1}(T) + \cdots + \dim V_{\lambda_m}(T)$.*

Доказательство.

- $1 \Leftrightarrow 2$. Заметим, что матрица оператора T в базисе $v_1, \ldots v_n$ имеет вид

$$\begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

тогда и только тогда, когда $T(v_j) = v_j\lambda_j$ для всех $j = 1, \ldots, n$.

- $2 \Rightarrow 3$. Предположим, что $v_1, \ldots, v_n$ — базис V, и каждый вектор v_j — собственный вектор оператора T. Обозначим $U_j = \langle v_j \rangle$. Очевидно, что каждое подпространство U_j одномерно и T-инвариантно. Из определения базиса следует, что вектор из V можно единственным образом записать в виде линейной комбинации элементов $v_1, \ldots, v_n$. Иными словами любой вектор из V можно единственным образом представить в виде суммы $u_1 + \cdots + u_n$, где $u_j \in U_j$. Это и значит, что $V = U_1 \oplus \cdots \oplus U_n$.

- $3 \Rightarrow 2$. Пусть $V = U_1 \oplus \cdots \oplus U_n$ для некоторых одномерных T-инвариантных подпространств $U_1, \ldots, U_n$. Выберем в каждом U_j по ненулевому вектору v_j. Из T-инвариантности U_j следует, что v_j — собственный вектор оператора T. Каждый вектор из V можно единственным образом представить в виде суммы $u_1 + \cdots + u_n$, где $u_j \in U_j$, то есть, единственным образом представить в виде суммы кратных v_j. Поэтому $v_1, \ldots, v_n$ — базис V.

- $2 \Rightarrow 4$. Пусть у V есть базис, состоящий из собственных векторов. Тогда любой вектор V является линейной комбинацией собственных, и потому $V = V_{\lambda_1}(T) + \cdots + V_{\lambda_m}(T)$. Осталось применить предложение 8.2.7.

- $4 \Rightarrow 5$. Достаточно применить следствие 6.5.6.

- $5 \Rightarrow 2$. Выберем базис в каждом подпространстве $V_{\lambda_j}(T)$. Собрав эти базисы вместе, получим набор $v_1, \ldots, v_n$, состоящий из собственных векторов оператора T. По предположению их количество n равно $\dim V$. Покажем, что этот набор линейно независим. Предположим, что $v_1 a_1 + \cdots + v_n a_n = 0$ для некоторых $a_1, \ldots, a_n \in k$. Пусть u_j — сумма всех слагаемых вида $v_k a_k$, для которых $v_k \in V_{\lambda_j}$. Тогда каждый вектор u_j лежит в V_{λ_j}, и сумма $u_1 + \cdots + u_m = 0$. Из теоремы 8.1.5 следует, что все слагаемые этой суммы равны нулю. Но каждое слагаемое u_j является суммой элементов вида $v_k a_k$, где v_k образуют базис пространства V_{λ_j}. Поэтому все коэффициенты a_k равны нулю. Мы получили, что набор $v_1, \ldots, v_n$ линейно независим. Его можно дополнить до базиса, но, с другой стороны, количество векторов в этом наборе уже равно размерности пространства V. Поэтому $v_1, \ldots, v_n$ — базис V.

□

Пример 8.3.3. Пусть оператор T на двумерном пространстве k^2 задан формулой $v \mapsto A \cdot v$, где

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

Иными словами, A — матрица оператора T в стандартном базисе пространства k^2. Матрица A верхнетреугольна, поэтому собственные числа оператора T — это ее диагональные элементы. Таким образом, у T есть ровно одно собственное число: 0. Несложное вычисление показывает, что все собственные векторы имеют вид $\begin{pmatrix} * \\ 0 \end{pmatrix}$. Поэтому у k^2 нет базиса, состоящего из собственных векторов, а значит, оператор T не диагонализуем.

Таким образом, не любой оператор можно привести к диагональному виду. Но, во всяком случае, это возможно, если у оператора достаточно много различных собственных чисел.

Следствие 8.3.4. *Пусть* $T\colon V \to V$ *— линейный оператор на* n*-мерном векторном пространстве* V*. Предположим, что у* T *есть* n *различных собственных чисел. Тогда оператор* T *диагонализуем.*

Доказательство. У оператора T есть n собственных векторов $v_1, \ldots, v_n$, соответствующих различным собственным числам. По теореме 8.1.5 они линейно независимы. Но их количество равно размерности пространства V, и потому они образуют базис V. По теореме 8.3.2 из этого следует, что T диагонализуем. □

8.4 Корневое разложение

Литература: [F], гл. XII, § 6, п. 2; [K2], гл. 2, § 4, п. 3; [KM], ч. 1, § 9.

Для нахождения правильного базиса в пространстве V нам понадобится некоторое расширение понятия собственного вектора. Напомним, что собственные векторы — это в точности ненулевые элементы $\operatorname{Ker}(T - \mathrm{id}_V \lambda)$. Посмотрим теперь на $\operatorname{Ker}(T - \mathrm{id}_V \lambda)^j$ при различных $j = 1, 2, \ldots$.

Лемма 8.4.1. *Для любого оператора* $T\colon V \to V$ *имеется возрастающая цепочка вложенных подпространств*

$$0 = \operatorname{Ker}(T^0) \leqslant \operatorname{Ker}(T) \leqslant \operatorname{Ker}(T^2) \leqslant \operatorname{Ker}(T^3) \leqslant \ldots.$$

Более того, если $\operatorname{Ker}(T^j) = \operatorname{Ker}(T^{j+1})$ *для некоторого натурального* j*, то* $\operatorname{Ker}(T^{j+m}) = \operatorname{Ker}(T^{j+m+1})$ *для всех* $m \geqslant 0$*.*

Доказательство. Пусть $v \in \operatorname{Ker}(T^i)$. Это значит, что $T^i(v) = 0$. Но тогда и $T^{i+1}(v) = T(T^i(v)) = T(0) = 0$. Мы показали, что $\operatorname{Ker}(T^i) \subseteq \operatorname{Ker}(T^{i+1})$. Докажем второе утверждение индукцией по m. База $m = 0$ очевидна. Пусть теперь $m > 0$. Мы уже знаем, что $\operatorname{Ker}(T^{j+m}) \subseteq \operatorname{Ker}(T^{j+m+1})$; осталось доказать обратное включение. Пусть $v \in \operatorname{Ker}(T^{j+m+1})$. Это означает, что $T^{j+m+1}(v) = 0$. Но $T^{j+m+1}(v) = T^{j+1}(T^m(v)) = 0$. Поэтому $T^m(v) \in \operatorname{Ker}(T^{j+1}) = \operatorname{Ker}(T^j)$, и тогда $0 = T^j(T^m(v)) = T^{j+m}(v)$, что и требовалось. □

Итак, мы построили бесконечную цепочку возрастающих подпространств и показали, что если два элемента в ней совпали, то начиная с этого места цепочка «стабилизируется». В конечномерном пространстве V, разумеется, невозможна бесконечная цепочка *строго* возрастающих подпространств.

Предложение 8.4.2. *Пусть* $T\colon V \to V$ *— линейный оператор на конечномерном пространстве* V*, и* $\dim(V) = n$*. Тогда* $\operatorname{Ker}(T^n) = \operatorname{Ker}(T^{n+1}) = \cdots = \operatorname{Ker}(T^{n+j}) = \ldots$.

Доказательство. Предположим, что $\operatorname{Ker}(T^n) \neq \operatorname{Ker}(T^{n+1})$. Посмотрим на включение $\operatorname{Ker}(T^0) \leqslant \operatorname{Ker}(T)$. Если в нем имеет место равенство, то (по лемме 8.4.1) и $\operatorname{Ker}(T^n) = \operatorname{Ker}(T^{n+1})$. Значит, $\operatorname{Ker}(T^0) \neq \operatorname{Ker}(T)$. Аналогично,

$$\operatorname{Ker}(T) \neq \operatorname{Ker}(T^2) \neq \operatorname{Ker}(T^3) \neq \cdots \neq \operatorname{Ker}(T^n) \neq \operatorname{Ker}(T^N n + 1).$$

Но тогда $\dim(\mathrm{Ker}(T)) \geqslant 1$, $\dim(\mathrm{Ker}(T^2)) \geqslant 2$, $\ldots$, $\dim(\mathrm{Ker}(T^{n+1})) \geqslant n+1$. Но $\mathrm{Ker}(T^{n+1})$ — подпространство в V, и не может иметь размерность, большую n. Получили противоречие. Мы показали, что $\mathrm{Ker}(T^n) = \mathrm{Ker}(T^{n+1})$, а по лемме 8.4.1 из этого следует и всех следующих подпространств в нашей цепочке. □

Следующее предложение оказывается ключом к разложению пространства в прямую сумму подпространств, на каждом из которых ситуацию проще исследовать.

Предложение 8.4.3. *Пусть* $T\colon V \to V$ *— линейный оператор на пространстве размерности* n*. Тогда* $V = \mathrm{Ker}(T^n) \oplus \mathrm{Im}(T^n)$*.*

Доказательство. Покажем сначала, что $\mathrm{Ker}(T^n) \cap \mathrm{Im}(T^n) = 0$. Действительно, пусть $v \in \mathrm{Ker}(T^n) \cap \mathrm{Im}(T^n)$. Тогда $v = T^n(u)$; с другой стороны, $T^n(v) = T^n(T^n(u)) = 0$. Поэтому $u \in \mathrm{Ker}(T^{2n}) = \mathrm{Ker}(T^n)$ (по предложению 8.4.2), откуда $v = T^n(u) = 0$.

Мы показали, что сумма $\mathrm{Ker}(T^n)+\mathrm{Im}(T^n) \leqslant V$ прямая. По следствию 6.5.6 тогда $\dim(\mathrm{Ker}(T^n)+\mathrm{Im}(T^n)) = \dim \mathrm{Ker}(T^n) + \dim \mathrm{Im}(T^n)$. По теореме о гомоморфизме 7.3.8 эта сумма размерностей равна $\dim V$, и потому $\mathrm{Ker}(T^n) \oplus \mathrm{Im}(T^n) = V$. □

Выше мы разобрались с диагональными операторами за счет того, что для них имеет место разложение в прямую сумму инвариантных T-подпространств вида $V = V_{\lambda_1} \oplus \cdots \oplus V_{\lambda_m}$, где $\lambda_1, \ldots, \lambda_m$ — все различные собственные числа оператора T. Сейчас мы покажем, что для произвольного оператора имеет место аналогичное разложение, если собственные подпространства заменить на чуть большие *корневые*.

Определение 8.4.4. Пусть $T\colon V \to V$ — линейный оператор, и $\lambda \in k$ — его собственное число. Ненулевой вектор $v \in V$ называется **корневым вектором** оператора T, соответствующим собственному числу λ, если $(T - \mathrm{id}_V \lambda)^j(v) = 0$ для некоторого натурального j.

Замечание 8.4.5. Предположим, что $(T-\mathrm{id}_V \lambda)^j(v) = 0$ для некоторого j. По предложению 8.4.2 тогда и $(T - \mathrm{id}_V \lambda)^n(v) = 0$, где $n = \dim(V)$. Поэтому корневые векторы — это на самом деле в точности ненулевые элементы $\mathrm{Ker}(T - \mathrm{id}_V \lambda)^n$.

Определение 8.4.6. Множество всех корневых векторов оператора T, соответствующих собственному числу λ, вместе с нулем, называется **корневым подпространством** и обозначается через $V(\lambda, T)$. Зачастую из контекста понятно, о каком операторе идет речь, и мы пишем $V(\lambda)$ вместо $V(\lambda, T)$. По замечанию 8.4.5 это действительно подпространство: $V(\lambda, T) = \mathrm{Ker}(T - \mathrm{id}_V \lambda)^n$, где $n = \dim(V)$.

Теорема 8.4.7. *Пусть* $T\colon V \to V$ *— линейный оператор,* $\lambda_1, \ldots, \lambda_m$ *— его попарно различные собственные числа,* $v_1, \ldots, v_m$ *— соответствующие им корневые векторы. Тогда* $v_1, \ldots, v_m$ *линейно независимы.*

Доказательство. Предположим, что $v_1, \ldots, v_m$ линейно зависимы. По лемме 6.3.12 найдется индекс j такой, что $v_j = v_1 a_1 + \cdots + v_{j-1} a_{j-1}$ для некоторых $a_1, \ldots, a_{j-1} \in k$. Выберем наименьшее такое j. Вектор v_j является корневым, соответствующим собственному числу λ_j. Возьмем

наименьшую степень d оператора $(T - \mathrm{id}_V \lambda_j)$, которая не переводит этот вектор в 0. Иными словами, пусть $(T - \mathrm{id}_V \lambda_j)^d(v_j) \neq 0$ и $(T - \mathrm{id}_V \lambda_j)^{d+1}(v_j) = 0$. Обозначим $(T - \mathrm{id}_V \lambda_j)^d(v_j) = w$. Тогда $(T - \mathrm{id}_V \lambda_j)(w) = 0$, и поэтому $Tw = w\lambda_j$. Более того, $(T - \mathrm{id}_V \lambda)(w) = T(w) - w\lambda = w(\lambda_j - \lambda)$ для всех $\lambda \in k$. Поэтому $(T - \mathrm{id}_V \lambda)^k(w) = w(\lambda_i - \lambda)^k$ для всех натуральных k.

Пусть $\dim V = n$. Применим к нашей линейной зависимости оператор $(T - \mathrm{id}_V \lambda_1)^n \dots (T - \mathrm{id}_V \lambda_{j-1})^n (T - \mathrm{id}_V \lambda_j)^d$. В левой части получим

$$(T - \mathrm{id}_V \lambda_1)^n \dots (T - \mathrm{id}_V \lambda_{j-1})^n (T - \mathrm{id}_V \lambda_j)^d (v_j).$$

Сначала к вектору v_j применяется оператор $(T - \mathrm{id}_V \lambda_j)^d$, и получается вектор w, а потом применяются по очереди операторы вида $(T - \mathrm{id}_V \lambda_i)^n$ для $i \neq j$. Но выше мы выяснили, как они действуют: такой оператор просто умножает w на $(\lambda_j - \lambda_i)^n$. Поэтому результат равен $(\lambda_j - \lambda_1)^n \dots (\lambda_j - \lambda_{j-1})^n w$ и отличен от нуля.

В правой же части происходит следующее: при вычислении действия оператора $(T - \mathrm{id}_V \lambda_1)^n \dots (T - \mathrm{id}_V \lambda_{j-1})^n (T - \mathrm{id}_V \lambda_j)^d$ на v_i (где $1 \leqslant i \leqslant j - 1$) можно переставить скобки так, чтобы сначала действовала скобка $(T - \mathrm{id}_V \lambda_i)^n$. Но $(T - \mathrm{id}_V \lambda_i)^n(v_i) = 0$ по определению корневого вектора. Поэтому каждое слагаемое в правой части равно нулю. Мы получили, что ненулевой вектор равен нулевому; это противоречие, которое завершает доказательство. □

Лемма 8.4.8. *Пусть* $T\colon V \to V$ *— линейный оператор,* $p \in k[x]$ *— многочлен. Тогда подпространства* $\mathrm{Ker}(p(T))$ *и* $\mathrm{Im}(p(T))$ T*-инвариантны.*

Доказательство. Пусть $v \in \mathrm{Ker}(p(T))$, то есть, $p(T)(v) = 0$. Тогда

$$p(T)(T(v)) = (p(T) \cdot T)(v) = (T \cdot p(T))(v) = T(p(T)(v)) = T(0) = 0.$$

Мы получили, что $T(v) \in \mathrm{Ker}(p(T))$, и потому $\mathrm{Ker}(p(T))$ действительно T-инвариантно.

Пусть теперь $v \in \mathrm{Im}(p(T))$, то есть, $v = p(T)(u)$ для некоторого $u \in V$. Тогда $T(v) = T(p(T)(u)) = p(T)(T(u)) \in \mathrm{Im}(p(T))$, что и требовалось. □

Теперь мы готовы показать, что пространство раскладывается в прямую сумму корневых. Для этого нам понадобится следующее определение.

Определение 8.4.9. Линейный оператор $T\colon V \to V$ называется **нильпотентным**, если $T^j = 0$ для некоторого натурального j.

Теорема 8.4.10. *Пусть* $T\colon V \to V$ *— линейный оператор на конечномерном пространстве* V *над алгебраически замкнутым полем* k, $\lambda_1, \dots, \lambda_m$ *— все его (попарно различные) собственные числа. Тогда*

1. $V = V(\lambda_1, T) \oplus \dots \oplus V(\lambda_m, T)$;
2. *каждое из подпространств* $V(\lambda_j, T)$ *является* T*-инвариантным;*
3. *оператор* $(T - \mathrm{id}_V \lambda_j)|_{V(\lambda_j, T)}$ *на корневом подпространстве* $V(\lambda_j, T)$ *нильпотентен.*

Доказательство. Пусть $\dim(V) = n$. Заметим сначала, что $V(\lambda_j, T) = \operatorname{Ker}(T - \mathrm{id}_V \lambda_j)^n$, и его T-инвариантность следует из леммы 8.4.8, примененной к многочлену $p(x) = (x - \lambda_j)^n$.

Далее, если $v \in V(\lambda_j, T)$, то $(T - \mathrm{id}_V \lambda_j)^n(v) = 0$. Поэтому оператор $(T - \mathrm{id}_V \lambda_j)^n$ тождественно равен 0 на подпространстве $V(\lambda_j, T)$, откуда следует нильпотентность оператора $(T - \mathrm{id}_V \lambda_j)|_{V(\lambda_j, T)}$.

Осталось показать, что V раскладывается в прямую сумму корневых. Будем доказывать это индукцией по n. Случай $n = 1$ очевиден. Пусть теперь $n > 1$, и нужный результат верен для всех пространств меньшей размерности. По предложению 8.2.1 у T есть собственное число; поэтому $m \geqslant 1$. По лемме 8.4.3 тогда $V = \operatorname{Ker}(T - \mathrm{id}_V \lambda_1)^n \oplus \operatorname{Im}(T - \mathrm{id}_V \lambda_1)^n$. Первое подпространство в прямой сумме — это в точности $V(\lambda_1, T)$, а второе давайте обозначим через U. Пространство $V(\lambda_1, T)$ нетривиально, и потому размерность U строго меньше размерности V. Кроме того, подпространство U является T-инвариантным по лемме 8.4.8. Значит, к оператору $T|_U$, действующему на пространстве U, можно применить предположение индукции, и получить, что

$$U = V(\mu_1, T|_U) \oplus \cdots \oplus V(\mu_k, T|_U),$$

где $\mu_1, \ldots, \mu_k$ — собственные числа оператора $T|_U$. Покажем, что любое собственное число λ оператора $T|_U$ является и собственным числом оператора T. Действительно, если $T|_U(u) = u\lambda$ для некоторого ненулевого вектора $u \in U$, то и $T(u) = u\lambda$. Заметим также, что у оператора $T|_U$ не может быть собственного числа λ_1: если $T|_U(u) = u\lambda_1$ то $T(u) = u\lambda_1$, и потому $u \in \operatorname{Ker}(T - \mathrm{id}_V \lambda_1)^n$, и из разложения в прямую сумму $V = \operatorname{Ker}(T - \mathrm{id}_V \lambda_1)^n \oplus U$ следует, что $u = 0$.

Мы получили, что $\mu_1, \ldots, \mu_k$ — это какие-то из чисел $\lambda_2, \ldots, \lambda_m$. Возьмем какое-нибудь одно из $\mu_1, \ldots, \mu_k$; пусть это λ_j. Несложно понять, что $V(\lambda_j, T|_U) \leqslant V(\lambda_j, T)$: действительно, если $u \in U$ — корневой вектор для собственного числа λ_j оператора $T|_U$, то тем более u является корневым вектором для собственного числа λ_j оператора T.

Вернемся к общей картине. По теореме 8.4.7 сумма корневых подпространств прямая; получаем, что $V(\lambda_1, T) \oplus \ldots V(\lambda_m, T) \leqslant V$. С другой стороны, мы показали, что $V = V(\lambda_1, T) \oplus U$, и U раскладывается в прямую сумму слагаемых, каждое из которых содержится в каком-то $V(\lambda_j, T)$. Поэтому

$$\begin{aligned} V &= V(\lambda_1, T) \oplus U \\ &= V(\lambda_1, T) \oplus V(\mu_1, T|_U) \oplus \cdots \oplus V(\mu_k, T|_U) \\ &\leqslant V(\lambda_1, T) \oplus V(\lambda_2, T) \oplus \cdots \oplus V(\lambda_m, T), \end{aligned}$$

и мы получили включение в обратную сторону. □

Следствие 8.4.11. *Пусть $T\colon V \to V$ — линейный оператор на конечномерном пространстве V над алгебраически замкнуты м полем k. Тогда у пространства V есть базис, состоящий из корневых векторов оператора T.*

Доказательство. Выберем базисы в каждом из подпространств вида $V(\lambda_j, T)$ и объединим их. □

8.5 Характеристический и минимальный многочлены

Определение 8.5.1. Пусть V — векторное пространство над алгебраически замкнутым полем k, $T\colon V \to V$ — линейный оператор, $\lambda \in k$ — его собственное число. Размерность соответствующего корневого подпространства $V(\lambda, T)$ называется **кратностью собственного числа** λ. Иными словами, кратность собственного числа λ оператора T равна $\dim(\operatorname{Ker}(T - \mathrm{id}_V \lambda)^{\dim(V)})$.

Замечание 8.5.2. Иногда то, что мы называем кратностью, в литературе называется *алгебраической кратностью*, в то время как размерность собственного подпространства $V_\lambda(T)$ называется *геометрической кратностью* λ. После этого доказывается теорема о том, что геометрическая кратность не превосходит алгебраической кратности, которая при наших определениях очевидна (собственное подпространство содержится в корневом).

Следствие 8.5.3. *Сумма кратностей всех собственных чисел оператора* $T\colon V \to V$ *равна* $\dim(V)$.

Доказательство. Тривиально следует из теоремы 8.4.10 и следствия 6.5.6. □

Определение 8.5.4. Пусть V — векторное пространство над алгебраически замкнутым полем k, $T\colon V \to V$ — линейный оператор. Пусть $\lambda_1, \ldots, \lambda_m$ — все его [попарно различные] собственные числа, а $d_1, \ldots, d_m$ — их кратности, соответственно. Многочлен $(x - \lambda_1)^{d_1} \ldots (x - \lambda_s)^{d_m}$ называется **характеристическим многочленом** оператора T.

Предложение 8.5.5. *Степень характеристического многочлена оператора* $T\colon V \to V$ *равна* $\dim(V)$*, а его корни — в точности собственные числа оператора* T.

Доказательство. Очевидно из определения и следствия 8.5.3. □

Теорема 8.5.6 (Гамильтона–Кэли). *Пусть* V *— векторное пространство над алгебраически замкнутым полем* k, $T\colon V \to V$ *— линейный оператор,* $q \in k[x]$ *— его характеристический многочлен. Тогда* $q(T) = 0$.

Доказательство. Пусть $\lambda_1, \ldots, \lambda_m$ — все собственные числа оператора T, а $d_1, \ldots, d_m$ — их кратности. По теореме 8.4.10 ограничения вида $(T - \mathrm{id}_V \lambda_j)|_{V(\lambda_j, T)}$ нильпотентны, а по предложению 8.4.2 тогда $(T - \mathrm{id}_V \lambda_j)^{d_j}|_{V(\lambda_j, T)} = 0$.

Любой вектор из V является суммой векторов из $V(\lambda_1, T), \ldots, V(\lambda_m, T)$ (по теореме 8.4.10), поэтому достаточно доказать, что $q(T)(v_j) = 0$ для любого $v_j \in V(\lambda_j, T)$. По определению

$$q(T) = (T - \mathrm{id}_V \lambda_1)^{d_1} \ldots (T - \mathrm{id}_V \lambda_m)^{d_m}.$$

Операторы в правой части являются многочленами от оператора T, и потому коммутируют друг с другом. Переставим их так, чтобы множитель $(T - \mathrm{id}_V \lambda_j)^{d_j}$ оказался последним. Но $(T - \mathrm{id}_V \lambda_j)^{d_j}(v_j) = 0$, и потому $q(T)(v_j) = 0$, что и требовалось. □

Определение 8.5.7. Пусть $T\colon V \to V$ — линейный оператор на векторном пространстве V. Многочлен $p \in k[x]$ минимальной степени со старшим коэффициентом 1, для которого $p(T) = 0$, называется **минимальным многочленом** оператора T. Иными словами, многочлен $p \in k[x]$ со старшим коэффициентом 1 называется минимальным многочленом оператора T, если

- $p(T) = 0$;
- если $f \in k[x]$ — многочлен со старшим коэффициентом 1, для которого $f(T) = 0$, то $\deg f \geqslant \deg p$.

Покажем, что это определение осмысленно: у каждого оператора T (на конечномерном пространстве V) существует единственный минимальный многочлен. Пусть $\dim(V) = n$. Рассмотрим множество операторов $\mathrm{id}_V, T, T^2, \ldots, T^{n^2}$. В нем $n^2 + 1$ элемент, в то время как размерность пространства всех линейных операторов на V равна n^2 (по теореме 7.5.6). Значит, указанный набор операторов линейно зависим. Выберем минимальное m, для которого операторы $\mathrm{id}_V, T, T^2, \ldots, T^m$ линейно зависимы. Тогда T^m выражается через $\mathrm{id}_V, T, T^2, \ldots, T^{m-1}$:

$$T^m = \mathrm{id}_V\, a_0 + Ta_1 + \cdots + T^{m-1}a_{m-1}$$

для некоторых $a_0, \ldots, a_{m-1} \in k$. Пусть $p \in k[x]$ — следующий многочлен:

$$p = x^m - a_{m-1}x^{m-1} - \cdots - a_1x - a_0.$$

Тогда $p(T) = 0$. Предположим, что f — еще один многочлен той же степени m со старшим коэффициентом 1, для которого $f(T) = 0$. Тогда многочлен $f - p$ имеет меньшую степень, но $(f - p)(T) = f(T) - p(T) = 0$, что противоречит выбору m.

Следующее предложение полностью описывает многочлены $f \in k[x]$, для которых $f(T) = 0$.

Предложение 8.5.8. *Пусть $T\colon V \to V$ — линейный оператор, $f \in k[x]$ — некоторый многочлен. Равенство $f(T) = 0$ равносильно тому, что f делится на минимальный многочлен оператора T.*

Доказательство. Пусть p — минимальный многочлен оператора T. Если f делится на p, то есть, $f = pq$ для некоторого многочлена $q \in k[x]$, то $f(T) = p(T)q(T) = 0$. Обратно, если $f(T) = 0$, поделим с остатком f на p: $f = pq + r$ для $q, r \in k[x]$, причем $\deg(r) < \deg(p)$. Но $r(T) = f(T) - p(T)q(T) = 0$, что противоречит минимальности многочлена p. □

Следствие 8.5.9. *Пусть V — векторное пространство над алгебраически замкнутым полем k, $T\colon V \to V$ — линейный оператор. Тогда характеристический многочлен оператора T делится на его минимальный многочлен.*

Доказательство. Немедленно следует из теоремы Гамильтона–Кэли 8.5.6 и предложения 8.5.8. □

Предложение 8.5.10. *Пусть* T *— линейный оператор на* V*. Корни минимальноого многочлена оператора* T *— это в точности все собственные числа этого оператора.*

Доказательство. Пусть p — минимальный многочлен оператора T. Если $\lambda \in k$ — корень p, то $p(x) = (x - \lambda)q$ для некоторого многочлена $q \in k[x]$ со старшим коэффициентом 1. Из равенства $p(T)$ следует, что $(T - \mathrm{id}_V \lambda)(q(T)(v)) = 0$ для всех $v \in V$. Заметим, что степень q меньше степени минимального многочлена оператора T, и потому $q(T) \neq 0$. Поэтому найдется вектор $v \in V$, для которого $q(T)(v) \neq 0$. Но тогда равенство $(T - \mathrm{id}_V \lambda)(q(T)(v)) = 0$ означает, что λ — собственное число оператора T (а $q(T)(v)$ — соответствующий ему собственный вектор).

Обратно, пусть $\lambda \in k$ — собственное число оператора T. Тогда найдется ненулевой вектор $v \neq 0$, для которого $T(v) = \lambda v$. Применяя несколько раз T к обеим частям этого равенства, получаем, что $T^j(v) = \lambda^j v$ для всех $j \geqslant 0$. Поэтому $p(T)(v) = p(\lambda)(v)$; с другой стороны, $p(T)(v) = 0$. При этом вектор v отличен от нуля, значит, $p(\lambda) = 0$. □

8.6 Жорданов базис для нильпотентного оператора

Литература: [F], гл. XII, § 6, пп. 2–4; [K2], гл. 2, § 4, пп. 4–6; [KM], ч. 1, § 9; [vdW], гл. XII, §§ 88, 89.

Напомним, что по теореме 8.4.10 изучение оператора T сводится к изучению нильпотентных операторов. Теперь мы готовы построить хороший базис для нильпотентного оператора.

Теорема 8.6.1. *Пусть* V *— векторное пространство над полем* k*,* $N\colon V \to V$ *— нильпотентный оператора. Тогда найдутся векторы* $v_1, \ldots, v_s \in V$ *и натуральные числа* $m_1, \ldots, m_s$ *такие, что*

- *векторы*

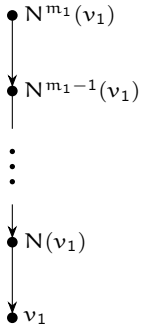
$$\begin{aligned} &N^{m_1}(v_1), \ldots, N(v_1), v_1,\\ &N^{m_2}(v_2), \ldots, N(v_2), v_2,\\ &\ldots\\ &N^{m_s}(v_s), \ldots, N(v_s), v_s \end{aligned}$$

  образуют базис V*;*

- $N^{m_1+1}(v_1) = \cdots = N^{m_s+1}(v_s) = 0$.

Замечание 8.6.2. Полученный базис удобно схематично изображать в виде ориентированного графа, вершины которого символизируют векторы базиса, а ребра выражают действие оператора N. Набор $N^{m_1}(v_1), \ldots, N(v_1), v_1$ тогда представляется в виде цепочки из $m_1 + 1$

вершины:



Очевидно, что подпространство, порожденное векторами из одной такой цепочки, N-инвариантно. Матрица ограничения оператора N на это подпространство (в этом базисе) имеет размер $(m_1 + 1) \times (m_1 + 1)$ и выглядит так:

$$\begin{pmatrix} 0 & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 \end{pmatrix}$$

Базис, о котором идет речь в теореме — набор из s таких цепочек (возможно, разной длины). Матрица оператора N в таком базисе, стало быть, имеет блочно-диагональный вид, и на диагонали стоят блоки указанного вида.

Доказательство теоремы 8.6.1. Будем доказывать теорему индукцией по размерности пространства V. Случай $\dim(V) = 1$ тривиален: нильпотентный оператор на одномерном пространстве должен быть нулевым, и мы можем положить $s = 1$, выбрать любой ненулевой вектор $v_1 \in V$ и $m_1 = 0$.

Пусть теперь $\dim(V) > 1$. Рассмотрим подпространство $\mathrm{Im}(N) \leqslant V$. Если оно совпадает с V, то оператор N обратим, что противоречит его нильпотентности. Поэтому $\mathrm{Im}(N)$ — подпространство в V меньшей размерности. Если случилось так, что $\mathrm{Im}(N)$ — нулевое пространство, то оператор N нулевой, и потому можно выбрать произвольный базис $v_1, \dots, v_s$ пространства V и положить $m_1 = \dots = m_s = 0$; на этом доказательство заканчивается.

Если же $\mathrm{Im}(N) \neq 0$, то к нему можно применить предположение индукции. Значит, мы можем выбрать векторы $v_1, \dots, v_s \in \mathrm{Im}(N)$ и натуральные числа $m_1, \dots, m_s$ так, что заключение теоремы выполнено (для подпространства $\mathrm{Im}(N)$). Для каждого вектора $v_i \in \mathrm{Im}(N)$ можно выбрать $u_i \in V$ так, что $v_i = N(u_i)$. Переписав заключение теоремы в терминах

векторов u_i, получаем, что набор

$$\begin{aligned}
&N^{m_1+1}(u_1), \ldots, N^2(u_1), N(u_1),\\
&N^{m_2+1}(u_2), \ldots, N^2(u_2), N(u_2),\\
&\ldots\\
&N^{m_s+1}(u_s), \ldots, N^2(u_s), N(u_s)
\end{aligned}$$

образует базис пространства $\mathrm{Im}(N)$, в то время как $N^{m_1+2}(u_1) = \cdots = N^{m_s+2}(u_s) = 0$. Какие же векторы можно добавить, чтобы получить базис всего пространства V, имеющий нужный вид «цепочек» векторов? Первое предположение — попытаться добавить векторы $u_1, \ldots, u_s$. Покажем, что полученный набор

$$\begin{aligned}
&N^{m_1+1}(u_1), \ldots, N^2(u_1), N(u_1), u_1,\\
&N^{m_2+1}(u_2), \ldots, N^2(u_2), N(u_2), u_2,\\
&\ldots\\
&N^{m_s+1}(u_s), \ldots, N^2(u_s), N(u_s), u_s
\end{aligned}$$

будет линейно зависим. Действительно, рассмотрим линейную комбинацию этих векторов, равную нулю. Подействуем на эту линейную комбинацию оператором N. Мы получим линейную комбинацию векторов

$$\begin{aligned}
&N^{m_1+2}(u_1), \ldots, N^2(u_1), N(u_1),\\
&N^{m_2+2}(u_2), \ldots, N^2(u_2), N(u_2),\\
&\ldots\\
&N^{m_s+2}(u_s), \ldots, N^2(u_s), N(u_s),
\end{aligned}$$

однако, мы знаем, что векторы $N^{m_1+2}(u_1), \ldots, N^{m_s+2}(u_s)$ равны нулю. Поэтому остается линейная комбинация в точности тех векторов, про которые мы знаем, что они образуют базис $\mathrm{Im}(N)$. Разумеется, из этого следует, что все коэффициенты в ней равны нулю. Возвращаясь к исходной линейной комбинации, видим, что все коэффициенты в ней, кроме, возможно, коэффициентов при векторах $N^{m_1+1}(u_1), \ldots, N^{m_s+1}(u_s)$, равны нулю. Но тогда остается линейная комбинация, состоящая только из указанных векторов, равная нулю. Эти векторы тоже входят в состав выбранного по предположению индукции базиса $\mathrm{Im}(N)$, и потому линейно независимы. Значит, и коэффициенты при них в исходной линейной комбинации также равны нулю.

Итак, мы показали, что векторы

$$\begin{aligned}
&N^{m_1+1}(u_1), \ldots, N^2(u_1), N(u_1), u_1,\\
&N^{m_2+1}(u_2), \ldots, N^2(u_2), N(u_2), u_2,\\
&\ldots\\
&N^{m_s+1}(u_s), \ldots, N^2(u_s), N(u_s), u_s
\end{aligned}$$

линейно независимы. Образуют ли они базис пространства V? Вообще говоря, не обязательно. Поэтому дополним их как-нибудь векторами $w_1, \dots, w_t$ до базиса V. Это еще не нужный нам базис пространства V: нужно его слегка подправить. Заметим, что $N(w_j) \in \mathrm{Im}(N)$ для всех j, и потому $N(w_j)$ является линейной комбинацией векторов

$$\begin{aligned}
&N^{m_1+1}(u_1), \dots, N^2(u_1), N(u_1),\\
&N^{m_2+1}(u_2), \dots, N^2(u_2), N(u_2),\\
&\dots\\
&N^{m_s+1}(u_s), \dots, N^2(u_s), N(u_s),
\end{aligned}$$

образующих, как мы знаем, базис пространства $\mathrm{Im}(N)$. Каждая такая линейная комбинация, очевидно, имеет вид $N(x_j)$, где x_j — линейная комбинация векторов

$$\begin{aligned}
&N^{m_1}(u_1), \dots, N(u_1), u_1,\\
&N^{m_2}(u_2), \dots, N(u_2), u_2,\\
&\dots\\
&N^{m_s}(u_s), \dots, N(u_s), u_s.
\end{aligned}$$

Мы нашли векторы $x_j \in V$ такие, что $N(w_j) = N(x_j)$. Положим $u_{s+j} = w_j - x_j$. Теперь мы утверждаем, что векторы

$$\begin{aligned}
&N^{m_1+1}(u_1), \dots, N^2(u_1), N(u_1), u_1,\\
&\dots\\
&N^{m_s+1}(u_s), \dots, N^2(u_s), N(u_s), u_s,\\
&u_{s+1},\\
&\dots\\
&u_{s+t}
\end{aligned}$$

образуют нужный нам базис пространства V. Напомним, что мы стартовали с базиса, в котором вместо векторов u_{s+j} были векторы w_j, и вычли из каждого w_j некоторую линейную комбинацию x_j предыдущих векторов из того же базиса. Нетрудно видеть, что такая замена обратима, и потому полученный набор векторов также будет базисом пространства V. Кроме того, выполнено и второе условие из заключения теоремы:

$$N^{m_1+2}(u_1) = \dots = N^{m_s+2}(u_s) = N(u_{s+1}) = \dots = N(u_{s+t}),$$

поскольку $N(u_{s+j}) = N(w_j - x_j) = N(w_j) - N(x_j) = 0$. □

8.7 Жорданова форма

Литература: [F], гл. XII, § 6, п. 4; [K2], гл. 2, § 4, пп. 1, 2; [KM], ч. 1, § 9; [vdW], гл. XII, § 87.

Теперь мы готовы сформулировать основной результат о линейных операторах на конечномерных векторных пространствах над алгебраически замкнутым полем.

Определение 8.7.1. Матрица вида

$$J_n(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & \lambda \end{pmatrix}$$

размера $n \times n$ называется **жордановым блоком**. Блочно-диагональная матрица, в которой каждый блок является жордановым блоком, называется **жордановой матрицей**. Пусть $T\colon V \to V$ — линейный оператор. Базис пространства V называется **жордановым базисом** для оператора T, если матрица T в этом базисе является жордановой. Эта матрица тогда называется **жордановой формой** оператора T.

Для доказательства основной теоремы нам понадобится следующая лемма:

Лемма 8.7.2. *Пусть V — векторное пространство над полем k, $T\colon V \to V$ — линейный оператор, и пусть $V = U_1 \oplus \dots \oplus U_m$ — разложение пространства в прямую сумму подпространств, каждое из которых T-инвариантно. Тогда*

$$\dim(\operatorname{Ker}(T)) = \dim(\operatorname{Ker}(T|_{U_1})) + \dots + \dim(\operatorname{Ker}(T|_{U_m}))$$

и

$$\dim(\operatorname{Im}(T)) = \dim(\operatorname{Im}(T|_{U_1})) + \dots + \dim(\operatorname{Im}(T|_{U_m})).$$

Доказательство. Очевидно, что $\operatorname{Ker}(T|_{U_i}) \leqslant \operatorname{Ker}(T)$. Кроме того, каждое $\operatorname{Ker}(T|_{U_i})$ является подпространством в U_i. Сумма $U_1 + \dots + U_m$ прямая, потому и сумма $\operatorname{Ker}(T|_{U_1}) + \dots + \operatorname{Ker}(T|_{U_m})$ прямая. Покажем, что $\operatorname{Ker}(T) \leqslant \operatorname{Ker}(T|_{U_1}) + \dots + \operatorname{Ker}(T|_{U_m})$. Действительно, пусть $v \in \operatorname{Ker}(T)$, и $v = u_1 + \dots + u_m$, где $u_i \in U_i$. Тогда $0 = T(v) = T(u_1) + \dots + T(u_m)$. При этом каждый вектор $T(u_i)$ лежит в U_i в силу T-инвариантности подпространства U_i. Из определения прямой суммы теперь следует, что каждое $T(u_i)$ равно нулю, то есть, $u_i \in \operatorname{Ker}(T|_{U_i})$, и нужное включение доказано.

Таким образом, $\operatorname{Ker}(T) = \operatorname{Ker}(T|_{U_1}) \oplus \dots \oplus \operatorname{Ker}(T|_{U_m})$. Вычисляя размерности, получаем первое из требуемых равенств. После этого второе следует по теореме о гомоморфизме 7.3.8. □

Теорема 8.7.3. *Пусть k — алгебраически замкнутое поле, V — конечномерное векторное пространство над k, T — линейный оператор на V. Тогда в V существует жорданов базис для T. Более того, жорданова форма оператора T единственна с точностью до перестановки жордановых блоков.*

Доказательство. По теореме 8.4.10 пространство V раскладывается в прямую сумму корневых подпространств оператора T. Более того, если $\lambda_i \in k$ — собственное число оператора T, то ограничение оператора $T - \mathrm{id}_V \lambda_i$ на корневое подпространство $V(\lambda_i, T)$ нильпотентно. К

этой ситуации можно применить теорему 8.6.1 и выбрать базис в $V(\lambda_i, T)$, в котором матрица оператора $(T - \mathrm{id}_V \lambda_i)|_{V(\lambda_i,T)}$ имеет вид, описанный в замечании 8.6.2. Матрица оператора $T|_{V(\lambda_i,T)}$ в выбранном базисе получается прибавлением к ней скалярной матрицы с λ_i на диагонали. Получаем, что матрица оператора $T|_{V(\lambda_i,T)}$ имеет жорданов вид (а именно, состоит из блоков $J_{m_1+1}(\lambda_i), \dots, J_{m_s+1}(\lambda_i$, где $m_1, \dots, m_s$ как в теореме 8.4.10). Проделав указанную процедуру для всех собственных чисел, мы получим базис во всем пространстве V, в котором матрица оператора T жорданова.

Осталось показать единственность жордановой формы. Заметим, что на диагонали в жордановой формы обязаны стоять собственные числа оператора T. Поэтому достаточно показать, что для каждого собственного числа λ оператора T размеры блоков вида $J_?(\lambda)$, встречающиеся в любой его жордановой форме, определены однозначно (не зависят от выбора этой формы). Для этого мы выразим количества блоков вида $J_1(\lambda), J_2(\lambda), \dots$ через числа, которые никак не зависят от выбора базиса в пространстве V.

А именно, пусть оператор T приведен к жордановой форме (некоторым выбором базиса). Фиксируем некоторое собственное число λ оператора T, и пусть n_m — количество блоков вида $J_m(\lambda)$ в этой форме. Будем считать, что максимальный размер блока такого вида равен s, и потому $n_{s+1} = n_{s+2} = \dots = 0$.

Посмотрим на размерность ядра оператора $T - \mathrm{id}_V \lambda$. Матрица этого оператора блочно-диагональна и составлена из блоков вида $J_?(\lambda_i - \lambda)$, где λ_i — все собственные числа оператора T. По лемме 8.7.2 достаточно просуммировать размерности ядер этих блоков. Если $\lambda_i \neq \lambda$, то блок вида $J_?(\lambda_i - \lambda)$ обратим по предложению 8.2.4, и вносит нулевой вклад в суммарную размерность ядра. В то же время, если $\lambda_i = \lambda$, то каждый блок вида $J_t(\lambda_i - \lambda) = J_t(0)$ имеет ранг $t - 1$ и размер t, поэтому вности вклад 1 в суммарную размерность ядра. Суммируя, получаем, что размерность ядра оператора $T - \mathrm{id}_V \lambda$ равна количеству блоков вида $J_?(\lambda)$ в жордановой форме оператора T, то есть, $n_1 + n_2 + \dots + n_s$:

$$\dim \mathrm{Ker}(T - \mathrm{id}_V \lambda) = n_1 + n_2 + n_3 + \dots + n_s.$$

Теперь посчитаем размерность ядра оператора $(T - \mathrm{id}_V \lambda)^2$. Снова можно применить лемму 8.7.2, и снова блоки в матрице оператора T вида $J_?(\lambda_i)$ при $\lambda_i \neq \lambda$ вносят нулевой вклад в суммарную размерность ядра. Посмотрим теперь на блок вида $J_t(\lambda)$. Матрица оператора $(T - \mathrm{id}_V \lambda)^2$ равна $(J_t(\lambda) - E_t\lambda)^2$. Нетрудно видеть, что при возведении в квадрат матрица вида

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

превращается в матрицу вида

$$\begin{pmatrix} 0 & 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \dots & 0 \end{pmatrix}.$$

Ранее мы посчитали, что каждый блок $J_t(\lambda)$ вносит вклад 1 в размерность $\operatorname{Ker}(T - \mathrm{id}_V \lambda)$. Теперь видно, что блоки размера 2 и больше вносят вклад еще на 1 больше в размерность $\operatorname{Ker}(T - \mathrm{id}_V \lambda)^2$. В то же время, блоки размера 1×1 при возведении в квадрат не меняются, и потому вносят тот же вклад, что и раньше. Мы получаем, что *разность* размерностей ядер операторов $(T - \mathrm{id}_V \lambda)^2$ и $T - \mathrm{id}_V \lambda$ равна количеству блоков размера 2 и больше:

$$\dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^2 - \dim \operatorname{Ker}(T - \mathrm{id}_V \lambda) = n_2 + n_3 + \dots + n_s.$$

Посчитаем размерность ядра оператора $(T - \mathrm{id}_V \lambda)^3$. Аналогичные рассуждения показывают, что блоки размера 1 и 2 с собственным числом λ при возведении в куб дают то же, что и про возведении в квадрат, а вот у блоков размера 3 и больше единицы «сдвигаются» на диагональ выше, и потому они вносят вклад на 1 больше, чем в размерность ядра оператора $(T - \mathrm{id}_V \lambda)^2$. Поэтому

$$\dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^3 - \dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^2 = n_3 + \dots + n_s.$$

Продолжая увеличивать степень, мы дойдем до последней:

$$\dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^s - \dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^{s-1} = n_s.$$

Полученные равенства можно воспринимать как систему линейных уравнений на $n_1, \dots, n_s$. Нетрудно видеть теперь, что (как и обещано) числа $n_1, \dots, n_s$ выражаются через размерности ядер степеней оператора $(T - \mathrm{id}_V \lambda)$, то есть, через параметры, которые никак не зависят от выбора базиса. Вычитая каждую строчку из предыдущей, можно написать и явную формулу:

$$n_m = 2 \dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^m - \dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^{m-1} - \dim \operatorname{Ker}(T - \mathrm{id}_V \lambda)^{m+1}.$$

Поэтому количество блоков размера m с собственным числом λ в жордановой форме оператора T не зависит от выбора жорданова базиса. □

8.8 Комплексификация

Жорданова форма дает ответ к задаче классификации линейных операторов на конечномерном пространстве над алгебраически замкнутым полем. Этот результат можно пытаться обобщать на разные контексты. Например, можно задуматься о классификации операторов на бесконечномерных пространствах. Наш подход существенно опирался на матричные вычисления, которые не переносятся на бесконечномерный случай, поэтому мы не будем этого делать. Второе направление обобщения — попробовать посмотреть на случай незамкнутого поля.

Действительно, хотя случай алгебраически замкнутого поля уже полезен для приложений (в большинстве неалгебраических приложений встречается случай поля комплексных чисел $\mathbb{C}$), естественный интерес представляют операторы над полем вещественных чисел. Мы продемонстрируем, как основные понятия и факты об операторах переносятся с $\mathbb{C}$ на $\mathbb{R}$.

Итак, пусть V — векторное пространство над полем вещественных чисел $\mathbb{R}$. Мы детально изучили пространства и операторы над полем $\mathbb{C}$, поэтому первое, что нужно попробовать сделать — свести один случай к другому. А именно, мы построим по V пространство $V_{\mathbb{C}}$ над полем комплексных чисел, и покажем, что любой базис в V превращается в базис пространства $V_{\mathbb{C}}$, а любой линейный оператор на V — в линейный оператор на $V_{\mathbb{C}}$.

Рассмотрим множество $V \times V$. По определению оно состоит из всевозможных упорядоченных пар (u, v), где $u, v \in V$. Мы же будем записывать пару (u, v) в виде $u + vi$ и воспринимать как один вектор. Сейчас мы введем на $V \times V$ структуру векторного пространства над полем комплексных чисел $\mathbb{C}$. Сложение определить несложно: $(u_1 + v_1 i) + (u_2 + v_2 i) = (u_1 + u_2) + (v_1 + v_2)i$ для всех $u_1, v_1, u_2, v_2 \in V$. Определим умножение на скаляр $a + bi \in \mathbb{C}$ следующим образом: $(u + vi)(a + bi) = (au - bv) + (av + bu)i$. Видно, что это определение совершенно естественно, и получается простым раскрытием скобок с учетом тождества $i^2 = -1$. Тем не менее, мы должны проверить, что все свойства из определения векторного пространства выполняются. К счастью, эта проверка совсем несложна, и мы оставляем ее читателю в качестве упражнения. Отметим лишь, что роль нулевого элемента играет вектор $0 = 0 + 0i$.

Определение 8.8.1. Полученное векторное пространство над $\mathbb{C}$ мы будем обозначать через $V_{\mathbb{C}}$ и называть **комплексификацией** пространства V.

Исходное векторное пространство V мы будем считать подмножеством в $V_{\mathbb{C}}$: если $v \in V$, то $v + 0i \in V_{\mathbb{C}}$.

Предложение 8.8.2. *Пусть V — векторное пространство над $\mathbb{R}$. Если $v_1, \ldots, v_n$ — базис V (как пространства над $\mathbb{R}$), то $v_1, \ldots, v_n$ — базис $V_{\mathbb{C}}$ (как пространства над $\mathbb{C}$).*

Доказательство. Заметим, что линейная оболочка векторов $v_1, \ldots, v_n$ в $V_{\mathbb{C}}$ содержит векторы $v_1, \ldots, v_n$ и векторы $v_1 i, \ldots, v_n i$. Любой элемент $u \in V$ есть линейная комбинация векторов $v_1, \ldots, v_n$, и для любого $v \in V$ вектор vi есть линейная комбинация векторов $v_1 i, \ldots, v_n i$. Поэтому любой элемент $u + vi \in V_{\mathbb{C}}$ лежит в линейной оболочке $v_1, \ldots, v_n$. Покажем, что $v_1, \ldots, v_n$ линейно независимы в $V_{\mathbb{C}}$. Если $a_1 + b_1 i, \ldots, a_n + b_n i \in \mathbb{C}$ таковы, что $v_1(a_1 + b_1 i) + \cdots + v_n(a_n + b_n i) = 0$, то, раскрывая скобки и приравнивая отдельно «вещественные» и «мнимые» части, получаем, что $v_1 a_1 + \cdots + v_n a_n = 0$ и $v_1 b_1 + \cdots + v_n b_n = 0$. Из линейной независимости векторов $v_1, \ldots, v_n$ в V следует, что $a_1 = \cdots = a_n = b_1 = \cdots = b_n = 0$. Поэтому $v_1, \ldots, v_n$ линейно независимы в $V_{\mathbb{C}}$. □

Следствие 8.8.3. *Размерность $V_{\mathbb{C}}$ как векторного пространства над $\mathbb{C}$ равна размерности V как векторного пространства над $\mathbb{R}$.*

Доказательство. Сразу следует из предложения 8.8.2. □

Определение 8.8.4. Пусть V — векторное пространство над $\mathbb{R}$, T — линейный оператор на V. Определим оператор $T_{\mathbb{C}}$ на пространстве $V_{\mathbb{C}}$ следующим образом:

$$T_{\mathbb{C}}(u + vi) = T(u) + T(v)i$$

для всех $u, v \in V$. Этот оператор называется **комплексификацией** оператора T.

Неформально говоря, оператор $T_{\mathbb{C}}$ действует отдельно на вещественную и мнимую часть вектора $u + vi$ оператором T. Несложно проверить, что эта формула действительно задает линейный оператор на пространстве $V_{\mathbb{C}}$.

Лемма 8.8.5. *Пусть V — векторное пространство над $\mathbb{R}$ с базисом $v_1, \ldots, v_n$, $T\colon V \to V$ — линейный оператор. Тогда матрица оператора T в базисе $v_1, \ldots, v_n$ совпадает с матрицей оператора $T_{\mathbb{C}}$ в том же базисе.*

Доказательство. Упражнение. □

Наш первый результат можно считать аналогом предложения 8.2.1, которое утверждало, что у любого оператора на конечномерном пространстве над алгебраически замкнутым полем есть одномерное инвариантное подпространство.

Предложение 8.8.6. *У любого оператора на (ненулевом) конечномерном векторном пространстве над $\mathbb{R}$ есть инвариантное подпространство размерности 1 или 2.*

Доказательство. Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор. Его комплексификация $T_{\mathbb{C}}\colon V_{\mathbb{C}} \to V_{\mathbb{C}}$ имеет собственное число (по предложению 8.2.1) $a + bi$, где $a, b \in \mathbb{R}$. Пусть $u + vi$ — соответствующий ему собственный вектор; $u, v \in V$, при этом u и v не равны одновременно нулю. Это означает, что $T_{\mathbb{C}}(u + vi) = (u + vi)(a + bi)$. Используя определение $T_{\mathbb{C}}$ и умножения в пространстве $V_{\mathbb{C}}$, получаем

$$T(u) + T(v)i = (ua - vb) + (va + ub)i.$$

Поэтому $T(u) = ua - vb$ и $T(v) = va + ub$. Пусть U — линейная оболочка векторов u, v в V. Тогда U — подпространство в V размерности 1 или 2, и полученные равенства показывают, что U инвариантно относительно оператора T. □

Напомним, что мы определили минимальный многочлен оператора над произвольным полем k (см. определение 8.2.1).

Предложение 8.8.7. *Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор. Тогда минимальный многочлен оператора $T_{\mathbb{C}}$ равен минимальному многочлену оператора T.*

Доказательство. Пусть $p \in \mathbb{R}[x]$ — минимальный многочлен оператора T. Сейчас мы покажем, что он удовлетворяет определению минимального многочлена оператора $T_{\mathbb{C}}$. Сначала необходимо показать, что $p(T_{\mathbb{C}}) = 0$. Напомним, что по определению $T_{\mathbb{C}}(u + vi) = T(u) + T(v)i$.

Применяя к этому равенству оператор $T_{\mathbb{C}}$, получаем, что $(T_{\mathbb{C}})^n(u + vi) = T^n(u) + T^n(v)i$. Поэтому $p(T_{\mathbb{C}}) = (p(T))_{\mathbb{C}} = 0$.

Пусть теперь $q \in \mathbb{C}[x]$ — некоторый многочлен со старшим коэффициентом 1, для которого $q(T_{\mathbb{C}}) = 0$. Нам нужно показать, что степень q не меньше, чем степень p. Заметим, что $(q(T_{\mathbb{C}}))(u) = 0$ для всех $u \in V$. Обозначим через r многочлен, j-й коэффициент которого равен вещественной части j-го коэффициента многочлена q. Очевидно, что старший коэффициент r также равен единице. Из равенства $(q(T_{\mathbb{C}}))(u) = 0$ немедленно следует, что $(r(T))(u) = 0$. Это выполнено для всех $u \in V$, и потому $r(T)$ — нулевой оператор. В силу минимальности p из этого следует, что $\deg r \geqslant \deg p$. Но $\deg r = \deg q$, откуда $\deg q \geqslant \deg p$, что и требовалось. □

Теперь посмотрим на собственные числа комплексификации $T_{\mathbb{C}}$. Каждое собственное число может оказаться вещественным, а может — невещественным. Оказывается, вещественные собственные числа $T_{\mathbb{C}}$ — это собственные числа исходного оператора T.

Предложение 8.8.8. *Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор, $\lambda \in \mathbb{R}$. Число λ является собственным числом оператора $T_{\mathbb{C}}$ тогда и только тогда, когда λ является собственным числом оператора T.*

Доказательство. По предложению 8.5.10 собственные числа оператора T (которые вещественны по определению) — это в точности (вещественные) корни минимального многочлена оператора T. С другой стороны (снова по предложению 8.5.10), вещественные собственные числа оператора $T_{\mathbb{C}}$ — это в точности вещественные корни минимального многочлена оператора $T_{\mathbb{C}}$. По предложению 8.8.7 эти минимальные многочлены совпадают. □

Следующее предложение утверждает, что $T_{\mathbb{C}}$ ведет себя симметрично по отношению к собственному числу λ и сопряженному к нему $\overline{\lambda}$.

Предложение 8.8.9. *Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор, $\lambda \in \mathbb{C}$, j — натуральное число, и $u, v \in V$. Тогда*

$$(T_{\mathbb{C}} - \mathrm{id}_{V_{\mathbb{C}}}\lambda)^j(u + vi) = 0 \iff (T_{\mathbb{C}} - \mathrm{id}_{V_{\mathbb{C}}}\overline{\lambda})^j(u - vi) = 0.$$

Доказательство. Будем доказывать утверждение индукцией по j. В случае $j = 0$ слева и справа стоит тождественный оператор, поэтому мы получаем утверждение, что равенство $u + vi = 0$ равносильно равенству $u - vi = 0$, что очевидно. Пусть теперь $j \geqslant 1$, и мы доказали результат для $j - 1$. Предположим, что $(T_{\mathbb{C}} - \mathrm{id}\,\lambda)^j(u + vi) = 0$. Это означает, что $(T_{\mathbb{C}} - \mathrm{id}\,\lambda)^{j-1}((T_{\mathbb{C}} - \mathrm{id}\,\lambda)(u + vi)) = 0$. Пусть $\lambda = a + bi$, где $a, b \in \mathbb{R}$. Тогда

$$(T_{\mathbb{C}} - \mathrm{id}\,\lambda)(u + vi) = (T(u) - ua + vb) + (T(v) - va - ub)i.$$

Значит, наше равенство можно записать в виде

$$(T_{\mathbb{C}} - \mathrm{id}\,\lambda)^{j-1}((T(u) - ua + vb) + (T(v) - va - ub)i) = 0.$$

По предположению индукции из него следует, что

$$(T_{\mathbb{C}} - \mathrm{id}\,\overline{\lambda})^{j-1}((T(u) - ua + vb) - (T(v) - va - ub)i) = 0.$$

Но прямое вычисление показыват, что

$$(T(u) - ua + vb) - (T(v) - va - ub)i = (T_{\mathbb{C}} - \mathrm{id}\,\overline{\lambda})(u + vi).$$

Мы получили, что $(T_{\mathbb{C}} - \mathrm{id}\,\overline{\lambda})^j(u + vi) = 0$, что и требовалось.

Заменив в приведенном рассуждении λ на $\overline{\lambda}$, а v на $-v$, мы получим и обратное следствие. □

Важным следствием предложения 8.8.9 является тот факт, что невещественные собственные числа оператора $T_{\mathbb{C}}$ ходят парами.

Следствие 8.8.10. *Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор, $\lambda \in \mathbb{C}$. Число λ является собственным числом оператора $T_{\mathbb{C}}$ тогда и только тогда, когда $\overline{\lambda}$ является собственным числом оператора $T_{\mathbb{C}}$.*

Доказательство. Достаточно положить $j = 1$ в предложении 8.8.9. □

Нетрудно проверить, что и кратности сопряженных собственных чисел λ и $\overline{\lambda}$ совпадают.

Следствие 8.8.11. *Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор, $\lambda \in \mathbb{C}$ — собственное число оператора $T_{\mathbb{C}}$. Тогда кратность λ как собственного числа $T_{\mathbb{C}}$ равна кратности $\overline{\lambda}$ как собственного числа $T_{\mathbb{C}}$.*

Доказательство. По определению кратность собственного числа — это размерность соответствующего корневого подпространства. Пусть $u_1 + v_1 i, \ldots, u_m + v_m i$ — базис корневого подпространства $V(\lambda, T_{\mathbb{C}})$, где $u_1, \ldots, u_m, v_1, \ldots, v_m \in V$. Покажем, что тогда векторы $u_1 - v_1 i, \ldots, u_m - v_m i$ образуют базис корневого подпространства $V(\overline{\lambda}, T_{\mathbb{C}})$. Проверим сначала, что они лежат в этом подпространстве: по определению корневого вектора $(T_{\mathbb{C}} - \mathrm{id}\,\lambda)^{\dim(V)}(u_j + v_j i) = 0$, и по предложению 8.8.9 тогда $(T_{\mathbb{C}} - \mathrm{id}\,\overline{\lambda})^{\dim(V)}(u_j - v_j i) = 0$.

Несложно проверить и линейную независимость векторов $u_1 - v_1 i, \ldots, u_m - v_m i$: если $(u_1 - v_1 i)\mu_1 + \cdots + (u_m - v_m i)\mu_m = 0$, то прямые вычисления показывают, что $(u_1 + v_1 i)\overline{\mu_1} + \cdots + (u_m + v_m i)\overline{\mu_m} = 0$, и потому все коэффициенты $\mu_1, \ldots, \mu_m$ равны нулю.

Наконец, нужно проверить, что это система образующих корневого подпространства $V(\overline{\lambda}, T_{\mathbb{C}})$. Пусть $u + vi \in V(\overline{\lambda}, T_{\mathbb{C}})$. Тогда (снова по предложению 8.8.9) $u - vi \in V(\lambda, T_{\mathbb{C}})$. Значит, $u - vi$ является линейной комбинацией векторов $u_1 + v_1 i, \ldots, u_m + v_m i$:

$$u - vi = (u_1 + v_1 i)\mu_1 + \cdots + (u_m + v_m i)\mu_m.$$

Но тогда $u + vi$ является линейной комбинацией векторов $u_1 - v_1 i, \ldots, u_m - v_m i$:

$$u + vi = (u_1 - v_1 i)\overline{\mu_1} + \cdots + (u_m - v_m i)\overline{\mu_m}.$$

□

Приведем еще один вариант переноса предложения 8.2.1 на случай вещественных пространств.

Предложение 8.8.12. *У линейного оператора на пространстве нечетной размерности над* $\mathbb{R}$ *есть собственное число.*

Доказательство. Пусть V — векторное пространство над $\mathbb{R}$ нечетной размерности, $T\colon V \to V$ — линейный оператор. По следствию 8.8.11 невещественные собственные числа оператора $T_{\mathbb{C}}$ встречаются с одинаковой кратностью. Поэтому сумма кратностей всех невещественных собственных чисел оператора $T_{\mathbb{C}}$ четна. С другой стороны, сумма кратностей всех собственных чисел оператора $T_{\mathbb{C}}$ равна размерности пространства $V_{\mathbb{C}}$ (по теореме 8.5.3), и потому равна размерности пространства V (по следствию 8.8.3), то есть, нечетна. Поэтому у $T_{\mathbb{C}}$ есть вещественное собственное число, и по предложению 8.8.8 оно является собственным числом оператора T. □

8.9 Вещественная жорданова форма

Введем понятие характеристического многочлена вещественного оператора. Для этого нам понадобится следующее предложение.

Предложение 8.9.1. *Пусть* V *— векторное пространство над* $\mathbb{R}$, $T\colon V \to V$ *— линейный оператор. Тогда все коэффициенты характеристического многочлена оператора* $T_{\mathbb{C}}$ *вещественны.*

Доказательство. Пусть λ — невещественное собственное число оператора $T_{\mathbb{C}}$, имеющее кратность m. По следствию 8.8.11 число $\overline{\lambda}$ также является собственным числом оператора $T_{\mathbb{C}}$ кратности m. Поэтому в характеристическом многочлене оператора $T_{\mathbb{C}}$ присутствуют множители $(x-\lambda)^m$ и $(x-\overline{\lambda})^m$. Перемножая эти два множителя, получаем

$$(x-\lambda)^m(x-\overline{\lambda})^m = ((x-\lambda)(x-\overline{\lambda}))^m = (x^2-(\lambda+\overline{\lambda})x+\lambda\overline{\lambda})^m.$$

Мы получили многочлен с вещественными коэффициентами, поскольку $\lambda+\overline{\lambda} = 2\operatorname{Re}(\lambda)$ и $\lambda\overline{\lambda} = |\lambda|^2$. Характеристический многочлен оператора $T_{\mathbb{C}}$ является произведением пар скобок указанного вида и скобок вида $(x-t)^d$ для вещественных собственных чисел t оператора $T_{\mathbb{C}}$ кратности d. Поэтому в произведении получаем многочлен с вещественными коэффициентами. □

Определение 8.9.2. Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор. **Характеристическим многочленом** оператора T называется характеристический многочлен оператора $T_{\mathbb{C}}$.

С таким определением совсем несложно доказать аналог предложения 8.5.5.

Предложение 8.9.3. *Пусть* V *— векторное пространство над* $\mathbb{R}$, $T\colon V \to V$ *— линейный оператор. Тогда характеристический многочлен* T *лежит в* $\mathbb{R}[x]$, *его степень равна* $\dim V$, *а его корни — это в точности все вещественные собственные числа оператора* T.

Доказательство. Характеристический многочлен лежит в $\mathbb{R}[x]$ по предложению 8.9.1, имеет степень $\dim V$ по предложению 8.5.5 и следствию 8.8.3, и имеет нужные корни по предложению 8.5.5 и предложению 8.8.8. □

Несложно получить и аналог теоремы Гамильтона–Кэли 8.5.6.

Теорема 8.9.4 (Гамильтона–Кэли). *Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор. Пусть q — характеристический многочлен оператора T. Тогда $q(T) = 0$.*

Доказательство. По теореме 8.5.6 имеем $q(T_{\mathbb{C}}) = 0$, откуда следует, что $q(T) = 0$ (см. рассуждение в начале доказательства предложения 8.8.7). □

Теперь мы готовы сформулировать аналог теоремы о жордановой форме для вещественных операторов.

Определение 8.9.5. Вещественным жордановым блоком называется матрица вида

$$J_n(c) = \begin{pmatrix} c & 1 & 0 & \dots & 0 & 0 \\ 0 & c & 1 & \dots & 0 & 0 \\ 0 & 0 & c & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & c & 1 \\ 0 & 0 & 0 & \dots & 0 & c \end{pmatrix}$$

размера $n \times n$, где $c \in \mathbb{R}$, или матрица вида

$$J_n(\lambda) = \begin{pmatrix} a & b & 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ -b & a & 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & a & b & 1 & 0 & \dots & 0 & 0 \\ 0 & 0 & -b & a & 0 & 1 & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & a & b & \dots & 0 & 0 \\ 0 & 0 & 0 & 0 & -b & a & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & a & b \\ 0 & 0 & 0 & 0 & 0 & 0 & \dots & -b & a \end{pmatrix}$$

размера $(2n) \times (2n)$, где $\lambda = a + bi$, $a, b \in \mathbb{R}$, причем $b > 0$. Блочно-диагональная матрица, в которой каждый блок является вещественным жордановым блоком, называется **вещественной жордановой матрицей**. Пусть V — векторное пространство над $\mathbb{R}$, $T\colon V \to V$ — линейный оператор. Базис пространства V называется **вещественным жордановым базисом** для оператора T, если матрица T в этом базисе является вещественной жордановой. Эта матрица тогда называется **вещественной жордановой формой** оператора T.

Теорема 8.9.6. *Пусть V — конечномерное векторное пространство над $\mathbb{R}$, T — линейный оператор на V. Тогда в V существует вещественный жорданов базис для T. Более того, вещественная жорданова форма оператора T единственна с точностью до перестановки вещественных жордановых блоков.*

Набросок доказательства. Поясним, откуда берутся вещественные жордановы блоки вида $J_n(\lambda)$ для комлпексных чисел $\lambda = a + bi$, $b \neq 0$. Рассмотрим комплексификацию $T_{\mathbb{C}}$ оператора T. Мы знаем, что в $V_{\mathbb{C}}$ существует базис, в котором матрица оператора $T_{\mathbb{C}}$ имеет жорданов вид. Теперь мы хотим перейти от этого базиса к базису пространства V так, чтобы матрица оператора T в нем выглядела не очень отлично от матрицы $T_{\mathbb{C}}$ в жордановом базисе.

Пусть λ — невещественное собственное число оператора $T_{\mathbb{C}}$, $\lambda = a + bi$. Мы выяснили, что тогда и $\overline{\lambda}$ является собственным числом оператора $T_{\mathbb{C}}$. Поменяв при необходимости λ и $\overline{\lambda}$ местами, можем считать, что $b > 0$. Оказывается, тогда и все размеры жордановых блоков, соответствующих числам λ и $\overline{\lambda}$, совпадают. Действительно, в доказательстве теоремы 8.7.3 мы выразили эти размеры блоков через размерности операторов вида $(T_{\mathbb{C}} - \mathrm{id}\,\lambda)^j$. Рассуждение, аналогичное доказательству следствия 8.8.11, показывает, что эти размерности для чисел λ и $\overline{\lambda}$, совпадают; поэтому и размеры блоков совпадают.

Более того, рассмотрим какой-нибудь жорданов блок вида $J_m(\lambda)$. Пусть $u_1 + v_1 i, \ldots, u_m + v_m i$ — соответствующие базисные векторы. Тогда векторы $u_1 - v_1 i, \ldots, u_m - v_m i$ линейно независимы, порождают $T_{\mathbb{C}}$-инвариантное подпространство и в ограничении на это подпространство получаем жорданов блок вида $J_m(\overline{\lambda})$. Таким образом, жордановы блоки, соответствующие невещественным собственным числам оператора $T_{\mathbb{C}}$, разбиваются на «сопряженные» пары. Посмотрим на подпространство в V, порожденное векторами $u_1, v_1, \ldots, u_m, v_m$. Мы утверждаем, что эти векторы линейно независимы, и матрица оператора T, ограниченного на это подпространство, как раз равна вещественному жордановому блоку вида $J_m(\lambda)$.

Действительно, например, мы знаем, что $T_{\mathbb{C}}(u_1 + v_1 i) = (u_1 + v_1 i)(a + bi)$ Раскрывая скобки, получаем, что $T(u_1) = u_1 a - v_1 b$ и $T(v_1) = u_1 b + v_1 a$. Это объясняет первые два столбика в матрице $J_m(\lambda)$. Далее, $T_{\mathbb{C}}(u_2 + v_2 i) = (u_2 + v_2 i)(a + bi) + (u_1 + v_1 i)$. Раскрывая скобки, получаем, что $T(u_2) = u_2 a - v_2 b + u_1$ и $T(v_2) = u_2 b + v_2 a + v_1$. Это объясняет третий и четвертый столбики в матрице $J_m(\lambda)$, и так далее.

Таким образом, можно взять пару комплексных жордановых блоков вида $J_m(\lambda)$ и $J_m(\overline{\lambda})$ и, слегка поменяв базис в соответствующем пространстве размерности $2m$, получить вещественный базис, в котором эти блоки «склеятся» и превратятся в один вещественный жорданов блок $J_m(\lambda)$ размера $2m$. Осталось аккуратно разобраться с вещественными собственными числами: показать, что можно выбрать базис в корневом подпространстве вида $V(c, T_{\mathbb{C}})$ для $c \in \mathbb{R}$ так, что он будет базисом в V, в котором матрица [ограничения] оператора T будет вещественным жордановым блоком вида $J_m(c)$. □

9 Эвклидовы и унитарные пространства

9.1 Эвклидовы пространства

Литература: [F], гл. XIII, § 1, п. 1; [K2], гл. 3, § 1, п. 1; [KM, ч. 2, § 2, пп. 1–3; § 5, п. 1.

Определение 9.1.1. Пусть V — векторное пространство над полем k. Отображение $B\colon V\times V\to k$ называется **билинейной формой**, если оно линейно по каждому аргументу. Иными словами,

$$\begin{aligned}
&B(u_1+u_2,v)=B(u_1,v)+B(u_2,v),\\
&B(u\alpha,v)=B(u,v)\alpha,\\
&B(u,v_1+v_2)=B(u,v_1)+B(u,v_2),\\
&B(u,v\alpha)=B(u,v)\alpha
\end{aligned}$$

для всех $u,v,u_1,u_2,v_1,v_2\in V$ и $\alpha\in k$. Если $B(u,v)=0$, то говорят, что вектор u **ортогонален** вектору v относительно формы B. Обозначение: $u\perp v$.

Определение 9.1.2. Форма B называется **симметрической**, если $B(u,v)=B(v,u)$ для всех $u,v\in V$. Форма B называется **кососимметрической**, если $B(u,v)=-B(v,u)$ для всех $u,v\in V$. Форма B называется **симплектической**, если $B(u,u)=0$ для всех $u\in V$.

Замечание 9.1.3. Симплектическая форма является кососимметрической. Действительно, для любых $u,v\in V$ тогда выполнено $0=B(u+v,u+v)=B(u,u)+B(u,v)+B(v,u)+B(v,v)=B(u,v)+B(v,u)$. Обратное, вообще говоря, неверно. В самом деле, из кососимметричности формы сразу следует, что $B(u,u)=-B(u,u)$, откуда $2B(u,u)=0$ для всех $u\in V$. Если характеристика поля k не равна 2, то $2\in k^*$ и каждая кососимметрическая форма является симплектической. Если же k — поле характеристики 2, то эти два класса форм не совпадают.

Пример 9.1.4. В эвклидовом пространстве $V=\mathbb{R}^n$ над полем $\mathbb{R}$ определены длины векторов и углы между векторами. Поэтому естественно определить *эвклидово скалярное произведение* формулой $(u,v)=|u|\cdot|v|\cdot\cos(\varphi)$, где $|u|$, $|v|$ — длины векторов u, v соответственно, а φ — угол между векторами u и v. Это скалярное произведение симметрично и для любого вектора $v\in V$ выполнено $(v,v)\geqslant 0$. Более того, равенство $(v,v)=0$ выполнено только для $v=0$.

Нас интересует алгебра, поэтому мы будем пользоваться чисто алгебраическими определениями билинейных форм, не ссылающимися на понятия «длины» и «угла»; наоборот, чуть позже мы *определим* слова «длина» и «угол» в терминах билинейных форм.

Пример 9.1.5. Пусть k — произвольное поле, $V=k^n$ — пространство столбцов высоты n над k. Определим форму $B\colon V\times V\to k$ формулой $B(u,v)=u_1v_1+\cdots+u_nv_n$. Иными словами, $B(u,v)=u^Tv$. Нетрудно видеть, что эта форма билинейна

$$\begin{aligned}
&B(u_1+u_2,v)=(u_1+u_2)^Tv=u_1^Tv+u_2^Tv=B(u_1,v)+B(u_2,v)\\
&B(u\lambda,v)=(u\lambda)^Tv=\lambda(u^Tv)=\lambda B(u,v)\\
&B(u,v_1+v_2)=u^T(v_1+v_2)=u^Tv_1+u^Tv_2=B(u,v_1)+B(u,v_2)\\
&B(u,v\lambda)=u^T(v\lambda)=\lambda(u^Tv)=\lambda B(u,v)
\end{aligned}$$

и симметрична

$$B(u,v) = B(u,v)^T = (u^T v)^T = v^T u = B(v,u).$$

Возьмем теперь в предыдущем примере в качестве k поле вещественных чисел $\mathbb{R}$. Заметим, что скалярное произведение вектора на себя является неотрицательным числом: $B(u,u) = u_1^2 + \cdots + u_n^2 \geqslant 0$; более того, $B(u,u) = 0$ только для $u = 0$.

Определение 9.1.6. Пусть V — векторное пространство над $\mathbb{R}$. Билинейная форма $B\colon V \times V \to \mathbb{R}$ называется **неотрицательно определенной**, если $B(u,u) \geqslant 0$ для всех $u \in V$. Форма B называется **положительно определенной**, если она неотрицательно определена и из $B(u,u) = 0$ следует, что $u = 0$.

Определение 9.1.7. Векторное пространство V над полем $\mathbb{R}$ вместе с положительно определенной симметрической билинейной формой $B\colon V \times V \to \mathbb{R}$ называется **эвклидовым пространством**, а форма B называется **эвклидовым скалярным произведением** на V.

Замечание 9.1.8. Любое подпространство $W \leqslant V$ эвклидова пространства (V, B) само является эвклидовым пространством относительно скалярного произведения $B|_{W\times W}\colon W \times W \to \mathbb{R}$, которое мы часто будем обозначать той же буквой B. Действительно, нетрудно проверить, что $B|_{W\times W}$ — симметрическая билинейная форма, и положительная определенность формы $B|_{W\times W}$ сразу следует из положительной определенности формы B.

9.2 Унитарные пространства

Литература: [F], гл. XIII, § 1, пп. 1, 3, [K2], гл. 3, § 2, п. 2; [KM], ч. 2, § 2, пп. 1–3; § 6, п. 1.

В связи с возникновением квантовой механики в первой половине XX века большое практическое значение стало придаваться векторным пространствам над полем комплексных чисел $\mathbb{C}$. Что будет аналогом положительно определенных билинейных форм в этом случае? Заметим, что прямой перенос определения на комплексный случай не работает: если V — векторное пространство над полем $\mathbb{C}$ и $B\colon V \times V \to \mathbb{C}$ — билинейная форма, то $B(iv, iv) = -B(v,v)$ для всех $v \in V$.

Определение 9.2.1. Отображение $B\colon V \times V \to \mathbb{C}$ называется **полуторалинейной формой**, если оно *линейно* по второму аргументу и *полулинейно* по первому аргументу:

$$\begin{aligned}
&B(u, v_1 + v_2) = B(u, v_1) + B(u, v_2)\\
&B(u, v\lambda) = B(u, v)\lambda\\
&B(u_1 + u_2, v) = B(u_1, v) + B(u_2, v)\\
&B(u\lambda, v) = \overline{\lambda}B(u, v)
\end{aligned}$$

для всех $u, v, u_1, u_2, v_1, v_2 \in V$ и всех $\lambda \in \mathbb{C}$.

Аналог условия симметричности формы также должен отличаться от билинейного случая, поскольку теперь $B(u, v\lambda) = \lambda B(u, v)$, но $B(v\lambda, u) = \overline{\lambda}B(v, u)$.

Определение 9.2.2. Полуторалинейная форма $B\colon V \times V \to \mathbb{C}$ называется **эрмитовой**, если для всех $u, v \in V$ выполнено $B(u, v) = \overline{B(v, u)}$.

Замечание 9.2.3. Заметим, что если B — эрмитова форма на V, то $B(u, u) = \overline{B(u, u)}$ для всех $u \in V$, поэтому $B(u, u)$ — вещественное число.

Пример 9.2.4. Пусть $V = \mathbb{C}^n$ — пространство столбцов высоты n над k. Определим форму $B\colon V \times V \to \mathbb{C}$ формулой $B(u, v) = \overline{u_1}v_1 + \cdots + \overline{u_n}v_n$. Иными словами, $B(u, v) = \overline{u}^T v$. Нетрудно видеть, что эта форма полуторалинейная

$$\begin{aligned}
&B(u, v_1 + v_2) = u^T(v_1 + v_2) = \overline{u}^T v_1 + \overline{u}^T v_2 = B(u, v_1) + B(u, v_2)\\
&B(u, v\lambda) = \overline{u}^T(v\lambda) = \lambda(\overline{u}^T v) = \lambda B(u, v)\\
&B(u_1 + u_2, v) = \overline{(u_1 + u_2)}^T v = \overline{u_1}^t v + \overline{u_2}^T v = B(u_1, v) + B(u_2, v)\\
&B(u\lambda, v) = \overline{(u\lambda)}^T v = \overline{\lambda}(\overline{u}^T v) = \overline{\lambda} B(u, v)
\end{aligned}$$

и эрмитова

$$\overline{B(u, v)} = \overline{B(u, v)}^T = \overline{(\overline{u}^T v)}^T = \overline{v^T \overline{u}} = \overline{v}^T u = B(v, u).$$

Заметим, что $B(u, u) = \overline{u_1}u_1 + \cdots + \overline{u_n}u_n = |u_1|^2 + \cdots + |u_n|^2 \geqslant 0$; более того, $B(u, u) = 0$ только для $u = 0$.

Определение 9.2.5. Пусть V — векторное пространство над $\mathbb{C}$. Эрмитова форма $B\colon V \times V \to \mathbb{C}$ называется **неотрицательно определенной**, если $B(u, u) \geqslant 0$ для всех $u \in V$. Форма B называется **положительно определенной**, если она неотрицательно определена и из $B(u, u) = 0$ следует, что $u = 0$.

Определение 9.2.6. Векторное пространство V над полем $\mathbb{C}$ вместе с положительно определенной эрмитовой формой $B\colon V \times V \to \mathbb{C}$ называется **унитарным пространством**, а форма B называется **эрмитовым скалярным произведением** на V.

Замечание 9.2.7. Как и в эвклидовом случае (см. замечание 9.1.8), любое подпространство $W \leqslant V$ унитарного пространства (V, B) само является унитарным пространством относительно скалярного произведения $B|_{W \times W}\colon W \times W \to \mathbb{C}$, которое мы часто будем обозначать той же буквой B.

В дальнейшем мы будем параллельно развивать теорию эвклидовых и унитарных пространств; мы будем обозначать через k поле $\mathbb{R}$ или $\mathbb{C}$. Заметим, что и для эвклидовых, и для унитарных пространств выполнены тождества $B(u, v\lambda) = B(u, v)\lambda$ и $B(u\lambda, v) = \overline{\lambda}B(u, v)$; отличие лишь в том, что для эвклидовых пространств константа λ является вещественной, поэтому $\overline{\lambda} = \lambda$. Кроме того, условия симметричности и эрмитовости также можно записать в единообразном виде: $B(u, v) = \overline{B(v, u)}$.

9.3 Норма

Литература: [F], гл. XII, § 1, пп. 1–3, [K2], гл. 3, § 1, п. 2; § 2, п. 2; [KM], ч. 2, § 2, п. 4; § 5, пп. 2–5; § 6, пп. 4–7.

Определение 9.3.1. Пусть (V, B) — эвклидово или унитарное пространство, $v \in V$. Будем называть число $||v|| = \sqrt{B(v,v)}$ **длиной** v.

Лемма 9.3.2. *Пусть* (V, B) — *эвклидово или унитарное пространство,* $u, v, \in V$. *Тогда*

1. *(Однородность нормы).* $||\lambda v|| = |\lambda| \cdot ||v||$ *для любого* $\lambda \in k$.
2. *(Теорема Пифагора). Если* $B(u,v) = 0$, *то* $||u+v||^2 = ||u||^2 + ||v||^2$.
3. *(Неравенство Коши–Буняковского–Шварца).* $|B(u,v)| \leqslant ||u|| \cdot ||v||$, *причем равенство достигается тогда и только тогда, когда векторы* u *и* v *пропорциональны.*
4. *(Неравенство треугольника).* $||u|| + ||v|| \geqslant ||u+v||$;

Доказательство. Заметим, что для $v = 0$ все утверждения леммы очевидны. Поэтому далее мы будем считать, что $v \neq 0$.

Однородность нормы следует из полуторалинейности:

$$||\lambda v||^2 = B(\lambda v, \lambda v) = \lambda\overline{\lambda}B(v,v) = |\lambda|^2 \cdot ||v||^2.$$

Заметим, что $||u+v||^2 = B(u+v, u+v) = B(u,u) + B(u,v) + \overline{B(u,v)} + B(v,v)$, и при $B(u,v) = 0$ получаем в точности теорему Пифагора.

Для доказательства неравенства Коши–Буняковского–Шварца положим

$$w = u - \frac{B(u,v)}{B(v,v)}v$$

и заметим, что

$$B(w,v) = B(u - \frac{B(u,v)}{B(v,v)}v, v) = B(u,v) - \frac{B(u,v)}{B(v,v)}B(v,v) = 0.$$

Это означает, что векторы v и w ортогональны. Поэтому и вектор $\frac{B(u,v)}{B(v,v)}v$ ортогонален вектору w. Применим к этой паре векторов теорему Пифагора:

$$||u||^2 = ||w||^2 + ||\frac{B(u,v)}{B(v,v)}v||^2 = ||w||^2 + \frac{|B(u,v)|^2}{||v||^2} \geqslant \frac{|B(u,v)|^2}{||v||^2},$$

откуда $|B(u,v)| \leqslant ||u|| \cdot ||v||$. Если достигается равенство, то $||w|| = 0$, откуда $w = 0$ и u пропорционально v; обратно, если u пропорционально v, то в неравенстве Коши–Буняковского–Шварца имеет место равенство.

Посмотрим на выражение для $B(u+v, u+v)$:

$$\begin{aligned}||u+v||^2 &= B(u+v, u+v)\\ &= B(u,u) + B(u,v) + \overline{B(u,v)} + B(v,v)\\ &= ||u||^2 + 2\,\mathrm{Re}(B(u,v)) + ||v||^2 \leqslant ||u||^2 + 2|B(u,v)| + ||v||^2\\ &\leqslant ||u||^2 + 2||u|| \cdot ||v|| + ||v||^2\\ &= (||u|| + ||v||)^2.\end{aligned}$$

Извлекая корень из обеих частей, получаем неравенство треугольника. □

Определение 9.3.3. Пусть (V, B) — эвклидово пространство. Лемма 9.3.2 показывает, что для ненулевых векторов $u, v \in V$ выражение $\frac{B(u,v)}{||u||\cdot||v||}$ лежит на отрезке $[-1, 1]$ и потому является косинусом некоторого однозначно определенного угла $\varphi \in [0, \pi]$. Этот угол называется **углом между векторами** u и v. Обозначение: $\varphi = \angle(u, v)$. Обратите внимание, что это определение не работает для унитарного пространства: $B(u, v)$ может оказаться комплексным. Однако, имеет смысл рассматривать выражение $\frac{|B(u,v)|}{||u||\cdot||v||}$; оно лежит на отрезке $[0, 1]$ и потому является косинусом некоторого однозначно определенного угла $\varphi \in [0, \frac{\pi}{2}]$.

Замечание 9.3.4. Заметим, что угол $\angle(u, v)$ равен $\pi/2$ тогда и только тогда, когда $B(u, v) = 0$, то есть, когда векторы u и v ортогональны в смысле определения 9.1.1.

9.4 Матрица Грама

Литература: [F], гл. XIII, § 1, п. 4; [KM], ч. 2, § 2, пп. 2–3; [KM], ч. 2, § 3, п. 8.

Пусть (V, B) — конечномерное пространство над полем k с формой, билинейной в случае $k = \mathbb{R}$ и полуторалинейной в случае $k = \mathbb{C}$. Пусть $\mathcal{E} = (e_1, \ldots, e_n)$ — базис V. Запишем векторы $u, v \in V$ в этом базисе: $u = e_1u_1 + \cdots + e_nu_n$, $v = e_1v_1 + \cdots + e_nv_n$. Подставим эти выражения в $B(u, v)$:

$$B(u,v) = B(e_1u_1 + \cdots + e_nu_n, e_1v_1 + \cdots + e_nv_n) = \sum_{i,j=1}^{n} B(e_iu_i, e_jv_j) = \sum_{i,j=1}^{n} \overline{u_i}v_jB(e_i, e_j).$$

Это означает, что форма B полностью определяется своими значениями на базисных векторах. Полученное выражение можно записать в матричной форме:

$$B(u,v) = \overline{[u]}^{T}(B(e_i, e_j))_{i,j=1}^{n}[v],$$

где через $[u], [v]$ мы обозначаем столбцы координат векторов u, v в базисе $\mathcal{E}$. Матрица, составленная из скалярных произведений $B(e_i, e_j)$ базисных векторов, называется **матрицей Грама** формы B в базисе $\mathcal{E}$. Обозначим ее через G. Мы получили, что $B(u, v) = \overline{[u]}^{T}G[v]$ для всех $u, v \in V$.

Пока мы использовали только билинейность/полуторалинейность формы B. Если форма B симметрична/эрмитова, то $\overline{B(v,u)} = \overline{B(v,u)}^T = (\overline{[v]}^T G[u])^T = \overline{[u]}^T G^T \overline{[v]} = \overline{[u]}^T \overline{G}^T [v]$. Сравним это с выражением $B(u,v) = \overline{[u]}^T G[v]$:

$$\overline{[u]}^T \overline{G}^T [v] = \overline{[u]}^T G[v] \quad \text{для всех } u, v \in V.$$

Подставляя в качестве u, v базисные векторы $e_1, \dots, e_n$, получаем, что матрицы $\overline{G}^T$ и G совпадают:

$$\overline{G}^T = G.$$

Для случая эвклидова пространства, конечно, это равенство означает, что $G^T = G$.

Определение 9.4.1. Матрица A над произвольным полем называется **симметрической**, если $A^T = A$. Матрица A над полем комплексных чисел называется **эрмитовой**, если $\overline{A}^T = A$.

Таким образом, мы показали, что матрица Грама симметрической билинейной формы является симметрической, а матрица Грама эрмитовой билинейной формы является эрмитовой.

Обратно, по любой симметрической матрице над $\mathbb{R}$ можно построить симметрическую билинейную форму, а по любой эрмитовой матрице над $\mathbb{C}$ — эрмитову полуторалинейную форму. Действительно, мы можем обобщить примеры 9.1.5 и 9.2.4. Пусть $G \in M(n,k)$ — симметрическая или эрмитова матрица. На пространстве столбцов $V = k^n$ высоты n определим форму $B\colon V \times V \to k$ равенством

$$B(u,v) = \overline{u}^T G v.$$

Нетрудно проверить, что эта форма билинейна в случае $k = \mathbb{R}$ и полуторалинейна в случае $k = \mathbb{C}$:

$$\begin{aligned}
&B(u, v_1 + v_2) = \overline{u}^T G(v_1 + v_2) = \overline{u}^T G v_1 + \overline{u}^T G v_2 = B(u, v_1) + B(u, v_2)\\
&B(u, v\lambda) = \overline{u}^T G(v\lambda) = (\overline{u}^T G v)\lambda = B(u,v)\lambda\\
&B(u_1 + u_2, v) = \overline{u_1 + u_2}^T G v = \overline{u_1}^T G v + \overline{u_2}^T G v = B(u_1, v) + B(u_2, v)\\
&B(u\lambda, v) = \overline{u\lambda}^T G v = \overline{\lambda}(\overline{u}^T G v) = \overline{\lambda} B(u,v)
\end{aligned}$$

Кроме того, для симметрической матрицы G имеем

$$B(v,u) = B(v,u)^T = (v^T G u)^T = u^T G^T v = u^T G v = B(u,v),$$

а для эрмитовой —

$$\overline{B(v,u)} = \overline{B(v,u)}^T = (\overline{v}^T G u)^T = \overline{u}^T \overline{G}^T v = \overline{u}^T G v = B(u,v).$$

Поэтому форма B является симметрической или эрмитовой соответственно. По определению исходная матрица G является матрицей Грама полученной формы B в стандартном базисе пространства столбцов.

Естественно поставить вопрос: как меняется матрица Грама при замене базиса в пространстве V? Напомним, что если $\mathcal{E} = \{e_1, \dots, e_n\}$ и $\mathcal{F} = \{f_1, \dots, f_n\}$ — два базиса в пространстве V, то *матрица перехода* $(\mathcal{E} \rightsquigarrow \mathcal{F})$ от базиса $\mathcal{E}$ к базису $\mathcal{F}$ устроена так: в столбце с номером j стоят координаты вектора f_j в базисе $\mathcal{E}$ (см. определение 7.9.1).

Теорема 9.4.2 (Преобразование матрицы Грама при замене базиса). *Пусть $\mathcal{E}, \mathcal{F}$ — два базиса конечномерного пространства V над полем k, $C = (\mathcal{E} \rightsquigarrow \mathcal{F})$ — матрица перехода от $\mathcal{E}$ к $\mathcal{F}$, $B\colon V \times V \to k$ — билинейная или полуторалинейная форма на V. Пусть $G_{\mathcal{E}}$ и $G_{\mathcal{F}}$ — матрицы Грама формы B в базисах $\mathcal{E}$ и $\mathcal{F}$ соответственно. Тогда*

$$G_{\mathcal{F}} = \overline{C}^{T} G_{\mathcal{E}} C.$$

Доказательство. Пусть $u, v \in V$. По теореме 7.9.3 координаты векторов в базисах $\mathcal{E}$, $\mathcal{F}$ связаны следующим образом: $[v]_{\mathcal{E}} = C \cdot [v]_{\mathcal{F}}$, $[u]_{\mathcal{E}} = C \cdot [u]_{\mathcal{F}}$. Поэтому

$$B(u, v) = \overline{[u]_{\mathcal{E}}}^{T} G_{\mathcal{E}} [v]_{\mathcal{E}} = \overline{C \cdot [u]_{\mathcal{F}}}^{T} G_{\mathcal{E}} C \cdot [v]_{\mathcal{F}} = \overline{[u]_{\mathcal{F}}}^{T} \overline{C}^{T} G_{\mathcal{E}} C \cdot [v]_{\mathcal{F}}$$

С другой стороны,

$$B(u, v) = \overline{[u]_{\mathcal{F}}}^{T} G_{\mathcal{F}} [v]_{\mathcal{F}}.$$

Получаем, что $\overline{[u]_{\mathcal{F}}}^{T} \overline{C}^{T} G_{\mathcal{E}} C \cdot [v]_{\mathcal{F}} = \overline{[u]_{\mathcal{F}}}^{T} G_{\mathcal{F}} [v]_{\mathcal{F}}$ для всех $u, v \in V$. Подставляя в качестве u, v всевозможные пары векторов базиса $\mathcal{F}$, получаем необходимое равенство матриц. □

Отметим, что матрица Грама скалярного произведения обратима.

Предложение 9.4.3. *Пусть (V, B) — эвклидово или унитарное пространство. Тогда матрица Грама формы B в любом базисе является обратимой.*

Доказательство. Выберем произвольный базис $\mathcal{E}$ пространства V и запишем матрицу Грама $G = G_{\mathcal{E}} \in M(n, k)$ скалярного произведения B в этом базисе. Если она необратима, то (по теореме Кронекера–Капелли 7.9.11) уравнение $GX = 0$ имеет ненулевое решение: найдется столбец $X_0 \in k^n \backslash \{0\}$, для которого $GX_0 = 0$. Такой столбец является столбцом координат некоторого ненулевого вектора $v_0 \in V$. Но тогда $B(v_0, v_0) = \overline{[v_0]_{\mathcal{E}}}^{T} \cdot G \cdot [v_0]_{\mathcal{E}} = \overline{X_0}^{T} G X_0 = 0$, что противоречит положительной определенности формы B. □

9.5 Процесс ортогонализации Грама–Шмидта

Литература: [F], гл. XIII, § 1, пп. 5, 6; § 2, п. 1; [K2], гл. 3, § 1, п. 3; § 2, п. 3; [KM], ч. 2, § 3, п. 6; § 4, пп. 2–4.

Определение 9.5.1. Пусть (V, B) — эвклидово или унитарное пространство. Базис $(e_1, \ldots, e_n)$ пространства V называется **ортогональным**, если все его векторы попарно ортогональны: $e_i \perp e_j$ при $i \neq j$. Этот базис называется **ортонормированным**, если он ортогонален и длина каждого вектора равна единице: $||e_i|| = 1$ для всех i.

Лемма 9.5.2. *Пусть (V, B) — эвклидово или унитарное пространство. Если ненулевые векторы $e_1, \ldots, e_n \in V$ попарно ортогональны, то они линейно независимо. Если, кроме того, $\dim V = n$, то векторы $e_1, \ldots, e_n$ образуют ортогональный базис.*

Доказательство. Предположим, что $e_1\lambda_1 + \cdots + e_n\lambda_n = 0$ — нетривиальная линейная комбинация этих векторов, равная нулю. Домножим это равенство скалярно на e_i:

$$B(e_i, e_1\lambda_1 + \cdots + e_n\lambda_n) = 0.$$

Пользуясь линейностью по второму аргументу и попарной ортогональностью векторов e_i, получаем равенство $\lambda_i B(e_i, e_i) = 0$. Так как $e_i \neq 0$, получаем, что $\lambda_i = 0$ для всех $i = 1, \ldots, n$.

Если $\dim V = n$, мы получаем n линейно независимых векторов в n-мерном векторном пространстве. Из предложения 6.5.3 следует, что они образуют базис (действительно, размерность их линейной оболочки совпадает с размерностью V, поэтому эта линейная оболочка равна V). □

Замечание 9.5.3. По определению матрица Грама формы B в базисе $\mathcal{E} = (e_1, \ldots, e_n)$ составлена из скалярных произведений $B(e_i, e_j)$. Поэтому базис $\mathcal{E}$ ортогонален тогда и только тогда, когда матрица Грама скалярного произведения в этом базисе диагональна; базис $\mathcal{E}$ ортонормирован тогда и только тогда, когда матрица Грама скалярного произведения в этом базисе единична.

Таким образом, если нам дано эвклидово или унитарное пространство, часто удобно выбрать в нем ортогональный базис: в нем скалярное произведение задается простыми формулами через координаты векторов (см. примеры 9.1.5 и 9.2.4: стандартные базисы пространства столбцов являются ортонормированными относительно рассматриваемых там форм).

Лемма 9.5.4 (Процесс ортогонализации Грама–Шмидта). *Пусть (V, B) — эвклидово или унитарное пространство, $e_1, \ldots, e_{n-1}$ — семейство попарно ортогональных ненулевых векторов, $v \notin \langle e_1, \ldots, e_{n-1}\rangle$. Тогда существует вектор $e_n \in V$ такой, что e_n ортогонален всем векторам $e_1, \ldots, e_{n-1}$ и, кроме того, $\langle e_1, \ldots, e_{n-1}, v\rangle = \langle e_1, \ldots, e_{n-1}, e_n\rangle$.*

Доказательство. Будем искать вектор e_n в виде

$$e_n = v - e_1\lambda_1 - e_2\lambda_2 - \cdots - e_{n-1}\lambda_{n-1}.$$

Подберем коэффициенты $\lambda_1, \ldots, \lambda_{n-1} \in k$ так, чтобы e_n был ортогонален каждому e_i, $i = 1, \ldots, n-1$. Посмотрим на скалярное произведение e_n и e_i. Поскольку e_i ортогонален всем векторам из $e_1, \ldots, e_{n-1}$, кроме e_i, получаем

$$B(e_i, e_n) = B(e_i, v) - B(e_i, e_i)\lambda_i.$$

Положим теперь $\lambda_i = \frac{B(e_i, v)}{B(e_i, e_i)}$; заметим, что $B(e_i, e_i) \neq 0$, поскольку $e_i \neq 0$. Мы добились того, что $e_n \perp e_i$ для всех $i = 1, \ldots, n-1$. Кроме того, v выражается через $e_1, \ldots, e_n$, поэтому $v \in \langle e_1, \ldots, e_n\rangle$, и e_n выражается через $e_1, \ldots, e_{n-1}, v$, поэтому $e_n \in \langle e_1, \ldots, e_{n-1}, v\rangle$. Это и означает равенство нужных линейных оболочек. □

Следствие 9.5.5. *Пусть (V, B) — эвклидово или унитарное пространство, и пусть $\mathcal{F} = (f_1, \ldots, f_n)$ — базис V. Тогда существует ортогональный базис $\mathcal{E} = (e_1, \ldots, e_n)$ пространства V такой, что $\langle e_1, \ldots, e_k\rangle = \langle f_1, \ldots, f_k\rangle$ для всех $k = 1, \ldots, n$.*

Доказательство. Индукция по n. Для $n = 1$ утверждение очевидно: достаточно взять $e_1 = f_1$. Пусть утверждение доказано для всех пространств размерности не выше $n-1$, и мы взяли пространство V размерности n. Рассмотрим в нашем пространстве V линейную оболочку векторов $f_1, \ldots, f_{n-1}$: $W = \langle f_1, \ldots, f_{n-1} \rangle$. По предположению индукции найдется ортогональный базис $e_1, \ldots, e_{n-1}$ пространства W такой, что $\langle e_1, \ldots, e_k \rangle = \langle f_1, \ldots, f_k \rangle$ для всех $k = 1, \ldots, n-1$.

Применим лемму 9.5.4 к набору $e_1, \ldots, e_{n-1}$ и вектору f_n. Мы найдем вектор e_n такой, что $e_1, \ldots, e_n$ — ортогональная система векторов, и $\langle e_1, \ldots, e_n \rangle = \langle f_1, \ldots, f_n \rangle = V$, то есть, $e_1, \ldots, e_n$ — базис V. Очевидно, что условие $\langle e_1, \ldots, e_k \rangle = \langle f_1, \ldots, f_k \rangle$ теперь выполняется для всех $k = 1, \ldots, n$. □

Следствие 9.5.6. *В любом [конечномерном] эвклидовом или унитарном пространстве существует ортогональный (и даже ортонормированный) базис.*

Доказательство. Применим следствие 9.5.5 к произвольному базису пространства V. Получим ортогональный базис $e_1, \ldots, e_n$. Положим $e'_i = e_i / ||e_i||$; легко видеть, что $||e'_i|| = 1$ и векторы $e'_1, \ldots, e'_n$ все еще попарно ортогональны. Мы получили ортонормированный базис пространства V. □

Следствие 9.5.7. *Пусть V — эвклидово или унитарное пространства, $W \leqslant V$ — подпространство в V. Любой ортогональный базис подпространства W можно дополнить до ортогонального базиса пространства V.*

Доказательство. Как и в доказательстве следствия 9.5.5, воспользуемся леммой 9.5.4 для индуктивного построения нужного базиса. □

9.6 Ортогональные и унитарные матрицы

Литература: [F], гл. XIII, § 1, п 7; [K2], гл. 3, § 1, п. 5; § 2, п. 4.

В этом разделе мы выясним, что матрица перехода между ортогональными базисами является ортогональной в эвклидовом случае и унитарной в унитарном случае.

Определение 9.6.1. Матрица $C \in M(n, \mathbb{R})$ называется **ортогональной**, если $C \cdot C^T = C^T \cdot C = E$. Матрица $C \in M(n, \mathbb{C})$ называется **унитарной**, если $C \cdot \overline{C}^T = \overline{C}^T \cdot C = E$.

Замечание 9.6.2. Конечно, условия ортогональности и унитарности матрицы записываются единообразно ($C \cdot \overline{C}^T = \overline{C}^T \cdot C = E$), если помнить, что $\overline{C} = C$ для $C \in M(n, \mathbb{R})$.

Лемма 9.6.3. *Для матрицы $C \in M(n, \mathbb{R})$ следующие условия равносильны:*

1. *C ортогональна*
2. *C^T ортогональна*
3. *столбцы C образуют ортонормированный базис в эвклидовом пространстве $\mathbb{R}^n$ со стандартным эвклидовым скалярным произведением (пример 9.1.5).*

4. *строки* C *образуют ортонормированный базис в эвклидовом пространстве* ${}^n\mathbb{R}$ *со стандартным эвклидовым скалярным произведением.*

Лемма 9.6.4. *Для матрицы* $C \in M(n, \mathbb{C})$ *следующие условия равносильны:*

1. C *унитарна*
2. $\overline{C}^T$ *унитарна*
3. *столбцы* C *образуют ортонормированный базис в унитарном пространстве* $\mathbb{C}^n$ *со стандартным эрмитовым скалярным произведением (пример 9.2.4).*
4. *строки* C *образуют ортонормированный базис в унитарном пространстве* ${}^n\mathbb{C}$ *со стандартным эрмитовым скалярным произведением.*

Доказательство. Мы докажем только вариант для унитарной матрицы.

$(1) \Leftrightarrow (2)$ Очевидно из определения.

$(1) \Rightarrow (3)$ Посмотрим на равенство $\overline{C}^T \cdot C = E$. Оно означает, что при умножении i-ой строки матрицы $\overline{C}^T$ на j-й столбец матрицы C мы получим $\delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$. То есть, при стандартном эрмитовом скалярном произведении i-го столбца матрицы C на ее j-й столбец получается δ_{ij}. Это означает, что столбцы матрицы C попарно ортогональны и, кроме того, длина каждого столбца равна 1. В частности, все столбцы ненулевые. По лемме 9.5.2 эти столбцы образуют ортонормированный базис в $\mathbb{C}^n$.

$(3) \Rightarrow (1)$ Мы знаем, что стандартное эрмитово скалярное произведение i-го столбца матрицы C на ее j-й столбец равно δ_{ij}. Но в точности это произведение стоит в позиции (i, j) матрицы $\overline{C}^T \cdot C$; поэтому $\overline{C}^T \cdot C = E$. Заметим, что $1 = \det(E) = \det(\overline{C}^T \cdot C) = \overline{\det}(C) \cdot \det(C)$, поэтому $\det(C)$ отличен от нуля и, стало быть, матрица C обратима. Из равенства $\overline{C}^T \cdot C = E$ теперь следует, что $C^{-1} = \overline{C}^T$, и поэтому $C \cdot \overline{C}^T = E$.

$(2) \Leftrightarrow (4)$ Применим только что доказанную равносильность $(1) \Leftrightarrow (3)$ к матрице C^T; осталось только заметить, что сопряжение не меняет выполнение свойства (3): если $e_1, \ldots, e_n$ — ортонормированный базис унитарного пространства $\mathbb{C}^n$, то и $\overline{e_1}, \ldots, \overline{e_n}$ — ортонормированный базис того же пространства.

□

Теорема 9.6.5. *Пусть* (V, B) *— эвклидово или унитарное пространство. Пусть* $\mathcal{E}$, $\mathcal{F}$ *— ортогональные базисы* V, *и* $C = (\mathcal{E} \rightsquigarrow \mathcal{F})$ *— матрица перехода между ними. Тогда матрица* C *ортогональна в случае эвклидова пространства и унитарна в случае унитарного пространства.*

Доказательство. По теореме 9.4.2 выполнено $G_{\mathcal{F}} = \overline{C}^T \cdot G_{\mathcal{E}} \cdot C$, где $G_{\mathcal{E}}$, $G_{\mathcal{F}}$ — матрицы Грама формы B в базисах $\mathcal{E}$, $\mathcal{F}$ соответственно. Но базисы $\mathcal{E}$, $\mathcal{F}$ ортогональны, поэтому $G_{\mathcal{E}} = G_{\mathcal{F}} = E$. Значит, $E = \overline{C}^T \cdot C$, и матрица C ортогональна в эвклидовом случае и унитарна в унитарном случае. □

9.7 Ортонормированные базисы

Введенное выше понятие ортонормированного базиса чрезвычайно полезно: в этом разделе мы увидим, что использование таких базисов упрощает вычисления.

Лемма 9.7.1. *Пусть* (V, B) *— эвклидово или унитарное пространство,* $e_1, \dots, e_n$ *— ортонормированный базис* V*,* $v \in V$ *— произвольный вектор, и* $v = e_1\alpha_1 + \dots + e_n\alpha_n$ *— его разложение по этому базису. Тогда* $\alpha_i = B(e_i, v)$ *и* $||v||^2 = |\alpha_1|^2 + \dots + |\alpha_n|^2$*.*

Доказательство. Домножим равенство $v = e_1\alpha_1 + \dots + e_n\alpha_n$ скалярно на e_i:

$$B(e_i, v) = B(e_i, e_1\alpha_1 + \dots + e_n\alpha_n).$$

Воспользовавшись линейностью B по второму аргументу и ортонормированностью базиса $e_1, \dots, e_n$, получаем, что $B(e_i, v) = B(e_i, e_i\alpha_i) = \alpha_i$. Заметим, что векторы $e_1\alpha_1, \dots, e_n\alpha_n$ попарно ортогональны и $||e_i\alpha_i|| = |\alpha_i|$. Доказательство завершается индукцией по n с применением теоремы Пифагора. □

Пусть (V, B) — конечномерное эвклидово или унитарное пространство, $u \in V$ — некоторый фиксированный вектор. Рассмотрим отображение $B(u, -)\colon V \to k$, $v \mapsto B(u, v)$. Линейность формы B по второму аргументу означает, что полученное отображение линейно, то есть, лежит в $\mathrm{Hom}_k(V, k)$. Оказывается, верно и обратное: любое линейное отображение из V в основное поле k имеет вид $B(u, -)$ для некоторого вектора $u \in V$.

Заметим, что если фиксированный вектор u поставить на второе место, то мы получим *полулинейное* отображение $B(-, u)\colon V \to k$ (оно обладает свойством аддитивности, а скаляр выносится с сопряжением). Аналогично, любое полулинейное отображение из V в k имеет вид $B(-, u)$ для некоторого вектора $u \in V$.

Теорема 9.7.2 (Теорема Риса). *Пусть* (V, B) *— конечномерное эвклидово или унитарное пространство. Если* $\varphi\colon V \to k$ *— линейное отображение, то существует единственный вектор* $u \in V$ *такой, что* $\varphi(v) = B(u, v)$ *для всех* $v \in V$*. Если* $\varphi\colon V \to k$ *— полулинейное отображение, то существует единственный вектор* $u \in V$ *такой, что* $\varphi(v) = B(v, u)$ *для всех* $v \in V$*.*

Доказательство. Пусть $\varphi\colon V \to k$ — линейное отображение. Выберем некоторый ортонормированный базис $e_1, \dots, e_n$ пространства V. Пусть $v \in V$ — произвольный вектор. Тогда по лемме 9.7.1

$$v = e_1B(e_1, v) + e_2B(e_2, v) + \dots + e_nB(e_n, v).$$

Применяя к этому равенству отображение φ и пользуясь его линейностью, получаем

$$\begin{aligned}\varphi(v) &= \varphi(e_1B(e_1, v) + e_2B(e_2, v) + \dots + e_nB(e_n, v)) \\ &= \varphi(e_1)B(e_1, v) + \varphi(e_2)B(e_2, v) + \dots + \varphi(e_nB(e_n) \\ &= B(e_1\overline{\varphi(e_1)} + e_2\overline{\varphi(e_2)} + \dots + e_n\overline{\varphi(e_n)}, v).\end{aligned}$$

Заметим, что первый аргумент полученного выражения не зависит от v. Положив $u = e_1\overline{\varphi(e_1)} + e_2\overline{\varphi(e_2)} + \cdots + e_n\overline{\varphi(e_n)}$, получаем, что $\varphi(v) = B(u, v)$ для произвольного $v \in V$. Осталось показать, что такой вектор u единственный. Предположим, что нашелся еще один вектор $u' \in V$ такой, что $\varphi(v) = B(u', v)$ для всех $v \in V$. Но тогда $B(u, v) = \varphi(v) = B(u', v)$, откуда $B(u - u', v) = 0$ для всех $v \in V$. В частности, это так для $v = u - u'$, и получаем $B(u - u', u - u') = 0$. Но форма B положительно определена, и потому $u - u' = 0$, то есть, $u = u'$.

Пусть теперь отображение $\varphi\colon V \to k$ полулинейно. Тогда отображение $\overline{\varphi}\colon V \to k$, $v \mapsto \overline{\varphi(v)}$, линейно, и к нему можно применить доказанное выше: существует единственный вектор $u \in V$ такой, что $\overline{\varphi}(v) = B(u, v)$ для всех $u \in V$. Но равенство $\overline{\varphi}(v) = B(u, v)$ равносильно равенству $\varphi(v) = B(v, u)$. □

Замечание 9.7.3. Заметим, что полученное выражение $u = e_1\overline{\varphi(e_1)} + \cdots + e_n\overline{\varphi(e_n)}$ для вектора u с виду зависит от выбора базиса $e_1, \ldots, e_n$. С другой стороны, мы показали, что вектор u с указанными свойствами единственный. Получается, что это выражение на самом деле одинаково во всех базисах пространства V.

9.8 Ортогональное дополнение

Литература: [F], гл. XIII, § 2, п. 2; [K2], гл. 3, § 1, п. 3; § 2, п. 3; [KM], ч. 2, § 3, пп. 1–2.

Определение 9.8.1. Пусть (V, B) — эвклидово или унитарное пространство, $U \subseteq V$ — произвольное подмножество. **Ортогональным дополнением** к подмножеству U в V называется $U^\perp = \{v \in V \mid \forall u \in U \;\; B(u, v) = 0\}$.

Предложение 9.8.2. *Пусть (V, B) — эвклидово или унитарное пространство, $U \subseteq V$ — подмножество в V. Тогда*

1. *$U^\perp$ является подпространством в V;*
2. *$\{0\}^\perp = V$, $V^\perp = \{0\}$;*
3. *$U \cap U^\perp \subseteq \{0\}$;*
4. *если $U \subseteq W$ — два подмножества в V, то $W^\perp \subseteq U^\perp$.*

Доказательство. 1. Если v_1, v_2 лежат в $U^\perp$, то для любого $u \in U$ выполнено $B(u, v_1) = B(u, v_2) = 0$. Поэтому для любых $\lambda_1, \lambda_2 \in k$ выполнено $B(u, v_1\lambda_1 + v_2\lambda_2) = B(u, v_1)\lambda_1 + B(u, v_2)\lambda_2 = 0$, и $v_1\lambda_1 + v_2\lambda_2 \in U^\perp$. Это доказывает, что $U^\perp \leqslant V$.

2. Любой вектор V ортогонален 0, поэтому $\{0\}^\perp = V$. Если вектор $v \in V$ ортогонален всем векторам из V, то, в частности, он ортогонален самому себе, то есть, $B(v, v) = 0$. В силу положительной определенности формы B из этого следует, что $v = 0$. Это доказывает, что $V^\perp = \{0\}$.

3. Пусть $v \in U \cap U^\perp$. Условие $v \in U^\perp$ означает, что $B(u, v) = 0$ для всех $u \in U$, в частности, для $u = v$. Поэтому $B(v, v) = 0$. В силу положительной определенности формы B получаем, что $v = 0$.

4. Пусть $v \in W^\perp$. Тогда $B(u, v) = 0$ для всех $u \in W$. В частности, это так для всех $u \in U$. Поэтому $v \in U^\perp$.

□

Предложение 9.8.3. *Пусть (V, B) — эвклидово или унитарное пространство, $U \leqslant V$ — конечномерное подпространство в V. Тогда*

1. $V = U \oplus U^\perp$;
2. *если, кроме того, V конечномерно, то* $\dim(U^\perp) = \dim(V) - \dim(U)$;
3. $(U \perp)^\perp = U$.

Доказательство. 1. Пусть $e_1, \ldots, e_m$ — некоторый ортонормированный базис подпространства U (такой существует по следствию 9.5.6). Возьмем произвольный вектор $v \in V$, обозначим

$$u = e_1 B(e_1, v) + \cdots + e_m B(e_m, v) \in U,$$

и положим $w = v - u$. Заметим, что $w \in U^\perp$. Действительно,

$$\begin{aligned} B(e_i, w) &= B(e_i, v - u) \\ &= B(e_i, v) - B(e_i, u) \\ &= B(e_i, v) - B(e_i, e_1 B(e_1, v) + \cdots + e_m B(e_m, v)) \\ &= B(e_i, v) - B(e_i, v) \\ &= 0 \end{aligned}$$

(мы воспользовались ортонормированностью базиса $e_1, \ldots, e_m$). Эта выкладка показывает, что w ортогонален каждому из векторов $e_1, \ldots, e_m$; поэтому w ортогонален и любой их линейной комбинации, то есть, любому вектору подпространства U. Итак, мы получили представление $v = u + w$, где $u \in U$, $w \in U^\perp$, для произвольного вектора $v \in V$. Это означает, что $V = U + U^\perp$. В предложении 9.8.2 мы уже показали, что $U \cap U^\perp \subseteq \{0\}$, и в нашем случае $U, U^\perp$ содержат 0, то есть, на самом деле $U \cap U^\perp = \{0\}$. По предложению 6.2.10 из этого следует, что $V = U \oplus U^\perp$.

2. По следствию 6.5.6 и по уже доказанному, имеем $\dim(V) = \dim(U) + \dim(U^\perp)$.

3. Покажем сначала, что $U \subseteq (U^\perp)^\perp$ (на самом деле, это верно даже без условия конечномерности U). Пусть $u \in U$; мы хотим проверить, что $u \in (U^\perp)^\perp$, то есть, что u ортогонален любому вектору из $U^\perp$. Пусть w — произвольный вектор из $U^\perp$. По определению это означает, что он ортогонален любому вектору из U, в частности, вектору u: $B(u, w) = 0$. Но тогда и $B(w, u) = 0$, то есть, u ортогонален w, что и требовалось.

Осталось проверить обратное включение: возьмем произвольный вектор $v \in (U^\perp)^\perp$ и покажем, что $v \in U$. По первому пункту мы можем представить v в виде $v = u + w$, где $u \in U$ и $w \in U^\perp$. Тогда $w = v - u$, и отсюда $B(w, w) = B(w, v - u)$. При этом $w \in U^\perp$, $v \in (U^\perp)^\perp$, и $u \in U \subseteq (U^\perp)^\perp$ (мы пользуемся уже доказанным включением). Значит, скалярное произведение w на $v - u$ равно нулю, откуда $B(w, w) = 0$, откуда следует, что $w = 0$. Поэтому $v = u \in U$, что и требовалось.

□

Определение 9.8.4. Пусть (V, B) — эвклидово или унитарное пространство, $U \leqslant V$ — конечномерное подпространство. Возьмем произвольный вектор $v \in V$. По предложению 9.8.3 существует единственное разложение вида $v = u + u'$, где $u \in U$, $u' \in U^\perp$. Так определенный вектор $u \in U$ мы будем называть **ортогональной проекцией** вектора v на подпространство U и обозначать через $\mathrm{pr}_U(v)$. Мы получили, таким образом, отображение $\mathrm{pr}_U \colon V \to V$, которое каждому вектору $v \in V$ сопоставляет его проекцию на подпространство U (рассмотренную как элемент объемлющего пространства V).

Теорема 9.8.5. *Пусть (V, B) — эвклидово или унитарное пространство, $U \leqslant V$ — конечномерное подпространство, $v \in V$.*

1. *Отображение $\mathrm{pr}_U \colon V \to V$ является линейным.*
2. *Если $v \in U$, то $\mathrm{pr}_U(v) = v$.*
3. *Если $v \in U^\perp$, то $\mathrm{pr}_U(v) = 0$.*
4. $\mathrm{Im}(\mathrm{pr}_U) = U$.
5. $\mathrm{Ker}(\mathrm{pr}_U) = U^\perp$.
6. $v - \mathrm{pr}_U(v) \in U^\perp$.
7. $\mathrm{pr}_U \circ \mathrm{pr}_U = \mathrm{pr}_U$.
8. $||\mathrm{pr}_U(v)|| \leqslant ||v||$.
9. *Если $e_1, \ldots, e_n$ — любой ортонормированный базис U, то* $\mathrm{pr}_U(v) = e_1 B(e_1, v) + \cdots + e_n B(e_n, v)$.

Доказательство. 1. Пусть $v_1, v_2 \in V$, причем $v_1 = u_1 + w_1$ и $v_2 = u_2 + w_2$, где $u_1, u_2 \in U$, $w_1, w_2 \in U^\perp$. Тогда $v_1 + v_2 = (u_1 + u_2) + (w_1 + w_2)$, и $u_1 + u_2 \in U$, $w_1 + w_2 \in U^\perp$. По определению $\mathrm{pr}_U(v_1) = u_1$, $\mathrm{pr}_U(v_2) = u_2$ и $\mathrm{pr}_U(v_1 + v_2) = u_1 + u_2 = \mathrm{pr}_U(v_1) + \mathrm{pr}_U(v_2)$. Мы показали аддитивность отображения pr_U. Если $v \in U$ и $v = u + w$ для $u \in U$, $w \in U^\perp$, то $v\lambda = u\lambda + w\lambda$, откуда следует и однородность pr_U.

2. Если $v \in U$, то $v = v + 0$, где $v \in U$, $0 \in U^\perp$.

3. Если $v \in U^\perp$, то $v = 0 + v$, где $0 \in U$, $v \in U^\perp$.

4. В пункте (2) мы показали, что $U \subseteq \mathrm{Im}(\mathrm{pr}_U)$. Обратное включение выполнено по определению отображения pr_U.

5. В пункте (3) мы показали, что $U^\perp \subseteq \mathrm{Ker}(\mathrm{pr}_U)$. Обратно, если $\mathrm{pr}_U(v) = 0$, то $v = 0 + w$, где $w \in U^\perp$.

6. По определению $v = u + w$, где $u \in U$, $w \in U^\perp$ и $u = \mathrm{pr}_U(v)$. Поэтому $v - \mathrm{pr}_U(v) = v - u = w \in U^\perp$.

7. Пусть $\mathrm{pr}_U(v) = u \in U$. Тогда $\mathrm{pr}_U(u) = u$ по пункту (2), что и требовалось.

8. $v = \mathrm{pr}_U(v) + w$, где $w \in U^\perp$, и потому векторы $\mathrm{pr}_U(v)$ и w ортогональны. По теореме Пифагора $||v||^2 = ||\mathrm{pr}_U(v)||^2 + ||w||^2$, откуда следует нужное неравенство.

9. Запишем $v = u + (v - u)$, где $u = e_1 B(e_1, v) + \cdots + e_n B(e_n, v)$. Как и в доказательстве пункта (1) предложения 9.8.3, получаем, что $v - u$ ортогонально каждому из $e_1, \ldots, e_n$, и потому $v - u \in U^\perp$, в то время как, очевидно, $u \in U$. По определению тогда $\mathrm{pr}_U(v) = u$, что и требовалось.

□

9.9 Сопряженные отображения

Литература: [F], гл. XIII, § 4, п. 2; [K2], гл. 3, § 3, п. 1; [KM], ч. 2, § 8, пп. 1–3.

Определение 9.9.1. Пусть (V, B) и (V', B') — эвклидовы или унитарные пространства, $\varphi\colon V \to V'$ — линейное отображение. Линейное отображение $\varphi^*\colon V' \to V$ называется **сопряженным** к отображению φ, если $B'(\varphi(v), v') = B(v, \varphi^*(v'))$ для всех векторов $v \in V$ и $v' \in V'$.

Покажем, что у каждого линейного отображения между эвклидовыми или унитарными пространствами имеется единственное сопряженное.

Предложение 9.9.2. *Пусть (V, B) и (V', B') — эвклидовы или унитарные пространства, $\varphi\colon V \to V'$ — линейное отображение. Существует линейное отображение $\varphi^*\colon V' \to V$ сопряженное к φ. Кроме того, такое линейное отображение единственно.*

Доказательство. Пусть $v' \in V'$. Рассмотрим отображение $f\colon V \to k$, которое сопоставляет вектору $v \in V$ скаляр $B'(\varphi(v), v')$. Покажем, что f — полулинейное отображение. Действительно, $f(v_1\lambda_1 + v_2\lambda_2) = B'(\varphi(v_1\lambda_1 + v_2\lambda_2), v') = B'(\varphi(v_1)\lambda_1 + \varphi(v_2)\lambda_2, v') = \overline{\lambda_1}B'(\varphi(v_1), v') + \overline{\lambda_2}B'(\varphi(v_2), v') = \overline{\lambda_1}f(v_1) + \overline{\lambda_2}f(v_2)$. По теореме Риса 9.7.2 найдется вектор $v_f \in V$ такой, что $B(v, v_f) = f(v) = B'(\varphi(v), v')$ для всех $v \in V$. Положим $\varphi^*(v') = v_f$.

Таким образом, для каждого $v' \in V'$ мы нашли вектор $\varphi^*(v') \in V$ такой, что $B(v, \varphi^*(v')) = B'(\varphi(v), v')$ для всех $v \in V$. Проверим, что полученное отображение $\varphi^*\colon V' \to V$ является линейным. Действительно.

$$\begin{aligned} B(v, \varphi^*(v'_1)\lambda_1 + \varphi^*(v'_2)\lambda_2) &= B(v, \varphi^*(v'_1))\lambda_1 + B(v, \varphi^*(v'_2))\lambda_2 \\ &= B'(\varphi(v), v'_1)\lambda_1 + B'(\varphi(v), v'_2))\lambda_2 \\ &= B'(\varphi(v), v'_1\lambda_1 + v'_2\lambda_2). \end{aligned}$$

С другой стороны, по определению φ^* выполнено $B(v, \varphi^*(v_1'\lambda_1 + v_2'\lambda_2)) = B'(\varphi(v), v_1'\lambda_1 + v_2'\lambda_2)$. Поэтому $B(v, \varphi^*(v_1'\lambda_1 + v_2'\lambda_2)) = B(v, \varphi^*(v_1')\lambda_1 - \varphi^*(v_2')\lambda_2)$ для всех $v \in V$, откуда следует, что $\varphi^*(v_1'\lambda_1 + v_2'\lambda_2) = \varphi^*(v_1')\lambda_1 - \varphi^*(v_2')\lambda_2$.

Осталось показать единственность отображения φ^* с указанным свойством. Но если $\widetilde{\varphi^*}$ — другое такое отображение, то $B(v, \varphi^*(v')) = B'(\varphi(v), v') = B(v, \widetilde{\varphi^*}(v'))$ для всех $v \in V$, $v' \in V'$. Из этого следует, что $\varphi^*(v') = \widetilde{\varphi^*}(v')$ для каждого v'. □

Предложение 9.9.3. *Пусть* (V, B) *и* (V', B') *— эвклидовы или унитарные пространства,* $\varphi, \psi\colon V \to V'$ *— линейные отображения,* $\lambda \in k$. *Тогда*

1. $(\varphi + \psi)^* = \varphi^* + \psi^*$;
2. $(\lambda\varphi)^* = \overline{\lambda}\varphi^*$;
3. $(\varphi^*)^* = \varphi$;
4. $(\mathrm{id}_V)^* = \mathrm{id}_V$;
5. *если* $\eta\colon V' \to V''$ *— еще одно линейное отображение (где* (V'', B'') *— эвклидово или унитарное пространство), то* $(\eta \circ \varphi)^* = \varphi^* \circ \eta^*$

Доказательство. 1. Пусть $v \in V$, $v' \in V'$. Тогда

$$\begin{aligned} B(v, (\varphi + \psi)^*(v')) &= B'((\varphi + \psi)(v), v') \\ &= B'(\varphi(v) + \psi(v), v') \\ &= B'(\varphi(v), v') + B(\psi(v), v') \\ &= B(v, \varphi^*(v')) + B(v, \psi^*(v')) \\ &= B(v, \varphi^*(v') + \psi^*(v')), \end{aligned}$$

откуда следует, что $(\varphi + \psi)^*(v') = \varphi^*(v') + \psi^*(v')$, что и требовалось.

2. Пусть $v \in V$, $v' \in V'$. Тогда

$$B(v, (\lambda\varphi)^*(v')) = B'(\lambda\varphi(v), v') = \overline{\lambda}B'(\varphi(v), v') = \overline{\lambda}B(v, \varphi^*(v')) = B(v, \overline{\lambda}\varphi^*(v')),$$

откуда $(\lambda\varphi)^*(v') = \overline{\lambda}\varphi^*(v')$, что и требовалось.

3. Пусть $v \in V$, $v' \in V'$. Тогда

$$B'(v', ((\varphi^*)^*(v)) = B(\varphi^*(v'), v) = \overline{B(v, \varphi^*(v'))} = \overline{B'(\varphi(v), v')} = B'(v', \varphi(v)),$$

откуда $((\varphi^*)^*(v) = \varphi(v)$, что и требовалось.

4. Пусть $v, w \in V$. Тогда

$$B(v, (\mathrm{id}_V)^*(w)) = B(\mathrm{id}_V(v), w) = B(v, w) = B(v, \mathrm{id}_V(w)),$$

откуда $(\mathrm{id}_V)^*(w) = \mathrm{id}_V(w)$, что и требовалось.

5. Пусть $v \in V$, $v'' \in V''$. Тогда

$$\begin{aligned} B(v, (\eta \circ \varphi)^*(v'')) &= B''((\eta \circ \varphi)(v), v'') \\ &= B''(\eta(\varphi(v)), v'') \\ &= B'(\varphi(v), \eta^*(v'')) \\ &= B(v, \varphi^*(\eta^*(v''))) \\ &= B(v, (\varphi^* \circ \eta^*)(v'')), \end{aligned}$$

откуда $(\eta \circ \varphi)^*(v'') = (\varphi^* \circ \eta^*)(v'')$, что и требовалось.

□

Выясним, как выглядит матрица сопряженного отображения в ортонормированных базисах.

Предложение 9.9.4. *Пусть (V, B), (V', B') — эвклидовы или унитарные пространства, $\mathcal{E}$ — ортонормированный базис пространства V, $\mathcal{E}'$ — ортонормированный базис пространства V'. Для любого линейного отображения $\varphi\colon V \to V'$ выполнено $[\varphi^*]_{\mathcal{E}',\mathcal{E}} = \overline{[\varphi]_{\mathcal{E},\mathcal{E}'}}^T$.*

Доказательство. Обозначим $A = [\varphi]_{\mathcal{E},\mathcal{E}'}$, $A^* = [\varphi^*]_{\mathcal{E}',\mathcal{E}}$. По основному свойству матрицы линейного отображения (теорема 7.4.5) для любых векторов $v \in V$, $v' \in V'$ выполнено $A \cdot [v]_{\mathcal{E}} = [\varphi(v)]_{\mathcal{E}'}$ и $A^* \cdot [v']_{\mathcal{E}'} = [\varphi^*(v')]_{\mathcal{E}}$. Матрицы Грама форм B и B' единичны, поэтому

$$\overline{[\varphi(v)]_{\mathcal{E}'}}^T \cdot [v']_{\mathcal{E}'} = B'(\varphi(v), v') = B(v, \varphi^*(v')) = \overline{[v]_{\mathcal{E}}}^T \cdot [\varphi^*(v')]_{\mathcal{E}}.$$

Подставляя сюда выражения для столбцов координат $\varphi(v)$ и $\varphi^*(v')$, получаем

$$\overline{A \cdot [v]_{\mathcal{E}}}^T \cdot [v']_{\mathcal{E}'} = \overline{[v]_{\mathcal{E}}}^T \cdot A^* \cdot [v']_{\mathcal{E}'},$$

откуда

$$\overline{[v]_{\mathcal{E}}}^T \cdot \overline{A}^T \cdot [v']_{\mathcal{E}'} = \overline{[v]_{\mathcal{E}}}^T \cdot A^* \cdot [v']_{\mathcal{E}'}.$$

Это равенство верно для всех $v \in V$, $v' \in V'$. Пусть теперь v пробегает все векторы базиса $\mathcal{E}$, а v' пробегает все векторы базиса $\mathcal{E}'$. Получаем равенство матриц $A^* = \overline{A}^T$. □

9.10 Самосопряженные операторы

Определение 9.10.1. Пусть (V, B) — эвклидово или унитарное пространство. Линейный оператор $T\colon V \to V$ называется **самосопряженным**, если $T^* = T$. Иными словами, T самосопряжен, если $B(T(v), w) = B(v, T(w))$ для всех $v, w \in V$.

Предложение 9.10.2. *Все собственные числа самосопряженного оператора вещественны.*

Доказательство. Пусть $T\colon V \to V$ — самосопряженный оператор, $\lambda \in k$ — собственное число оператора T, и $v \in V$ — соответствующий ему собственный вектор, то есть, $T(v) = v\lambda$ и $v \neq 0$. Тогда

$$\lambda ||v||^2 = \lambda B(v, v) = B(v, v\lambda) = B(v, T^*(v)) = B(T(v), v) = B(v\lambda, v) = \overline{\lambda} B(v, v) = \overline{\lambda} ||v||^2$$

При этом $||v||^2 \neq 0$, и потому $\lambda = \overline{\lambda}$. □

Следующие две леммы верны только для унитарных пространств, но не для эвклидовых (см. замечание 9.10.4).

Лемма 9.10.3. *Пусть V — унитарное пространство (внимание!), $T\colon V \to V$ — линейный оператора. Предположим, что $B(T(v), v) = 0$ для всех $v \in V$. Тогда $T = 0$.*

Доказательство. Пусть $u, v \in V$. Заметим, что

$$B(T(u), v) = \frac{B(T(u+v), u+v) - B(T(u-v), u-v) - iB(T(u+vi), u+vi) + iB(T(u-vi), u-vi)}{4}$$

(это можно проверить прямым вычислением). В правой части стоят выражения вида $B(T(w), w)$, которые по предположению равны нулю. Значит, $B(T(u), v) = 0$. В частности, это так для $v = T(u)$; получаем, что $T(u) = 0$ для всех $u \in V$, откуда $T = 0$. □

Замечание 9.10.4. Заметим, что лемма 9.10.3 неверна для эвклидовых пространств: линейный оператор $\mathbb{R}^2 \to \mathbb{R}^2$, осуществляющий поворот на $\pi/2$, служит контрпримером.

Лемма 9.10.5. *Пусть V — унитарное пространство (внимание!), $T\colon V \to V$ — линейный оператор. Оператор T самосопряжен тогда и только тогда, когда скалярное произведение $B(T(v), v)$ вещественно для всех $v \in V$.*

Доказательство. Пусть $v \in V$. Тогда

$$B(T(v), v) - \overline{B(T(v), v)} = B(T(v), v) - B(v, T(v)) = B(T(v), v) - B(T^*(v), v) = B((T - T^*)(v), v).$$

Если $B(T(v), v) \in \mathbb{R}$ для всех $v \in V$, то правая часть всегда равна нулю, и по лемме 9.10.3 из этого следует, что $T - T^* = 0$.

Обратно, если $T = T^*$, то правая часть всегда равна нулю, и потому $B(T(v), v) = \overline{B(T(v), v)}$ для всех $v \in V$, откуда $B(T(v), v) \in \mathbb{R}$. □

Замечание 9.10.6. Замечание 9.10.4 показывает, что на эвклидовом пространстве оператор T может удовлетворять тождеству $B(T(v), v) = 0$ для всех $v \in V$. Однако, этого не может случиться для самосопряженного оператора.

Лемма 9.10.7. *Пусть (V, B) — эвклидово или унитарное пространство, $T\colon V \to V$ — самосопряженный оператор. Если $B(T(v), v) = 0$ для всех $v \in V$, то $T = 0$.*

Доказательство. Для унитарного пространства это уже доказано в лемме 9.10.3. Если же V эвклидово, то

$$B(T(u), v) = \frac{B(T(u+v), u+v) - B(T(u-v), u-v)}{4}$$

для всех $u, v \in V$, что проверяется прямым вычислением с использованием равенств $B(T(v), u) = B(v, T(u)) = B(T(u), v)$ (здесь мы используем самосопряженность T). По предположению правая часть равна нулю, поэтому $B(T(u), v) = 0$ для всех $u, v \in V$; в частности, это так для $v = T(u)$, откуда следует, что $T = 0$. □

9.11 Нормальные операторы

Литература: [F], гл. XIII, § 4, п. 3; [K2], гл. 3, § 3, п. 7; [KM], ч. 2, § 8, п. 11.

Определение 9.11.1. Пусть (V, B) — эвклидово или унитарное пространство. Линейный оператор $T\colon V \to V$ называется **нормальным**, если он коммутирует со своим сопряженным: $T^* \circ T = T \circ T^*$.

Замечание 9.11.2. Очевидно, что любой самосопряженный оператор нормален.

Лемма 9.11.3 (Свойства нормальных операторов). *1. Тождественный оператор нормален.*

2. Сопряженный к нормальному оператору нормален.

Доказательство. Очевидно. □

Лемма 9.11.4. *Пусть (V, B) — эвклидово или унитарное пространство. Оператор $T\colon V \to V$ нормален тогда и только тогда, когда $||T(v)|| = ||T^*(v)||$ для всех $v \in V$.*

Доказательство. Заметим, что оператор $T^* \circ T - T \circ T^*$ самосопряжен. По лемме 9.10.7 равенство $T^* \circ T - T \circ T^*$ равносильно тому, что $B((T^* \circ T - T \circ T^*)(v), v) = 0$ для всех $v \in V$, что равносильно равенству $B(T^*(T(v)), v) = B(T(T^*(v)), v)$ для всех $v \in V$. Но $B(T^*(T(v)), v) = ||T(v)||^2$ и $B(T(T^*(v)), v) = ||T^*(v)||^2$. □

Предложение 9.11.5. *Пусть (V, B) — эвклидово или унитарное пространство, $T\colon V \to V$ — нормальный оператор, и $v \in V$ — собственный вектор оператора T, соответствующий собственному числу λ. Тогда v является и собственным вектором оператора T^*, соответствующим собственному числу $\overline{\lambda}$.*

Доказательство. Из нормальности T следует, что и оператор $T - \lambda \mathrm{id}_V$ нормален (проверьте это!). По лемме 9.11.4 тогда $||(T - \lambda \mathrm{id}_V)(v)|| = ||(T - \lambda \mathrm{id}_V)^*(v)||$. Но левая часть по предположению равна нулю, а правая часть равна $||(T^* - \overline{\lambda} \mathrm{id}_V)(v)||$. □

Предложение 9.11.6. *Пусть (V, B) — эвклидово или унитарное пространство, $T\colon V \to V$ — нормальный оператор. Тогда собственные векторы T, соответствующие различным собственным числам, ортогональны.*

Доказательство. Пусть $\lambda \neq \mu$ — два различных собственных числа оператора T, и пусть $u, v \in V$ — соответствующие им собственные векторы: $T(u) = u\lambda$, $T(v) = v\mu$. По предложению 9.11.5 теперь $T^*(u) = u\overline{\lambda}$. Поэтому $(\lambda - \mu)B(u, v) = B(u\overline{\lambda}, v) - B(u, v\mu) = B(T(u), v) - B(u, T^*(v)) = 0$. Поскольку $\lambda \neq \mu$, из этого равенства следует, что $B(u, v) = 0$, что и требовалось. □

9.12 Спектральные теоремы

Литература: [F], гл. XIII, § 5; [K2], гл. 3, § 3, пп. 3, 6; [KM], ч. 2, § 7, пп. 4–5; § 8, пп. 2–6, 8.

Теорема 9.12.1 (Спектральная теорема для нормальных операторов в унитарном пространстве). *Пусть* (V, B) *— унитарное пространство,* $T\colon V \to V$ *— линейный оператор. Следующие условия равносильны:*

1. *оператор* T *нормален;*
2. *у* V *есть ортонормированный базис, состоящий из собственных векторов оператора* T*;*
3. *матрица оператора* T *в некотором ортонормированном базисе* V *диагональна.*

Доказательство. Очевидно, что (2) $\Leftrightarrow$ (3) (см. также доказательство теоремы 8.3.2). Покажем, что из (3) следует (1). Пусть матрица t в некотором ортонормированном базисе $\mathcal{B}$ диагональна. По предложению 9.9.4 матрица T^* тогда получается из матрицы T транспонированием и сопряжением, и потому тоже диагональна. Но любые две диагональные матрицы коммутируют; поэтому T коммутирует с T^*, то есть, T нормален.

Пусть теперь выполняется (1): оператор T нормален. По теореме о жордановой форме 8.7.3 существует базис $\mathcal{B} = (v_1, \dots, v_n)$ пространства V, в котором матрица T верхнетреугольна. Применим к этому базису процесс ортогонализации Грама–Шмидта: мы получим ортонормированный базис $\mathcal{E} = (e_1, \dots, e_n)$. По предложению 8.2.2 верхнетреугольность матрицы T в базисе $\mathcal{B}$ равносильна тому, что все подпространства вида $\langle v_1, \dots, v_i \rangle$ являются T-инвариантными. Но в процессе ортогонализации мы получили базис, для которого $\langle e_1, \dots, e_i \rangle = \langle v_1, \dots, v_i \rangle$, а инвариантность этих подпространств равносильна верхнетреугольности матрицы T в ортонормированном базисе $\mathcal{E}$.

Итак, матрица оператора T в базисе $\mathcal{E}$ верхнетреугольна:

$$[T]_{\mathcal{E}} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ 0 & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{nn} \end{pmatrix}$$

Покажем, что она на самом деле не только верхнетреугольна, но и диагональна. Мы знаем, что матрица оператора T^* в том же базисе выглядит так:

$$[T^*]_{\mathcal{E}} = \overline{[T]_{\mathcal{E}}}^{\mathsf{T}} \begin{pmatrix} \overline{a_{11}} & 0 & \dots & 0 \\ \overline{a_{12}} & \overline{a_{22}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \overline{a_{1n}} & \overline{a_{2n}} & \dots & \overline{a_{nn}} \end{pmatrix}$$

Самое время воспользоваться нормальностью оператора T. Посмотрим внимательно, что стоит в левом верхнем углу матриц, полученных перемножением $[T]_{\mathcal{E}}$ и $[T^*]_{\mathcal{E}}$. Нетрудно видеть, что у

матрицы $[T^*]\cdot[T]$ в позиции $(1,1)$ стоит $|a_{11}|^2$, а у матрицы $[T]\cdot[T^*]$ — $|a_{11}|^2+|a_{12}|^2+\cdots+|a_{1n}|^2$, сумма квадратов модулей элементов первой строки матрицы $[T]$. Но эти выражения должны быть равны, и все входящие в них слагаемые — неотрицательные вещественные числа. Поэтому $a_{12}=\cdots=a_{1n}=0$. Значит, в первой строке матрицы $[T]$ на самом деле только один ненулевой элемент: диагональны. Вооружившись этим знанием, проследим теперь за позицией $(2,2)$. Перемножая матрицы в одном порядке, получаем $|a_{22}|^2$, а в другом — сумму квадратов элементов второй строки матрицы $[T]$. Из этого следует, что и во второй строке матрица $[T]$ не отличается от диагональной. Продолжая этот процесс, получаем, что $[T]_\varepsilon$ диагональна, что и требовалось. □

Теперь обратимся к случаю эвклидового пространства. Как мы знаем, жорданова форма для оператора на вещественном пространстве уже не обязана быть верхнетреугольной, поэтому для переноса спектральной теоремы на эвклидов случай придется действовать обходным путем. Сначала мы разберемся с самосопряженными операторами. Для этого нам понадобится следующая лемма, в основе которой лежит несложное вычисление, известное вам со школы:

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

Лемма 9.12.2. *Пусть $T\colon V\to V$ — самосопряженный линейный оператор на эвклидовом или унитарном пространстве V, и числа $b,c\in\mathbb{R}$ таковы, что $b^2-4c<0$. Тогда оператор $T^2+bT+c\,\mathrm{id}_V$ обратим.*

Доказательство. Пусть $v\in V$. Тогда

$$\begin{aligned}B((T^2+bT+c\,\mathrm{id}_V)(v),v)&=B(T^2(v),v)+bB(T(v),v)+cB(v,v)\\&=B(T(v),T(v))+bB(T(v),v)+c||v||^2\\&\geqslant||T(v)||^2-|b|\cdot||T(v)||\cdot||v||+c||v||^2\end{aligned}$$

в силу неравенства Коши–Буняковского–Шварца: $-||T(v)||\cdot||v||\leqslant B(T(v),v)\leqslant||T(v)||\cdot||v||$. Полученное выражение можно переписать так:

$$\left(||T(v)||-\frac{|b|\cdot||v||}{2}\right)^2+\left(c-\frac{b^2}{4}\right)||v||^2,$$

и видно, что оно (при нашем условии на b и c) неотрицательно. Поэтому оператор $T^2+bT+c\,\mathrm{id}$ инъективен, значит, и биективен. □

Замечание 9.12.3. Мы знаем, что у любого оператора на комплексном пространстве есть собственное число. Поэтому следующую лемму достаточно доказать только для случая эвклидово пространств.

Лемма 9.12.4. *Пусть $V\neq\{0\}$ — эвклидово пространство, $T\colon V\to V$ — самосопряженный линейный оператор. Тогда у T есть собственное число.*

Доказательство. Пусть $\dim(V) = n$. Рассмотрим минимальный многочлен оператора T:

$$f = a_0 + a_1 x + \cdots + a_n x^n \in k[x]$$

(см. определение 8.5.7). По теореме 4.4.4 его можно разложить на множители вида

$$f = c(x^2 + b_1 x + c_1) \dots (x^2 + b_M x c_M)(x - \lambda_1) \dots (x - \lambda_m),$$

где $c \neq 0$, b_j, c_j, λ_j — вещественные числа, причем $b_j^2 - 4c_j < 0$. Поэтому

$$0 = f(T)(v) = c(T^2 + b_1 T + c_1 \operatorname{id}) \dots (T^2 + b_M T + c_M \operatorname{id})(T - \lambda_1 \operatorname{id}) \dots (T - \lambda_m \operatorname{id})(v).$$

По лемме 9.12.2 множители вида $T^2 + b_j T + c_j \operatorname{id}$ обратимы. Поэтому

$$0 = (T - \lambda_1 \operatorname{id}) \dots (T - \lambda_m \operatorname{id})(v).$$

Значит, хотя бы один из операторов $T - \lambda_j \operatorname{id}$ неинъективен. Это и означает, что у T есть собственное число. □

Замечание 9.12.5. Позже мы увидим (см. 9.12.9), что в следующем предложении можно заменить условие самосопряженности оператора на условие нормальности.

Предложение 9.12.6. *Пусть $T\colon V \to V$ — самосопряженный оператор на эвклидовом или унитарном пространстве, и пусть $U \leqslant V$ — T-инвариантное подпространство. Тогда*

1. *подпространство $U^\perp$ также T-инвариантно;*
2. *оператор $T|_U$ самосопряжен;*
3. *оператор $T|_{U^\perp}$ самосопряжен.*

Доказательство. 1. Пусть $v \in U^\perp$. Нам хочется показать, что $T(v) \in U^\perp$. Возьмем любой вектор $u \in U$ и посмотрим на $B(T(v), u)$. Из самосопряженности T следует, что $B(T(v), u) = B(v, T(u))$. Но по условию $T(u) \in U$, значит, мы получили 0.

2. Если $u, v \in U$, то $B((T|_U)(u), v) = B(T(u), v) = B(u, T(v)) = B(u, (T|_U)(v))$.

3. Применим результат второго пункта к $U^\perp$ вместо U.

□

Теорема 9.12.7 (Спектральная теорема для самосопряженных операторов в эвклидовых пространствах). *Пусть (V, B) — эвклидово пространство, $T\colon V \to V$ — линейный оператор. Следующие условия равносильны:*

1. *оператор T самосопряжен;*
2. *у V есть ортонормированный базис, состоящий из собственных векторов оператора T;*

3. матрица оператора T *в некотором ортонормированном базисе* V *диагональна.*

Доказательство. Мы уже знаем, что $(2) \Leftrightarrow (3)$. Предположим, что выполняется (3): матрица оператора T в некотором базисе диагональна. Но диагональная матрица совпадает со своей транспонированной, поэтому $T = T^*$, откуда следует (1).

Теперь мы докажем. что из (1) следует (2) индукцией по размености пространства V. Если $\dim(V) = 1$, утверждение очевидно. Пусть теперь $\dim(V) > 1$, и оператора T самосопряжен. По лемме 9.12.4 у T есть собственное число и, стало быть, собственный вектор u. Поделив его на $||u||$, можно считать, что $||u|| = 1$. Подпространство $U = \langle u \rangle$ тогда является T-инвариантным, и по предложению 9.12.6 подпространство $U^\perp$ тоже T-инвариантно, и оператор $T|_{U^\perp}$ самосопряжен. По предположению индукции у $U^\perp$ есть ортонормальный базис, состоящий из собственных векторов оператора $T|_{U^\perp}$. Присоединив к нему u, получаем ортонормальный базис $U^\perp$, состоящий из собственных векторов оператора T. □

Теперь мы готовы описать нормальные операторы на двумерных эвклидовых пространствах.

Предложение 9.12.8. *Пусть* V *— эвклидово пространство размерности* 2, $T\colon V \to V$ *— линейный оператор. Следующие условия равносильны:*

1. T *нормален, но не самосопряжен;*
2. *матрица* T *в любом ортонормальном базисе* V *имеет вид*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

   где $b \neq 0$*;*

3. *матрица* T *в некотором ортонормальном базисе* V *имеет вид*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

   где $b > 0$*.*

Доказательство. $(1) \Rightarrow (2)$. Пусть e_1, e_2 — ортонормальный базис пространства V, и пусть матрица T в этом базисе имеет вид

$$\begin{pmatrix} a & c \\ b & d \end{pmatrix}.$$

Тогда $||T(e_1)||^2 = a^2 + b^2$, $||T^*(e_1)||^2 = a^2 + c^2$. По предложению 9.11.4 эти числа равны, откуда $c = \pm b$. Если $c = b$, то T самосопряжен (его матрица симметричны), поэтому $c = -b$, при этом $b \neq 0$. Перемножим теперь матрицы T и $T^* = T^{\mathsf{T}}$ в одном и в другом порядке. Результаты должны совпасть, но в правом верхнем углу у одной матрицы стоит bd, а у другой ab. Значит, $a = d$, и мы получили матрицу нужного вида.

(2) $\Rightarrow$ (3). Если в нашем базисе уже $b > 0$, то все доказано, а если нет — поменяем знак у второго базисного вектора.

(3) $\Rightarrow$ (1). Если T имеет указанный вид, то видно, что T не самосопряжен. Перемножая матрицы T и T^* видим, что T нормален. □

Предложение 9.12.9. *Пусть (V, B) — эвклидово или унитарное пространство, $T\colon V \to V$ — нормальный оператор, $U \leqslant V$ — T-инвариантное подпространство. Тогда*

1. *подпространство $U^\perp$ тоже T-инвариантно;*
2. *подпространство U T^*-инвариантно;*
3. *$(T|_U)^* = (T^*)|_U$;*
4. *операторы $T|_U$ и $T|_{U^\perp}$ нормальны.*

Доказательство. Пусть $e_1, \ldots, e_m$ — какой-нибудь ортонормированный базис U. Дополним его до ортонормированного базиса $\mathcal{B}$ пространства V векторами $f_1, \ldots, f_n$. Матрица оператора T имеет в этом базисе следующий вид:

$$[T]_{\mathcal{B}} = \begin{pmatrix} A & B \\ 0 & C \end{pmatrix},$$

где A — блок размера $m \times m$, а C — блок размера $n \times n$. Нетрудно понять, что $||T(e_j)||^2$ равняется сумме квадратов модулей элементов j-го столбца матрицы A. Складывая по всем j, получаем, что $\sum_j ||T(e_j)||^2$ равна сумме квадратов модулей всех элементов матрицы A. С другой стороны, $||T^*(e_j)||^2$ равна сумме квадратов модулей элементов j-й строки матрицы A и j-й строки матрицы B. Складывая по всем j, получаем, что $\sum_j ||T^*(e_j)||^2$ равна сумме квадратов модулей всех элементов матрицы A и всех элементов матрицы B. Из равенства $||T(e_j)|| = ||T^*(e_j)||$ (предложение 9.11.4) теперь следует, что B — нулевая матрица. Теперь из вида матрицы оператора T можно заключить, что $U^\perp$ T-инвариантно. Написав матрицу оператора T^*, можно заметить, что U еще и T^*-инвариантно.

Докажем (3). Пусть $S = T|_U\colon U \to U$. Возьмем $v \in U$. Тогда $B(u, S^*(v)) = B(S(u), v) = B(T(u), v) = B(u, T^*(v)$ для всех $u \in U$. Мы уже знаем, что $T^*(v) \in U$, поэтому из приведенного равенства следует, что $S^*(v) = T^*(v)$. Это выполнено для всех $v \in U$, потому $(T|_U)^* = (T^*)|_U$.

Наконец, для доказательства (4) можно заметить, что T коммутирует с T^*, и потому $T|_U$ коммутирует с $(T|_U)^* = (T^*)|_U$; подставляя $U^\perp$ вместо U, видим, что и $T|_{U^\perp}$ нормален. □

Теорема 9.12.10 (Спектральная теорема для нормальных операторов в эвклидовом пространстве). *Пусть (V, B) — эвклидово пространство, и пусть $T\colon V \to V$ — линейный оператор. Следующие условия равносильны:*

1. *оператор T нормален;*

2. *существует ортонормированный базис пространства* V, *в котором матрица оператора* T *блочно-диагональна, причем каждый блок имеет либо размер* 1×1, *либо размер* 2×2 *и вид*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

где $b > 0$.

Доказательство. $(2) \Rightarrow (1)$: несложно проверить, что матрица такого вида коммутирует со своей сопряженной.

Докажем $(1) \Rightarrow (2)$ индукцией по размерности V. Случай $\dim(V) = 1$ тривиален, а случай $\dim(V) = 2$ следует из спектральной теоремы 9.12.7 для самосопряженного оператора, и из предложения 9.12.8 для остальных.

Пусть теперь $\dim(V) > 2$. Если у оператора T есть одномерное инвариантное подпространство (иными словами, есть собственное число), обозначим его через U. Если же нет, то по предложению 8.8.6 у него есть двумерное инвариантное подпространство, и тогда мы обозначим его через U. Если $\dim(U) = 1$, выберем в U вектор нормы 1 — это будет ортонормированным базисом подпространства U; если же $\dim(U) = 2$, то оператор $T|_U$ нормален (по предложению 9.12.9), но не самосопряжен (иначе у $T|_U$ было бы собственное число по лемме 9.12.4), и в этом случае можно применить предложение 9.12.8.

В любом случае, мы нашли ортонормированный базис в инвариантном подпространстве U, причем подпространство $U^\perp$ T-инвариантно, и оператор $T|_{U^\perp}$ нормален (по предожению 9.12.9). По предположению индукции у $U^\perp$ есть ортонормированный базис с нужными свойствами; приписывая к нему выбранный базис U, получаем нужный базис всего пространства V. □

9.13 Самосопряженные, кососимметрические, унитарные, ортогональные операторы

Литература: [F], гл. XIII, § 5; [K2], гл. 3, § 3, пп. 3, 6; [KM], ч. 2, § 7, пп. 1–2, 4; § 8, пп. 2–6.

Сейчас мы применим знания, полученные при изучении нормальных операторов, к некоторым частным случаям.

Определение 9.13.1. Пусть (V, B) — эвклидово или унитарное пространство, $a\colon V \to V$ — линейный оператор. Оператор a называется **самосопряженным**, если он совпадает со своим сопряженным: $a = a^*$. Оператор a называется **кососимметрическим**, если он противоположен своему сопряженному: $a = -a^*$. Если выполняется равенство $a \circ a^* = a^* \circ a = \mathrm{id}_V$, то оператор a называется **унитарным** в случае унитарного пространства и **ортогональным** в случае эвклидового пространства.

Замечание 9.13.2. Нетрудно видеть, что самосопряженные, кососимметрические, унитарные, ортогональные операторы являются нормальными.

Теорема 9.13.3. *Пусть* (V, B) — *конечномерное унитарное пространство,* $a\colon V \to V$ — *линейный оператор.*

1. *Оператор* $\mathfrak{a}$ *является самосопряженным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* $\mathfrak{a}$ *диагональна, и все ее диагональные элементы вещественны.*

2. *Оператор* $\mathfrak{a}$ *является кососимметрическим тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* $\mathfrak{a}$ *диагональна, и все ее диагональные элементы — чисто мнимые комплексные числа.*

3. *Оператор* $\mathfrak{a}$ *является унитарным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* $\mathfrak{a}$ *диагональна, и все ее диагональные элементы — комплексные числа, равные по модулю* 1*.*

Доказательство. Если оператор самосопряженный, кососимметрический, нормальный, то по теореме 9.12.1 существует базис, в котором его матрица диагональна. Если он самосопряжен, то каждый диагональный блок 1×1 самосопряжен, поэтому в нем стоит комплексное число λ такое, что $\lambda = \overline{\lambda}$, то есть, $\lambda \in \mathbb{R}$. Аналогично, из кососимметричности следует, что λ чисто мнимое, а из унитарности — то, что $|\lambda|^2 = \lambda\overline{\lambda} = 1$.

Обратно, если все диагональные элементы матрицы имеют указанный вид, то прямая проверка показывает, что оператор $\mathfrak{a}$ обладает соответствующим свойством. □

Теорема 9.13.4. *Пусть* (V, B) *— конечномерное эвклидово пространство,* $\mathfrak{a}\colon V \to V$ *— линейный оператор.*

1. *Оператор* $\mathfrak{a}$ *является самосопряженным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* $\mathfrak{a}$ *диагональна.*

2. *Оператор* $\mathfrak{a}$ *является кососимметрическим тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* $\mathfrak{a}$ *имеет блочно-диагональный вид, и каждый блок выглядит как* (0) *или* $\begin{pmatrix} 0 & -b \\ b & 0 \end{pmatrix}$ *для* $b \in \mathbb{R}$*,* $\beta > 0$*.*

3. *Оператор* $\mathfrak{a}$ *является ортогональным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* $\mathfrak{a}$ *имеет блочно-диагональный вид, и каждый блок выглядит как* (1)*,* (-1) *или* $\begin{pmatrix} a & -b \\ b & a \end{pmatrix}$ *для* $a, b \in \mathbb{R}$*,* $b > 0$*,* $a^2 + b^2 = 1$*.*

Доказательство. Если оператор самосопряженный, кососимметрический, нормальный, то по теореме 9.12.10 существует базис, в котором его матрица блочно-диагональна, с блоками

вида

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix},$$

где $b > 0$. Если он самосопряжен, то каждый диагональный блок самосопряжен, что для блока 2×2 указанного вида означает, что $b = -b$, что невозможно. Поэтому остаются только блоки размера 1×1, что означает диагональность матрицы. Аналогично, из кососимметричности для блока 2×2 следует, что $a = 0$, а для блока (λ) размера 1×1 — что $\lambda = 0$. Наконец, из ортогональности для блока 2×2 следует, что $s^2 + b^2 = 1$, а для блока (λ) — что $\lambda^2 = 1$, откуда следует, что $\lambda = \pm 1$.

Обратно, если матрица оператора состоит из блоков указанного вида, нетрудно проверить, что оператор обладает соответствующим свойством. □

Определение 9.13.5. Пусть (V, B) — эвклидово или унитарное пространство, $a\colon V \to V$ — линейный оператор. Будем говорить, что оператор a **сохраняет скалярное произведение**, если $B(a(u), a(v)) = B(u, v)$ для любых $u, v \in V$. Оператор a называется **изометрией**, если $||a(v)|| = ||v||$ для всех $v \in V$.

Лемма 9.13.6. *Пусть $a\colon V \to V$ — линейный оператор на эвклидовом или унитарном пространстве (V, B). Следующие условия равносильны:*

1. *a ортогонален (в случае эвклидова пространства) или унитарен (в случае унитарного пространства);*
2. *a сохраняет скалярное произведение;*
3. *a является изометрией.*

Доказательство. $1 \Rightarrow 2$ Пусть a ортогонален/унитарен. Тогда $B(a(u), a(v)) = B(u, a^*(a(v)))$ по определению сопряженного оператора; из равенства $a^* \circ a = \mathrm{id}$ теперь следует, что $B(a(u), a(v)) = B(u, v)$.

$2 \Rightarrow 1$ Пусть $B(a(u), a(v)) = B(u, v)$ для всех $u, v \in V$. По определению сопряженного оператора $B(a(u), a(v)) = B(u, a^*(a(v)))$. Стало быть, $B(u, v) = B(u, a^*(a(v)))$ для всех $u, v \in V$. Значит, вектор $v - a^*(a(v))$ ортогонален всем векторам $u \in V$, откуда следует, что $v = a^*(a(v))$ для всех $v \in V$. Поэтому $a^* \circ a = \mathrm{id}$.

$2 \Rightarrow 3$ Если a сохраняет скалярное произведение, то, в частности, $B(a(v), a(v)) = B(v, v)$ для всех $v \in V$. Левая часть равна $||a(v)||^2$, а правая равна $||v||^2$. Извлекая [положительные] квадратные корни, получаем, что a является изометрией.

$3 \Rightarrow 2$ Если a является изометрией, то $B(a(u + \lambda v), a(u + \lambda v)) = B(u + \lambda v, u + \lambda v)$. Раскроем скобки:

$$\begin{aligned} &B(a(u), a(u)) + \overline{\lambda} B(a(v), a(u)) + \lambda B(a(u), a(v)) + \overline{\lambda}\lambda B(a(v), a(v)) \\ &= B(u, u) + \overline{\lambda} B(v, u) + \lambda B(u, v) + \overline{\lambda}\lambda B(v, v). \end{aligned}$$

Воспользуемся равенствами $B(a(x),a(x)) = B(x,x)$ и $B(x,y) = \overline{B(x,y)}$:

$$\lambda B(a(u),a(v)) + \overline{\lambda B(a(u),a(v))} = \lambda B(u,v) + \overline{\lambda B(u,v)}.$$

Подставляя $\lambda = 1$ и $\lambda = i$, получаем равенства

$$2\operatorname{Re}(B(a(u),a(v)) = 2\operatorname{Re}(B(u,v)), \quad 2\operatorname{Im}(B(a(u),a(v)) = 2\operatorname{Im}(B(u,v)).$$

Отсюда следует, что $B(a(u),a(v)) = B(u,v)$, что и требовалось.

□

Следствие 9.13.7 (Теорема Эйлера о вращениях трехмерного пространства). *Пусть $V = \mathbb{R}^3$ — трехмерное вещественное пространство со стандартным эвклидовым скалярным произведением, $a\colon \mathbb{R}^3 \to \mathbb{R}^3$ — изометрия на $\mathbb{R}^3$. Тогда в некотором ортогональном базисе матрица оператора a имеет вид*

$$\begin{pmatrix} \pm 1 & 0 & 0 \\ 0 & \cos(\varphi) & \sin(\varphi) \\ 0 & -\sin(\varphi) & \cos(\varphi) \end{pmatrix}$$

для некоторого угла φ. Если, кроме того, определитель оператора a равен 1, то элемент в левом верхнем углу такой матрицы равен 1.

Доказательство. По лемме 9.13.6 оператор a ортогонален. По теореме 9.13.4 найдется ортогональный базис V, в котором матрица оператора a имеет блочно-диагональный вид, и блоки имеют вид (± 1) или $\begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}$. Если там имеется блок размера 2, то теорема доказана. Если же все блоки имеют размер 1, то среди знаков ± 1 найдется два одинаковых, и их можно заменить на блок размера 2 вида $\begin{pmatrix} \cos(\varphi) & \sin(\varphi) \\ -\sin(\varphi) & \cos(\varphi) \end{pmatrix}$ для $\varphi = 0$ или $\varphi = \pi$. Последнее утверждение теоремы очевидно. □

Следствие 9.13.8 (Приведение вещественной квадратичной формы к диагональному виду при помощи ортогонального преобразования). *Пусть (V, B) — эвклидово пространство, и пусть $q\colon V \times V \to B$ — симметрическая билинейная форма. Существует ортогональный базис пространства V, в котором матрица Грама формы q имеет диагональный вид.*

Доказательство. Выберем некоторый ортонормированный базис $\mathcal{B}$ пространства V; пусть Q — матрица Грама формы q в этом базисе. Поскольку форма q симметрична, матрица Q является симметричной матрицей: $Q^T = Q$. Рассмотрим Q как матрицу некоторого оператора a на пространстве V; по предложению 9.9.4 оператор q самосопряжен. По теореме 9.13.4 существует ортонормированный базис $\mathcal{C}$ пространства V, в котором матрица оператора a диагональна. Это означает, что $C^{-1}QC = D$ — диагональная матрица, где C — матрица перехода от базиса $\mathcal{B}$ к базису $\mathcal{C}$ (см. теорему 7.9.5). Кроме того, поскольку C — матрица перехода между ортонормированными базисами, то C ортогональна (лемма 9.6.3): $C^T = C^{-1}$. Но тогда $D = C^T QC$, и по теореме 9.4.2 это означает, что D — матрица Грама квадратичной формы q в ортонормированном базисе $\mathcal{C}$. □

Замечание 9.13.9. Переформулируем утверждение первого пункта теоремы 9.13.4 на геометрическом языке. Если a — самосопряженный оператор на эвклидовом пространстве V, мы показали, что в некотором ортонормированном базисе его матрица A имеет диагональный вид. Пусть $\lambda_1, \ldots, \lambda_m$ — все различные собственные числа a; тогда у матрицы A на диагонали стоят числа $\lambda_1, \ldots, \lambda_m$ (возможно, некоторые встречаются по несколько раз). Очевидно, что собственное подпространство, соответствующее λ_i — это в точности линейная оболочка базисных векторов, соответствующих позициям, в которых на диагонали стоит λ_i. Поскольку базис ортонормирован, собственные подпространства, соответствующие различным собственным числам, попарно ортогональны; кроме того, их прямая сумма совпадает со всем пространством V (см. также раздел 8.3).

Таким образом, каждому самосопряженному оператору на V мы сопоставили разложение пространства V в ортогональную прямую сумму собственных подпространств, соответствующих различным собственным числам этого оператора. Обратно, если имеется разложение пространства V в ортогональную прямую сумму подпространств $V = \bigoplus_{i=1}^{m} V_m$ и заданы различные числа $\lambda_1, \ldots, \lambda_m$, то имеется единственный самосопряженный оператор a, который на векторе $v = \sum_{i=1}^{m} v_i$ (для $v_i \in V_i$) действует следующим образом: $a(v) = \sum_{i=1}^{m} \lambda_i v_i$. Если в каждом подпространстве V_i выбрать ортонормированный базис, то объединение этих базисов является ортонормированным базисом пространства V, и матрица оператора a в этом базисе диагональна; на диагонали стоят числа $\lambda_1, \ldots, \lambda_m$, и кратность λ_i равна размерности подпространства V_i.

Мы получили взаимно однозначное соответствие между самосопряженными операторами и разложениями $V = \bigoplus_{i=1}^{m} V_i$ с заданными попарно различными числами $\lambda_1, \ldots, \lambda_m$.

9.14 Положительно определенные операторы

Литература: [F], гл. XIII, § 4, п. 4; [K2], гл. 3, § 3, пп. 8, 9.

Пусть (V, B) — эвклидово или унитарное пространство, $a\colon V \to V$ — самосопряженный оператор на нем. Тогда в силу самосопряженности $B(a(v), v) = B(v, a(v))$ для любого $v \in V$; с другой стороны, $B(a(v), v) = \overline{B(v, a(v))}$. Поэтому выражение $B(a(v), v)$ всегда вещественно.

Определение 9.14.1. Самосопряженный оператор $a\colon V \to V$ на эвклидовом или унитарном пространстве V называется **неотрицательно определенным**, если $B(a(v), v) \geqslant 0$ для любого $v \in V$. Оператор a называется **положительно определенным**, если он неотрицательно определен и из $B(a(v), v) = 0$ следует, что $v = 0$.

Предложение 9.14.2. *Оператор $a\colon V \to V$ на эвклидовом или унитарном пространстве V неотрицательно определен тогда и только тогда, когда в некотором ортонормированном базисе матрица этого оператора диагональна, причем на диагонали стоят неотрицательные вещественные числа. Оператор a положительно определен тогда и только тогда, когда в некотором ортонормированном базисе матрица этого оператора диагональна, причем на диагонали стоят положительные вещественные числа.*

Доказательство. Если a неотрицательно определен, то он (по определению) самосопряжен, и по теоремам 9.13.3 и 9.13.4 существует ортонормированный базис $\mathcal{B} = (e_1, \ldots, e_n)$, в котором a имеет диагональную матрицу

$$[a]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}.$$

Предположим, что $\lambda_i < 0$. Тогда $a(e_i) = \lambda_i e_i$ и $B(a(e_i), e_i) = \lambda_i B(e_i, e_i) = \lambda_i < 0$, что противоречит неотрицательной определенности a. Если же a положительно определен, то и случай $\lambda_i = 0$ невозможен: если $\lambda_i = 0$, то $B(a(e_i), e_i) = \lambda_i = 0$, в то время как $e_i \neq 0$.

Обратно, пусть a в некотором ортонормированном базисе $\mathcal{B} = \{e_1, \ldots, e_n\}$ имеет диагональную матрицу с неотрицательными числами $\lambda_1, \ldots, \lambda_n$ на диагонали. По теоремам 9.13.3 и 9.13.4 мы уже знаем, что a самосопряжен. Разложим произвольный вектор v по базису $\mathcal{B}$: $v = \sum_i c_i e_i$. Тогда $a(v) = \sum_i c_i a(e_i) = \sum_i c_i \lambda_i e_i$. Поэтому

$$B(a(v), v) = B(\sum_i c_i \lambda_i e_i, \sum_j c_j e_j) = \sum_{i,j} \overline{c_i} \lambda_i c_j B(e_i, e_j) = \sum_i \lambda_i \overline{c_i} c_i B(e_i, e_i) = \sum_i \lambda_i |c_i|^2 \geqslant 0.$$

Если же все $\lambda_i > 0$ и оказалось, что $\sum_i \lambda_i |c_i|^2 = 0$, то и $c_i = 0$ для всех i, откуда $v = 0$. □

Замечание 9.14.3. Таким образом, положительно определенный оператор всегда является обратимым: его матрица в некотором базисе имеет ненулевой определитель. Кроме того, если неотрицательно определенный оператор обратим, то он положительно определен: у обратимой диагональной матрицы не может встретиться 0 на диагонали.

Теорема 9.14.4 (Извлечение квадратного корня в классе положительно определенных операторов). *Пусть* $a\colon V \to V$ *— положительно определенный оператор на эвклидовом или унитарном пространстве* V*. Существует единственный положительно определенный оператор* $b\colon V \to V$ *такой, что* $b^2 = a$*.*

Доказательство. По предложению 9.14.2 найдется базис $\mathcal{B} = (e_1, \ldots, e_n)$, такой, что

$$[a]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix},$$

причем λ_i — положительно вещественные числа. Рассмотрим оператор b, матрица которого в базисе $\mathcal{B}$ равна

$$[a]_{\mathcal{B}} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \ldots & 0 \\ 0 & \sqrt{\lambda_2} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \sqrt{\lambda_n} \end{pmatrix}.$$

Заметим, что $\sqrt{\lambda_i} > 0$ для всех i, поэтому (снова по предложению 9.14.2) оператор b положительно определен. Кроме того, очевидно, что $b^2 = a$.

Нам осталось показать, что такой оператор b единственный. Пусть $\widetilde{b}$ — другой оператор с теми же свойствами: $\widetilde{b}$ положительно определен и $\widetilde{b}^2 = a$. Воспользуемся замечанием 9.13.9 для оператора $\widetilde{b}$. А именно, пусть $\mu_1, \dots, \mu_n$ — собственные числа оператора $\widetilde{b}$ с учетом кратности. Тогда $\widetilde{b}$ приводится в некотором базисе к диагональному виду, и на диагонали стоят положительные числа $\mu_1, \dots, \mu_n$. Но тогда $a = \widetilde{b}^2$ в этом же базисе имеет диагональный вид, и на диагонали стоят числа $\mu_1^2, \dots, \mu_n^2$. Значит, собственные числа оператора a (с учетом кратности) равны $\mu_1^2, \dots, \mu_n^2$. С другой стороны, мы знаем, что они равны $\lambda_1, \dots, \lambda_n$. Мы знаем, что $\mu_i > 0$ для всех i, поэтому набор $\mu_1, \dots, \mu_n$ совпадает (с точностью до перестановки) с набором $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$.

Мы получили, что наборы собственных чисел операторов b и $\widetilde{b}$ совпадают. Осталось показать, что собственные подпространства для этих операторов, соответствующие одинаковым собственным числам, совпадают, и воспользоваться соответствием из замечания 9.13.9.

Пусть теперь V_i — собственное подпространство для оператора b, соответствующее собственному числу $\sqrt{\lambda_i}$. Оно натянуто на те векторы базиса $\mathcal{B}$, которым соответствуют номера столбиков, в которых в матрице b стоят числа $\sqrt{\lambda_i}$. После возведения в квадрат матрица остается диагональной, поэтому V_i является собственным подпространством оператора a, соответствующим собственному числу λ_i. Но то же самое рассуждение применимо и к оператору $\widetilde{b}$. Поэтому собственные подпространства для операторов b и $\widetilde{b}$, соответствующие $\sqrt{\lambda_i}$, совпадают. □

Следующая теорема является прямым обобщением того факта, что любое ненулевое комплексное число z можно (единственным образом) записать в тригонометрической форме (см. определение 3.3.1): $z = |z| \cdot (\cos(\varphi) + i\sin(\varphi))$. Здесь $|z|$ — положительное вещественное число, а $(\cos(\varphi) + i\sin(\varphi))$ — комплексное число, которое по модулю равно 1. Полярное разложение обобщает эту теорему на многомерный случай: слова «ненулевое число» нужно заменить на «обратимый оператор», слова «положительное вещественное число» на «положительно определенный оператор», а «комплексное число, равное по модулю 1» — на «унитарный оператор». Обратите внимание, что матрица 1×1 задается ровно одним числом, поэтому при подстановке в следующую теорему одномерного векторного пространства $V = \mathbb{C}$ действительно получается утверждение о тригонометрической форме комплексного числа. Вещественный случай еще проще: если $z \in \mathbb{R}\backslash\{0\}$, то $z = |z| \cdot (\pm 1)$; ортогональный оператор на одномерном пространстве может быть равен лишь 1 или -1.

Теорема 9.14.5 (Полярное разложение). *Пусть $a\colon V \to V$ — обратимый оператор на эвклидовом или унитарном пространстве. Тогда существуют операторы $p, u\colon V \to V$ такие, что $a = pu$, причем p — положительно определенный оператор, а u — ортогональный или унитарный. Более того, такие операторы единственны: если $a = p'u'$ для положительно определенного p и ортогонального/унитарного u, то $p = p'$ и $u = u'$.*

Доказательство. Рассмотрим оператор $c = a \circ a^*$. Заметим, что c самосопряжен: действительно, $c^* = (a \circ a^*)^* = a^{**} \circ a^* = a \circ a^* = c$. Кроме того, c неотрицательно определен:

$B(c(v), v) = B((a \circ a^*)(v), v) = B(a(a^*(v)), v) = B(a^*(v), a^*(v)) \geqslant 0$. Наконец, поскольку a обратим, то и a^* обратим (их матрицы в ортонормированном базисе транспонированны, поэтому из обратимости одной следует обратимость другой), значит, и c обратим; поэтому c положительно определен (см. замечание 9.14.3). По теореме 9.14.4 из c можно извлечь квадратный корень: найдется положительно определенный оператор p такой, что $p^2 = c = a \circ a^*$. В силу положительной определенности оператор p обратим. Обозначим теперь $u = p^{-1}a$. Тогда, очевидно, $a = pu$, и осталось проверить, что u — ортогональный/унитарный оператор. Заметим сначала, что $pp^{-1} = \mathrm{id}$, поэтому $(pp^{-1})^* = \mathrm{id}^* = \mathrm{id}$, откуда $(p^{-1})^* = p^{-1}$. Поэтому $u \circ u^* = p^{-1}a(p^{-1}a)^* = p^{-1}aa^*(p^{-1})^* = p^{-1}p^2p^{-1} = \mathrm{id}$, что и требовалось.

Наконец, если $pu = a = p'u'$, то $(pu)^* = (p'u')^*$, откуда $u^*p = (u')^*p'$. Из этого следует, что $(pu)(u^*p) = (p'u')((u')^*p^*)$, откуда $p^2 = (p')^2$, и в силу единственности извлечения квадратного корня (теорема 9.14.4), получаем, что $p = p'$, и, стало быть, $u = u'$. □

Замечание 9.14.6. Даже доказательство теоремы 9.14.5 напоминает доказательство факта про тригонометрическую форму записи комплексного числа: напомним, что модуль комплексного числа z определялся как $\sqrt{z \cdot \overline{z}}$ (см. определение 3.2.3); извлечение корня возможно в силу неотрицательности $z \cdot \overline{z}$.

10 Теория групп

10.1 Определения и примеры

Литература: [F], гл. I, § 3, п. 1, гл. X, § 1, пп. 1–2, § 5, п. 1; [K1], гл. 4, § 2, п. 1; [vdW], гл. 2, § 6; [Bog], гл. 1, § 1.

Мы уже встречали определение группы (см. определение 5.6.1):

Определение 10.1.1. Множество G с бинарной операцией $\circ\colon G \times G \to G$ называется **группой**, если выполняются следующие свойства:

- $a \circ (b \circ c) = (a \circ b) \circ c$ для всех $a, b, c \in G$; (**ассоциативность**);
- существует элемент $e \in G$ (**единичный элемент**) такой, что для любого $a \in G$ выполнено $a \circ e = e \circ a = a$;
- для любого $a \in G$ найдется элемент $a^{-1} \in G$ (называемый **обратным** к a) такой, что $a \circ a^{-1} = a^{-1} \circ a = e$.

Группа G называется **коммутативной**, или **абелевой**, если $a \circ b = b \circ a$ для всех $a, b \in G$.

В прошлом семестре мы некоторое время изучали *группу перестановок* $S(X)$ множества X (см. определение 5.6.2):

Определение 10.1.2. Множество всех биекций из X в X обозначается через $S(X)$ и называется **группой перестановок** множества X. Тождественное отображение $\mathrm{id}_X\colon X \to X$ называется **тождественной перестановкой**. Если $X = \{1, \ldots, n\}$, мы обозначаем группу $S(X)$ через S_n и называем ее **симметрической группой на n элементах**.

В разделе 5.6 мы видели, что группа S_n не является абелевой при $n \geqslant 3$.

На самом деле мы встречали и другие группы.

Примеры 10.1.3.

1. Пусть R — кольцо (см.определение 2.8.1). В частности, это означает что на R задана операция сложения. Из определения кольца сразу следует, что R относительно этой операции сложения является абелевой группой. Она называется **аддитивной группой кольца**. В частности, множества $\mathbb{Z}$, $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ являются абелевыми группами относительно сложения.

2. Пусть V — векторное пространство над полем k (см. определение 6.1.1). В частности, на V задана операция сложения. Относительно этой операции множество V является абелевой группой.

3. Пусть k — поле. Тогда умножение является ассоциативной, коммутативной операцией, единица поля является нейтральным элементом относительно этой операции, и у каждого ненулевого элемента имеется обратный. Это означает, что $k^* = k\backslash\{0\}$ является абелевой группой. Эта группа называется **мультипликативной группой поля** k. В частности, множества $\mathbb{Q}^*$, $\mathbb{R}^*$, $\mathbb{C}$ являются абелевыми группами относительно умножения.

4. Более общо, пусть R — ассоциативное кольцо с единицей (не обязательно коммутативное). Обозначим через R^* множество *двусторонне обратимых* элементов R, то есть, множество элементов $x \in R$ таких, что существует $y \in R$, для которого $xy = yx = 1$. Нетрудно проверить (сделайте это!), что множество R^* образует группу относительно умножения. Эта группа называется **группой обратимых элементов кольца** R. В частности, если R — поле, то все ненулевые элементы R [двусторонне] обратимы, и мы получаем мультипликативную группу поля из предыдущего примера. Простейший пример: $\mathbb{Z}^* = \{1, -1\}$.

5. Пусть k — некоторое поле, $n \geqslant 1$. Мы знаем, что множество квадратных матриц размера $n \times n$ образует кольцо относительно операций сложения и умножения матриц (см. замечание 5.3.5). Группа обратимых элементов этого кольца обозначается через $\mathrm{GL}(n, k)$ и называется **полной линейной группой**. Таким образом, $\mathrm{GL}(n, k)$ состоит из обратимых матриц размера $n \times n$, и это группа относительно операции умножения. В частности, при $n = 1$ получаем группу k^* обратимых элементов поля k (см. пример 3).

6. В продолжение предыдущего примера, рассмотрим подмножество $\mathrm{SL}(n, k) \subseteq \mathrm{GL}(n, k)$, состоящее из матриц с определителем 1. Напомним, что определитель произведения матриц равен произведению их определителей, и (см. теорему 5.8.5). Более того, если $x \in \mathrm{SL}(n, k)$ — матрица с определителем 1, то и обратная матрица x^{-1} имеет определитель 1. Поэтому множество $\mathrm{SL}(n, k)$ само является группой относительно операции умножения. Эта группа называется **специальной линейной группой**.

7. Пусть $\mathbb{T} = \{z \in \mathbb{C} \mid |z| = 1\}$ — множество комплексных чисел с модулем 1. Это группа по умножению (поскольку модуль комплексного числа мультипликативен, см. предложение 3.2.6). Она часто называется **группой углов**. Ниже (см. пример 10.5.2 (4)) мы приведем другое ее описание, не использующее комплексных чисел.

8. Наиболее архетипичный пример группы выглядит так: рассмотрим все обратимые преобразования (*автоморфизмы*) некоторого объекта в себя (и/или сохраняющих *нечто*). Это группа относительно композиции: действительно, композиция преобразований объекта в себя (сохраняющих *нечто*) является преобразованием объекта в себя (сохраняющим *нечто*); композиция преобразований всегда ассоциативна; тождественное преобразование должно сохранять *нечто* и потому является нейтральным элементом; наконец, мы потребовали обратимость, поэтому и с обратными элементами нет проблемы. Рассмотренные выше примеры все сводятся к этому. Симметрическая группа — это просто группа обратимых преобразований *множества* без всякой дополнительной структуры. $\mathrm{GL}(n, k)$ — группа преобразований векторного пространства (сохраняющих структуру векторного пространства — сложение и умножение на скаляры — то есть, *линейных*). $\mathrm{SL}(n, k)$ — группа линейных преобразований определителя 1, то есть, *сохраняющих ориентированный объем* (мы узнаем, что это такое, в главе 11). Даже группу целых чисел по сложению можно интерпретировать схожим образом: рассмотрим целое число x

как сдвиг вещественной прямой (с отмеченными целыми точками) на x вправо (если x отрицательно, получаем сдвиг влево). Композиция таких сдвигов в точности соответствует сложению целых чисел. Такой *геометрический взгляд* на теорию групп чрезвычайно продуктивен: более того, Давид Гильберт продемонстрировал, что синтетическая геометрия (эвклидова, геометрия Лобачевского, проективная) целиком вкладывается в теорию групп.

10.2 Подгруппы

Литература: [F], гл. X, § 1, пп. 3–4, § 3, п. 6; [vdW], гл. 2, § 7; [Bog], гл. 1, § 1.

Ситуация, описанная в примере 10.1.3 (6), встречается достаточно часто:

Определение 10.2.1. Пусть G — некоторая группа. Подмножество $H \subseteq G$ называется **подгруппой** группы G, если выполнены следующие условия:

1. если $h, h' \in H$, то $h \circ h' \in H$.
2. если $h \in H$, то $h^{-1} \in H$.

Обозначение: $H \leqslant G$.

Заметим, что если H — подгруппа группы G, то множество H само является группой относительно той же операции (точнее, относительно *ограничения* этой операции на H).

Примеры 10.2.2.
1. В любой группе G имеются подгруппы $\{e\} \leqslant G$ и $G \leqslant G$; подгруппа $\{e\}$ называется **тривиальной** и часто обозначается через 1 или 0 (если групповая операция в G записывается мультипликативно или аддитивно, соответственно).
2. Как мы уже видели выше, $\mathrm{SL}(n, k) \leqslant \mathrm{GL}(n, k)$.
3. Напомним, что все перестановки из S_n делятся на *четные* и *нечетные* (см. определение 5.6.7), причем произведение четных перестановок четно (теорема 5.6.12), и обратная к четной перестановке четна (следствие 5.6.13). Это означает, что множество четных перестановок образует подгруппу в S_n. Она обозначается через A_n и называется **знакопеременной группой**.
4. Рассмотрим аддитивную группу целых чисел $\mathbb{Z}$. Пусть $m \in \mathbb{N}$. Множество $m\mathbb{Z} = \{mx \mid x \in \mathbb{Z}\}$ является подгруппой в $\mathbb{Z}$. Действительно, $mx + my = m(x + y) \in m\mathbb{Z}$ и $-mx = m(-x) \in m\mathbb{Z}$. В частности, $0\mathbb{Z} = 0$, $1\mathbb{Z} = \mathbb{Z}$. Ниже мы увидим, что любая подгруппа $\mathbb{Z}$ имеет вид $m\mathbb{Z}$ для некоторого натурального m.

Теорема 10.2.3. *Любая подгруппа G аддитивной группы $\mathbb{Z}$ целых чисел имеет вид $m\mathbb{Z}$ для некоторого натурального m.*

Доказательство. Если $G = \{0\}$, можно взять $m = 0$. В противном случае выберем наименьший по модулю элемент из $G\backslash\{0\}$. Заменив при необходимости знак, можно считать, что этот элемент больше нуля. Обозначим его через m и покажем, что $G = m\mathbb{Z}$. Во-первых, для натурального x имеем $mx = \underbrace{m + \cdots + m}_{x} \in G$ и $m(-x) = (-m)x = \underbrace{(-m) + \cdots + (-m)}_{x} \in G$; поэтому $m\mathbb{Z} \subseteq G$. Обратно, пусть $g \in G$. Поделим с остатком g на m: $g = mq + r$. При этом $0 \leqslant r < |m| = m$. Поскольку $g \in G$ и $mq \in G$, получае, что $r = g - mq \in G$. Если $r \neq 0$, это противоречит минимальности m. Значит, $g = mq$ и мы показали, что $g \in m\mathbb{Z}$. Это доказывает обратное включение $G \subseteq m\mathbb{Z}$. □

Полезно знать, что пересечение произвольного (конечного или бесконечного) набора подгрупп группы G снова является подгруппой в G.

Лемма 10.2.4. *Пусть* $\{H_i\}_{i\in I}$ *— семейство подгрупп группы* G. *Обозначим* $H = \bigcap_{i\in I} H_i$. *Тогда* $H \leqslant G$.

Доказательство. Если $h, h' \in H$, то $h, h' \in H_i$ и $h^{-1} \in H_i$ для всех $i \in I$, и поэтому $hh', h^{-1} \in H_i$ для всех $i \in I$, откуда $hh', h^{-1} \in H$. □

Весьма важен следующий способ построения подгрупп: пусть X — произвольное *подмножество* группы G. Мы хотим «наименьшими усилиями» расширить X так, чтобы получилась подгруппа.

Определение 10.2.5. Пусть $X \subseteq G$ — подмножество группы G. Наименьшая подгруппа в G, содержащая X, называется **подгруппой, порожденной подмножеством** X, и обозначается через $\langle X\rangle$. Более подробно, $\langle X\rangle \leqslant G$ — такая подгруппа группы G, что $X \subseteq \langle X\rangle$ и для любой подгруппы $H \leqslant G$, содержащей X, выполнено $\langle X\rangle \leqslant H$.

Замечание 10.2.6. Для конечного множества $X = \{x_1, \ldots, x_n\}$ мы часто пишем $\langle x_1, \ldots, x_n\rangle$ вместо $\langle\{x_1, \ldots, x_n\}\rangle$.

Определение 10.2.5 хорошо всем, кроме одного: a priori совершенно не очевидно, что для данного подмножества $X \subseteq G$ существует подгруппа $\langle X\rangle \leqslant G$ с указанными удивительными свойствами. Следующее предложение показывает, что это действительно так.

Предложение 10.2.7. *Пусть* G *— группа,* $X \subseteq G$. *Пересечение всех подгрупп в* G, *содержащих* X, *является подгруппой в* G, *порожденной множеством* X.

Доказательство. По лемме 10.2.4 пересечение всех подгрупп в G, содержащих X, является подгруппой в G. Обозначим ее через $\langle X\rangle$ и проверим, что она удовлетворяет определению 10.2.5. Действительно, множество X содержится во всех пересекаемых подгруппах, поэтому содержится в $\langle X\rangle$. С другой стороны, если $H \leqslant G$ содержит X, то H является одной из пересекаемых подгрупп, поэтому полученное пересечение $\langle X\rangle$ содержится в H. □

Замечание 10.2.8. Обратите внимание на сходство предложения 10.2.7 и определения линейной оболочки 6.3.1. Понятие подгруппы, порожденной множеством элементов G, является точным аналогом понятия линейной оболочки множества элементов векторного пространства.

Лемма 10.2.9. *Пусть* G *— группа,* $X \subseteq G$*. Подгруппа, порожденная множеством* X *— это множество всех произведений элементов* X *и обратных к ним:*

$$\langle X \rangle = \{y_1 y_2 \dots y_n \mid y_i \in X \text{ или } y_i^{-1} \in X \text{ для всех } i = 1, \dots, n\}.$$

Доказательство. Обозначим правую часть равенства через Y. Докажем сначала, что $Y \subseteq \langle X \rangle$. Пусть $y = y_1 y_2 \dots y_n$ — некоторый элемент Y; мы знаем, что каждый y_i либо является элементом X, либо является обратным к элементу X. Если $H \leqslant G$ — произвольная подгруппа, содержащая X, то H содержит и элементы $y_1, \dots, y_n$, а потому содержит и их произведение y. Значит, y лежит в пересечении всех таких подгрупп H, которое равно $\langle X \rangle$ по предложению 10.2.7.

Для доказательства обратного включения заметим, что множество Y само является подгруппой в G, содержащей множество X. В силу определения 10.2.5 из этого следует, что $\langle X \rangle \leqslant Y$. □

Следующее понятие продолжает эту мысль, вводя аналог понятия *системы образующих* векторного пространства (см. определение 6.3.3).

Определение 10.2.10. Говорят, что группа G **порождается** множеством $X \subseteq G$, и что X — **система порождающих** (или **порождающее множество**) группы G, если $\langle X \rangle = G$.

Примеры 10.2.11.

1. Предложение 5.6.4 в точности показывает, что группа S_n порождается множеством всех транспозиций, а вместе с предложением 5.6.5 оно означает, что группа S_n порождается множеством всех элементарных транспозиций.

2. Группа целых чисел $(\mathbb{Z}, +)$ порождается одним элементом 1. Действительно, любое натуральное число n является суммой n единиц: $n = \underbrace{1 + 1 + \dots + 1}_{n}$, а любое отрицательное число $-n$ является суммой n минус единиц: $-n = \underbrace{(-1) + (-1) + \dots + (-1)}$.

10.3 Классы смежности и нормальные подгруппы

Литература: [F], гл. X, § 1, пп. 5, § 2; [K3], гл. 1, § 2, п. 1; [vdW], гл. 2, §§ 8–9; [Bog], гл. 1, § 2.

Определение 10.3.1. Пусть G — группа, $H \leqslant G$ — ее подгруппа, и $g \in G$. Множество

$$gH = \{gh \mid h \in H\}$$

называется **правым смежным классом элемента** g **по подгруппе** H. Аналогично, множество

$$Hg = \{hg \mid h \in H\}$$

называется **левым смежным классом элемента** g **по подгруппе** H.

Предложение 10.3.2. *Пусть* G *— группа,* $H \leqslant G$*. Любые два правых смежных класса по подгруппе* H *либо не пересекаются, либо совпадают. Таким образом, группа* G *разбивается на правые смежные классы. Аналогично, любые два левых смежных класса по подгруппе* H *либо не пересекаются, либо совпадают. Таким образом,* G *разбивается на левые смежные классы.*

Доказательство. Пусть $gH, g'H$ — два правых смежных класса. Предположим, что они пересекаются: $x \in gH \cap g'H$. Тогда $x = gh = g'h'$ для некоторых $h, h' \in H$, откуда $g = g'h'h^{-1}$. Если y — еще один элемент gH, $y = gh''$, то $y = g'h'h^{-1}h''$, поэтому $y \in g'H$. Аналогично, если $y \in g'H$, то $y \in gH$. Поэтому $gH = g'H$. Осталось заметить, что каждый элемент $g \in G$ лежит в некотором правом смежном классе, хотя бы, $g \in gH$. Доказательство для левых смежных классов совершенно аналогично. □

Предложение 10.3.2 чрезвычайно похоже на теорему 1.5.4 о разбиении на классы эквивалентности. Это не случайно: за смежными классами стоят достаточно естественные отношения эквивалентности.

Определение 10.3.3. Пусть G — группа, $H \leqslant G$. Введем на G отношения $\sim_H$ и ${}_H\sim$. Будем говорить, что $g \sim_H g'$, если $g^{-1}g' \in H$. Будем говорить, что $g {}_H\sim g'$, если $g'g^{-1} \in H$.

Лемма 10.3.4. *Отношения* $\sim_H$ *и* ${}_H\sim$ *являются отношениями эквивалентности; класс элемента* $g \in G$ *по отношению* $\sim_H$ *— это в точности правый смежный класс* gH*, а по отношению* ${}_H\sim$ *— левый смежный класс* Hg*.*

Доказательство. Мы докажем лемму только для $\sim_H$ и правых смежных классов; остальное совершенно аналогично. Проверим рефлексивность, симметричность и транзитивность отношения $\sim_H$: для $g \in G$ имеем $g^{-1}g = e \in H$, поэтому $g \sim_H g$. Если $g \sim_H g'$, то $g^{-1}g' \in H$, поэтому и $g'^{-1}g = (g^{-1}g')^{-1} \in H$, откуда $g' \sim_H g$. Наконец, если $g \sim_H g'$ и $g' \sim_H g''$, то $g^{-1}g' \in H$ и $g'^{-1}g'' \in H$, поэтому и их произведение $g^{-1}g'' = (g^{-1}g')(g'^{-1}g'') \in H$, откуда $g \sim_H g''$.

Заметим, что $y \in G$ лежит в классе элемента $g \in G$ тогда и только тогда, когда $g \sim_H y$ (см. определение 1.5.3). Это равносильно тому, что $g^{-1}y \in H$, то есть, что $g^{-1}y = h$ для некоторого $h \in H$. Это, в свою очередь, равносильно тому, что $y = gh$, то есть, что $y \in gH$. □

Определение 10.3.5. Пусть G — группа, $H \leqslant G$. Множество правых смежных классов G по H (оно же фактор-множество G по отношению эквивалентности $\sim_H$) обозначается через G/H. Множество левых смежных классов G по H (оно же фактор-множество G по отношению эквивалентности ${}_H\sim$) обозначается через $H\backslash G$.

Замечание 10.3.6. Отношения $\sim_H$ и ${}_H\sim$ являются прямыми аналогами сравнения по модулю подпространства (см. определение 7.7.1); однако, отсутствие коммутативности приводит к тому, что необходимо рассматривать два варианта обобщения: условие $v_1 - v_2 \in U$ из определения 7.7.1 мы заменяем на $v_1 v_2^{-1}$ в одном варианте и на v_2^{-1} в другом. Если группа G абелева, то $gH = Hg$ для всех $g \in G$, и отношения $\sim_H$, ${}_H\sim$ совпадают.

Продолжим аналогию с линейной алгеброй: следующим шагом в построении факторпространства было введение структуры векторного пространства на множестве классов эквивалентности по модулю подпространства (предложение 7.7.2). В случае групп отсутствие коммутативности приводит к фатальным последствиям: оказывается, что для произвольной подгруппы $H \leqslant G$ фактор-множество G/H не обязано снабжаться естественной структурой группы. Для того, чтобы G/H оказалось группой, необходимо наложить на H дополнительное условие *нормальности*.

Определение 10.3.7. Пусть G — группа. Подгруппа $H \leqslant G$ называется **нормальной** (обозначение: $H \trianglelefteqslant G$), если для любого элемента $g \in G$ его левый и правый смежный классы совпадают: $Hg = gH$.

Полезны следующие переформулировки нормальности.

Лемма 10.3.8. *Пусть G — группа, $H \leqslant G$. Следующие условия равносильны:*

1. *H нормальна в G;*
2. *$gHg^{-1} = H$ для всех $g \in G$;*
3. *$gHg^{-1} \subseteq H$ для всех $g \in G$.*

(Здесь $gHg^{-1} = \{ghg^{-1} \mid h \in H\}$).

Доказательство. $1 \Rightarrow 2$ Пусть $Hg = gH$ и $h \in H$. Рассмотрим элемент ghg^{-1}. По предположению элемент gh можно записать в виде $h'g$ для некоторого $h' \in H$. Поэтому $ghg^{-1} = (gh)g^{-1} = (h'g)g^{-1} = h' \in H$. Это значит, что $gHg^{-1} \subseteq H$. Обратно, для $h \in H$ запишем $h = hgg^{-1}$; по предположению элемент hg можно записать в виде gh' для некоторого $h' \in H$. Значит, $h = (hg)g^{-1} = gh'g^{-1} \in gHg^{-1}$. Отсюда $H \subseteq gHg^{-1}$, и необходимое равенство доказано.

$2 \Rightarrow 3$ Очевидно.

$3 \Rightarrow 1$ Пусть $gHg^{-1} \subseteq H$. Возьмем $h \in H$ и рассмотрим элемент gh. Мы знаем, что $ghg^{-1} = h' \in H$, откуда $gh = h'g$; поэтому $gH \subseteq Hg$. Обратно, рассмотрим элемент $hg \in Hg$. Применяя предположение к g^{-1}, получаем, что $g^{-1}Hg \subseteq H$. Значит, элемент $g^{-1}hg = h''$ лежит в H. Отсюда $hg = gh''$, и мы показали, что $Hg \subseteq gH$.

□

Определение 10.3.9. Пусть G — группа, $g, h \in G$. Элемент ghg^{-1} называется **сопряженным к h при помощи g**; говорят, что элементы h и ghg^{-1} **сопряжены**. Обозначение: $ghg^{-1} = {}^g h$.

Замечание 10.3.10. Из замечания 10.3.6 следует, что все подгруппы абелевой группы нормальны.

Примеры 10.3.11.

1. $\mathrm{SL}(n,k) \trianglelefteqslant \mathrm{GL}(n,k)$. Действительно, если $h \in \mathrm{SL}(n,k)$ и $g \in \mathrm{GL}(n,k)$, то $\det(ghg^{-1}) = \det(g)\cdot\det(h)\cdot\det(g^{-1}) = \det(h) = 1$, поэтому ${}^g h \in \mathrm{SL}(n,k)$.

2. $A_n \trianglelefteqslant S_n$. Это доказывается совершенно аналогично предыдущему примеру, с заменой определителя на знак перестановки. Нормальность в обоих этих примерах также следует из леммы 10.4.5.

3. Любая подгруппа индекса 2 нормальна. Мы докажем это чуть позже.

10.4 Гомоморфизмы групп

Литература: [F], гл. X, § 3, п. 1; [K1], гл. 4, § 2, пп. 3–4; [vdW], гл. 2, § 10; [Bog], гл. 1, § 3.

Определение 10.4.1. Пусть G, H — группы. Отображение $\varphi\colon G \to H$ называется **гомоморфизмом групп**, если $\varphi(xy) = \varphi(x)\varphi(y)$ для всех $x, y \in G$.

Лемма 10.4.2. *Пусть $\varphi\colon G \to H$ — гомоморфизм групп. Тогда $\varphi(e_G) = e_H$ и $\varphi(x^{-1}) = \varphi(x)^{-1}$ для всех $x \in G$.*

Доказательство. Заметим, что $e_G \cdot e_G = e_G$. Поэтому $\varphi(e_G) = \varphi(e_G \cdot e_G) = \varphi(e_G)\cdot\varphi(e_G)$. Домножим обе части полученного равенства справа на $\varphi(e_G)^-$:

$$\varphi(e_G)\cdot\varphi(e_G)^{-1} = \varphi(e_G)\cdot\varphi(e_G)\cdot\varphi(e_G)^{-1} = \varphi(e_G).$$

С другой стороны, левая часть очевидным образом равна e_H. Поэтому $e_H = \varphi(e_G)$.

Пусть теперь $x \in G$. Тогда $e_H = \varphi(e_G) = \varphi(x\cdot x^{-1}) = \varphi(x)\cdot\varphi(x^{-1})$. Домножая обе части на $\varphi(x)^{-1}$ слева, видим, что $\varphi(x)^{-1} = \varphi(x^{-1})$. □

Примеры 10.4.3.

1. Пусть G, H — произвольные группы. Отображение $\mathrm{const}_e\colon G \to H$, $g \mapsto e$, переводящее все элементы группы G в нейтральный элемент группы H, является гомоморфизмом групп. Такой гомоморфизм называется **тривиальным**. Тождественное отображение $\mathrm{id}_G\colon G \to G$ также является гомоморфизмом групп по тривиальным причинам.

2. Пусть $G = (\mathbb{R}, +)$ — аддитивная группа поля $\mathbb{R}$, и $H = \mathbb{R}^*$ — мультипликативная группа поля $\mathbb{R}$. Определим отображение $\exp\colon (\mathbb{R},+) \to \mathbb{R}^*$ посредством формулы $\exp(x) = e^x$, где e — основание натуральных логарифмов. Это гомоморфизм групп, поскольку $e^{x+y} = e^x\cdot e^y$ для всех вещественных x, y.

3. Пусть теперь $G = (\mathbb{R}_{>0}, \cdot)$ — группа положительных вещественных чисел с операцией умножения, $H = (\mathbb{R}, +)$ — аддитивная группа поля $\mathbb{R}$. Рассмотрим отображение логарифма $\ln\colon (\mathbb{R}_{>0},\cdot) \to (\mathbb{R},+)$. Это гомоморфизм групп, поскольку $\ln(xy) = \ln(x) + \ln(y)$ для всех вещественных $x, y > 0$.

4. Пусть $G = S_n$, $H = \{\pm 1\} = \mathbb{Z}^*$ — группа обратимых элементов кольца целых чисел. Отображение знака $\mathrm{sgn}\colon S_n \to \{\pm 1\}$ является гомоморфизмом групп (теорема 5.6.12).

5. Пусть $G = H = \mathbb{Z}$ — аддитивная группа целых чисел, и $m \in \mathbb{Z}$. Определим отображение $\varphi\colon \mathbb{Z} \to \mathbb{Z}$ умножения на m формулой $\varphi(x) = mx$ для всех целых x. Нетрудно видеть, что φ является гомоморфизмом групп: $m(x+y) = mx + my$. Более общо, если R — произвольное кольцо, и $m \in R$, то отображение $\varphi\colon R \to R$, $x \mapsto mx$ является гомоморфизмом аддитивной группы R в себя по причине дистрибутивности.

6. Пусть $G = \mathrm{GL}(n,k)$ — группа обратимых матриц размера $n \times n$ над некоторым полем k, а $H = k^*$ — мультипликативная группа этого поля. Определитель является гомоморфизмом $\det\colon \mathrm{GL}(n,k) \mapsto k^*$, поскольку $\det(xy) = \det(x)\det(y)$ для всех $x, y \in \mathrm{GL}(n,k)$ (теорема 5.8.5).

Определение 10.4.4. Пусть $\varphi\colon G \to H$ — гомоморфизм групп. **Ядром** гомоморфизма φ называется множество $\mathrm{Ker}(\varphi) = \{x \in G \mid f\varphi(x) = e_H\}$ (полный прообраз единицы). **Образом** гомоморфизма φ называется его теоретико-множественный образ: $\mathrm{Im}(\varphi) = \{y \in H \mid y = \varphi(x)$ для некоторого $x \in G\}$.

Предложение 10.4.5. *Образ гомоморфизма $\varphi\colon G \to H$ является подгруппой в H, а его ядро — нормальной подгруппой в G: $\mathrm{Im}(\varphi) \leqslant H$, $\mathrm{Ker}(\varphi) \trianglelefteqslant G$.*

Доказательство. Пусть $h, h' \in \mathrm{Im}(\varphi)$. Это означает, что найдутся $g, g' \in G$ такие, что $\varphi(g) = h$ и $\varphi(g') = h'$. Тогда $\varphi(gg') = \varphi(g)\varphi(g') = hh'$, откуда следует, что и $hh' \in \mathrm{Im}(\varphi)$. Кроме того, $\varphi(g^{-1}) = \varphi(g)^{-1} = h^{-1}$, откуда $h^{-1} \in \mathrm{Im}(\varphi)$.

Пусть теперь $g, g' \in \mathrm{Ker}(\varphi)$. Это означает, что $\varphi(g) = e$ и $\varphi(g') = e$. Тогда $\varphi(gg') = \varphi(g)\varphi(g') = e \cdot e = e$, поэтому $gg' \in \mathrm{Ker}(\varphi)$. Кроме того, $\varphi(g^{-1}) = \varphi(g)^{-1} = e^{-1} = e$, поэтому и $g^{-1} \in \mathrm{Ker}(\varphi)$.

Наконец, если $x \in \mathrm{Ker}(\varphi)$, то $\varphi(gxg^{-1}) = \varphi(g)\varphi(x)\varphi(g^{-1}) = \varphi(g)\varphi(g^{-1}) = \varphi(gg^{-1}) = e$, то есть, gxg^{-1} тоже лежит в $\mathrm{Ker}(\varphi)$. Мы показали, что $g\,\mathrm{Ker}(\varphi)g^{-1} \subseteq \mathrm{Ker}(\varphi)$ для любого $g \in G$; по лемме 10.3.8 этого достаточно для доказательства нормальности $\mathrm{Ker}(\varphi) \trianglelefteqslant G$. □

Замечание 10.4.6. Сравните с предложениями 7.3.4 и 7.3.7. Здесь нужно быть аккуратнее: операция в группе, в отличие от сложения в векторном пространстве, не обязана быть коммутативной. Тем не менее, доказательство переносится дословно.

Замечание 10.4.7. Пусть $\varphi\colon G \to H$ — гомоморфизм групп. Образ $\mathrm{Im}(\varphi)$ измеряет отклонение гомоморфизма от сюръективности: φ сюръективно тогда и только тогда, когда $\mathrm{Im}(\varphi) = H$. Аналогично, следующая лемма показывает, что ядро $\mathrm{Ker}(\varphi)$ измеряет отклонение φ от инъективности.

Лемма 10.4.8. *Пусть $\varphi\colon G \to H$ — гомоморфизм групп. Он инъективен тогда и только тогда, когда $\mathrm{Ker}(\varphi) = e$.*

Доказательство. Если φ инъективен, то есть только один элемент $g \in G$ такой, что $\varphi(g) = e$, и мы знаем, что $\varphi(e) = e$. Обратно, если $\mathrm{Ker}(\varphi) = e$ и $g, g' \in G$ таковы, что $\varphi(g) = \varphi(g')$, то $\varphi(g^{-1}g') = \varphi(g)^{-1}\varphi(g') = e$, поэтому $g^{-1}g' \in \mathrm{Ker}(\varphi) = e$, откуда $g = g'$. □

Определение 10.4.9. Пусть G, H — группы. Отображение $f\colon G \to H$ называется **изоморфизмом групп**, если f — гомоморфизм групп, и существует гомоморфизм групп $f'\colon H \to G$ такой, что $f' \circ f = \mathrm{id}_G$ и $f \circ f' = \mathrm{id}_H$.

Лемма 10.4.10. *Гомоморфизм групп $f\colon G \to H$ является изоморфизмом тогда и только тогда, когда f биективен.*

Доказательство. Если f изоморфизм, то у него имеется обратное отображение f', и поэтому f биективен. Обратно, если $f\colon G \to H$ – гомоморфизм, являющийся биекцией, рассмотрим обратное отображение $f^{-1}\colon H \to G$. Покажем, что это тоже гомоморфизм групп. Нам нужно проверить, что для любых $h, h' \in H$ выполнено $f^{-1}(h)\cdot f^{-1}(h') = f^{-1}(hh')$. Обозначим $f^{-1}(h) = g$, $f^{-1}(h') = g'$; тогда по предположению $f(gg') = f(g)f(g') = hh'$, откуда $gg' = f^{-1}(hh')$, что и требовалось. □

10.5 Фактор-группы

Литература: [F], гл. X, § 1, п. 5, § 2, § 3, п. 2; [K3], гл. 1, § 4, пп. 1–2; [vdW], гл. 2, §§ 8, 10; [Bog], гл. 1, § 2.

Пусть G — группа, и $H \leqslant G$ — ее нормальная подгруппа. Рассмотрим множество G/H правых классов смежности G по H и введем на нем бинарную операцию: для $gH, g'H \in G/H$ положим $(gH)\cdot(g'H) = (gg')H$.

Теорема 10.5.1. *Эта операция корректно определена и превращает фактор-множество G/H в группу. Каноническая проекция $G \to G/H$ на фактор-множество является гомоморфизмом групп.*

Доказательство. Корректная определенность означает, что если мы рассмотрим других представителей $\widetilde{g} \in gH$ и $\widetilde{g'} \in g'H$, то результат их перемножения будет тот же: $(\widetilde{g}\widetilde{g'})H = (gg')H$. Действительно, запишем $\widetilde{g} = gh$, $\widetilde{g'} = g'h'$; тогда $\widetilde{g}\widetilde{g'} = ghg'h' = g(hg')h'$. По определению нормальности элемент hg' можно записать в виде $g'h''$ для некоторого $h'' \in H$; поэтому $\widetilde{g}\widetilde{g'} = gg'h''h' \in gg'H$. Это и означает, что $\widetilde{g}\widetilde{g'}$ лежит в том же классе, что gg'.

Теперь несложно проверить ассоциативность: $(gH\cdot g'H)\cdot g''H = (gg')H\cdot g''H = (gg')g''H = g(g'g'')H = gH\cdot(g'g'')H = gH\cdot(g'H\cdot g''H)$. Нейтральным элементом для G/H служит смежный класс eH, поскольку $eH\cdot gH = (eg)H = gH = (ge)H = gH\cdot eH$. Наконец, у каждого класса gH имеется обратный класс $g^{-1}H$: $gH\cdot g^{-1}H = eH = g^{-1}H\cdot gH$.

Наконец, утверждение о том, что каноническая проекция $\pi\colon G \to G/H$ является гомоморфизмом, напрямую следует из определения операции в G/H. Действительно, $\pi(x)\pi(y) = xH\cdot yH$, в то время как $\pi(xy) = (xy)H$. □

Примеры 10.5.2.

1. $G/G \cong \{e\}$. Действительно, имеется только один класс смежности G по G.
2. $G/\{e\} \cong G$: все классы смежности G по подгруппе $\{e\}$ одноэлементны и поэтому отождествляются с элементами G. Формула для операции в фактор-группе превращается в

$g\{e\} \cdot g'\{e\} = gg'\{e\}$, что после отождествления означает, что $g \cdot g'$ полагается равным gg'; поэтому операция в $G/\{e\}$ та же, что была в G.

3. Мы уже встречали группу $\mathbb{Z}/m\mathbb{Z}$: это аддитивная группа кольца вычетов по модулю m.

4. Рассмотрим аддитивную группу поля вещественных чисел $\mathbb{R}$ и подгруппу $2\pi\mathbb{Z} = \{2\pi n \mid n \in \mathbb{Z}\}$ в ней. Фактор-группу $\mathbb{R}/2\pi\mathbb{Z}$ естественно представлять как множество вещественных чисел «с точностью до целых кратных 2π». Например, в этой группе есть элемент $3\pi/2$ (точнее, образ элемента $3\pi/2 \in \mathbb{R}$ относительно канонической проекции) и элемент π. Их сумма равна $3\pi/2 + \pi = 5\pi/2 = \pi/2 \in \mathbb{R}/2\pi\mathbb{Z}$, поскольку сложение происходит «по модулю 2π». Нетрудно понять, что эта группа изоморфна группе $\mathbb{T}$ комплексных чисел модуля 1 (см. пример 10.1.3 (7)) — изоморфизм устанавливается взятием аргумента. Поэтому группа $\mathbb{R}/2\pi\mathbb{Z}$, как и группа $\mathbb{T}$, часто называется **группой углов**.

Теперь мы можем доказать аналог теоремы о гомоморфизме 7.7.3.

Теорема 10.5.3 (Теорема о гомоморфизме). *Пусть* G, H — *группы,* $\varphi\colon G \to H$ — *гомоморфизм групп. Тогда* $G/\operatorname{Ker}(\varphi) \cong \operatorname{Im}(\varphi)$.

Доказательство. Определим отображение $\widetilde{\varphi}\colon G/\operatorname{Ker}(\varphi) \to \operatorname{Im}(\varphi)$ правилом $\widetilde{\varphi}(g\operatorname{Ker}(\varphi)) = \varphi(g)$. Заметим, прежде всего, что $\varphi(g)$ действительно лежит в $\operatorname{Im}(\varphi)$. Далее, этот гомоморфизм корректно определен: если $g\operatorname{Ker}(\varphi) = g'\operatorname{Ker}(\varphi)$, то $g = g'x$ для некоторого $x \in \operatorname{Ker}(\varphi)$, поэтому $\varphi(g) = \varphi(g'x) = \varphi(g')\varphi(x) = \varphi(g')e = \varphi(g')$.

Проверим, что $\widetilde{\varphi}$ — изоморфизм групп. Для этого по лемме 10.4.10 достаточно проверить, что $\widetilde{\varphi}$ — биективный гомоморфизм групп. Пусть $g\operatorname{Ker}(\varphi), g'\operatorname{Ker}(\varphi) \in G/\operatorname{Ker}(\varphi)$. Тогда $\widetilde{\varphi}(g\operatorname{Ker}(\varphi))\widetilde{\varphi}(g'\operatorname{Ker}(\varphi)) = \varphi(g)\varphi(g')$ и $\widetilde{\varphi}(g\operatorname{Ker}(\varphi) \cdot g'\operatorname{Ker}(\varphi)) = \widetilde{\varphi}((gg')\operatorname{Ker}(\varphi)) = \varphi(gg')$. Получили одно и то же (поскольку φ — гомоморфизм групп).

Для доказательства биективности проверим инъективность и сюръективность. Инъективность: по лемме 10.4.8 достаточно показать, что ядро $\widetilde{\varphi}$ тривиально. Если $g\operatorname{Ker}(\varphi)$ лежит в этом ядре, то $\widetilde{\varphi}(g\operatorname{Ker}(\varphi)) = \varphi(g) = e$, поэтому $g \in \operatorname{Ker}(\varphi)$ и $g\operatorname{Ker}(\varphi) = e\operatorname{Ker}(\varphi)$, что и требовалось. Сюръективность: если $h \in \operatorname{Im}(\varphi)$, то найдется $g \in G$ такой, что $\varphi(g) = h$. Но тогда $\widetilde{\varphi}(g\operatorname{Ker}(\varphi)) = \varphi(g) = h$. □

10.6 Циклические группы

Литература: [F], гл. X, § 1, пп. 6–7; [K1], гл. 4, § 2, п. 2; [K3], гл. 1, § 2, п. 2; [vdW], гл. 2, § 7.

Пусть G — произвольная группа, $g \in G$. Определим отображение $\operatorname{pow}_g\colon \mathbb{Z} \to G$ следующим образом: целое число n отправим в $g^n \in G$. Иными словами, для натурального n положим $g^n = \underbrace{g \cdot \dots \cdot g}_{n}$ и $g^{-n} = \underbrace{g^{-1} \cdot \dots \cdot g^{-1}}_{n}$. Легко видеть, что при этом $g^{m+n} = g^m \cdot g^n$ для всех $m, n \in \mathbb{Z}$ поэтому отображение pow_g является гомоморфизмом групп. Его образ по предложению 10.4.5 является подгруппой в G.

Лемма 10.6.1. *Образ отображения* pow_g *совпадает с* $\langle g \rangle$ *(подгруппой, порожденная* g*).*

Доказательство. Прежде всего, $\mathrm{Im}(\mathrm{pow}_g)$ содержит g, поэтому и $\langle g\rangle \subseteq \mathrm{Im}(\mathrm{pow}_g)$. С другой стороны, любой элемент $\mathrm{Im}(\mathrm{pow}_g)$ имеет вид g^n для некоторого n, и содержится в $\langle g\rangle$, поскольку $\langle g\rangle$ — подгруппа в G. □

Определение 10.6.2. Группа G называется **циклической**, если она порождается одним элементом, то есть, найдется элемент $g \in G$ такой, что $G = \langle g\rangle$.

Наша ближайшая задача — описать все циклические группы.

Теорема 10.6.3 (Классификация циклических групп). *Любая циклическая группа изоморфна $\mathbb{Z}/m\mathbb{Z}$ для некоторого натурального m. В случае $m = 0$ получаем бесконечную циклическую группу $\mathbb{Z}$, в остальных случаях получаем циклическую группу из m элементов.*

Доказательство. Пусть G — циклическая группа, порожденная элементом $g \in G$. Рассмотрим отображение $\mathrm{pow}_g\colon \mathbb{Z} \to G$. По лемме 10.6.1 его образ совпадает с $\langle g\rangle = G$. По теореме о гомоморфизме 10.5.3 имеем $\mathbb{Z}/\operatorname{Ker}(\mathrm{pow}_g) \cong G$. По теореме 10.2.3 $\operatorname{Ker}(\mathrm{pow}_g)$, будучи подгруппой в $\mathbb{Z}$, имеет вид $m\mathbb{Z}$ для некоторого натурального m, что и требовалось доказать. □

Следствие 10.6.4. *Пусть G — произвольная группа, $g \in G$. Множество $\{g^n \mid n \in \mathbb{Z}\}$ является подгруппой в G, изоморфной группе $\mathbb{Z}/m\mathbb{Z}$ для некоторого $m \in \mathbb{N}$.*

Доказательство. Это множество — циклическая подгруппа $\langle g\rangle$; осталось применить к ней теорему 10.6.3. □

Определение 10.6.5. Если группа $\{g^n \mid n \in \mathbb{Z}\}$ изоморфна $\mathbb{Z}/m\mathbb{Z}$ и $m > 0$, говорят, что элемент g имеет **порядок** m. Если же эта группа изоморфна $\mathbb{Z}$, то говорят, что g имеет **бесконечный порядок**. Таким образом, порядок элемента g равен числу элементов в циклической подгруппе $\langle g\rangle$, порожденной g. Обозначение для порядка: $\mathrm{ord}_G(g) = m$ или ∞.

Иными словами, порядок элемента $g \in G$ — это наименьшее натуральное число m такое, что $g^m = 1$. Действительно, при гомоморфизме $\mathrm{pow}_g\colon \mathbb{Z} \to G$ в единицу переходят в точности элементы из подгруппы $m\mathbb{Z}$.

Замечание 10.6.6. Заметим, что порядок нейтрального элемента равен 1, и это единственный элемент порядка 1 в любой группе.

10.7 Теорема Лагранжа

Литература: [F], гл. X, § 1, пп. 5, 7; [K3], гл. 1, § 2, п. 1; [Bog], гл. 1, § 2.

Определение 10.7.1. Пусть G — группа, $H \leqslant G$. Количество правых смежных классов G по H называется **индексом** подгруппы H и обозначается через $|G : H|$.

Покажем, что в этом определении можно заменить правые смежные классы на левые смежные классы:

Лемма 10.7.2. *Пусть* G *— группа,* $H \leqslant G$*. Тогда множества левых смежных классов* G *по* H *и правых смежных классов* G *по* H *равномощны.*

Доказательство. Пусть $\{a_i H\}_{i \in I}$ — множество всех правых смежных классов (иными словами, мы выбрали в каждом правом смежном классе по представителю и занумеровали их элементами некоторого множества I, возможно, бесконечного). По предложению 10.3.2 каждый элемент группы G содержится ровно в одном множестве вида $a_i H$. Покажем, что набор $\{H a_i^{-1}\}_{i \in I}$ состоит из всех левых смежных классов, взятых ровно по одному разу (то есть, что a_i^{-1} — представители всех левых смежных классов G по H).

Действительно, пусть $g \in G$. Тогда $g \in H a_i^{-1}$ равносильно тому, что $g = h a_i^{-1}$ для некоторого H, откуда $g^{-1} = (h a_i^{-1})^{-1} = a_i h^{-1} \in a_i H$. Но это равенство выполнено ровно для одного индекса $i \in I$, поэтому g лежит ровно в одном множестве вида $H a_i^{-1}$, что и требовалось доказать. □

Замечание 10.7.3. По определению фактор-множество G/H состоит из правых смежных классов G по H, так что $|G : H| = |G/H|$.

Теорема 10.7.4 (Теорема Лагранжа). *Пусть* G *— конечная группа,* $H \leqslant G$*. Тогда* $|G| = |H| \cdot |G : H|$.

Доказательство. Докажем, что во всех правых смежных классах G по H поровну элементов. Заметим, что для каждого $g \in G$ отображение $H \to gH$, $h \mapsto gh$, задает биекцию между H и gH. Действительно, если $gh = gh'$, то $h = h'$, и в силу определения смежного класса это отображение сюръективно. Поэтому в каждом смежном классе столько же элементов, сколько в подгруппе H. Таким образом, элементы G разбиваются на $|G : H|$ смежных классов, в каждом по H элементов. Отсюда сразу следует требуемое равенство. □

Следствие 10.7.5. *Порядок конечной группы* G *делится на порядок любой ее подгруппы. В частности, порядок конечной группы* G *делится на порядок любого ее элемента.*

Доказательство. Первое утверждение очевидно; второе следует из первого, если рассмотреть подгруппу $\langle g \rangle$, порядок которой (по определению) равен порядку g. □

Следствие 10.7.6. *Пусть* G *— конечная группа. Тогда* $g^{|G|} = 1$ *для любого* $g \in G$.

В качестве примера приложения теоремы Лагранжа выведем из нее теорему Эйлера 2.12.1 (и, как следствие, малую теорему Ферма 2.12.2).

Теорема 10.7.7. *Пусть* m *— натуральное число,* $a \in \mathbb{Z}$ *и* $a \perp m$*. Тогда* $a^{\varphi(m)} \equiv 1 \pmod{m}$.

Доказательство. Рассмотрим кольцо $\mathbb{Z}/m\mathbb{Z}$. Множество $(\mathbb{Z}/m\mathbb{Z})^*$ его обратимых элементов образует группу по умножению (пример 10.1.3 (4)). Порядок этой группы равен $\varphi(m)$ (предложение 2.11.3). Класс $\overline{a}$ элемента a в $\mathbb{Z}/m\mathbb{Z}$ обратим, поскольку $a \perp m$ (предложение 2.8.12). Применение следствия 10.7.6 дает $\overline{a}^{\varphi(m)} = \overline{1}$, что в переводе на язык целых чисел и дает нужное равенство. □

Еще одно приложение теоремы Лагранжа — описание всех групп простого порядка.

Теорема 10.7.8. *Пусть* G *— конечная группа порядка* p*, где* p *— простое число. Тогда* G *изоморфна циклической группе* $\mathbb{Z}/p\mathbb{Z}$*.*

Доказательство. По теореме Лагранжа порядок любого элемента группы G должен быть делителем p, и в силу простоты p он равен либо 1 либо p. По замечанию 10.6.6 в G лишь один элемент имеет порядок 1; поэтому найдется элемент $g \in G$ порядка p. Но тогда подгруппа $\langle g \rangle$ состоит из p элементов и, стало быть, совпадает с G. Значит, G циклическая, порождена элементом g и (по теореме 10.6.3) изоморфна $\mathbb{Z}/p\mathbb{Z}$. □

10.8 Прямое произведение

Литература: [F], гл. X, § 4, пп. 1–2, [K3], гл. 1, § 4, п. 4.

Пусть G, H — две группы. Рассмотрим декартово произведение множеств $G \times H$ и введем на нем операцию: положим $(g, h) \cdot (g', h') = (gg', hh')$ для $g, g' \in G$, $h, h' \in H$. Нетрудно видеть, что $G \times H$ с такой операцией является группой: ассоциативность выполняется, поскольку она выполняется в группах G и H, нейтральным элементом служит пара (e, e), обратным элементом к паре (g, h) является элемент (g^{-1}, h^{-1}).

Определение 10.8.1. Множество $G \times H$ с такой операцией называется **прямым произведением** групп G и H.

Предложение 10.8.2. *Пусть* G, H *— группы. Рассмотрим отображения*

$$\begin{aligned} i_1 &\colon G \to G \times H,\ g \mapsto (g, e),\\ i_2 &\colon H \to G \times H,\ h \mapsto (e, h),\\ \pi_1 &\colon G \times H \to G,\ (g, h) \mapsto g,\\ \pi_2 &\colon G \times H \to H,\ (g, h) \mapsto h. \end{aligned}$$

1. i_1, i_2 *— инъективные, а* π_1, π_2 *— сюръективные гомоморфизмы групп;*
2. $\mathrm{Im}(i_1) = \mathrm{Ker}(\pi_2) = G \times \{e\}$, $\mathrm{Im}(i_2) = \mathrm{Ker}(\pi_1) = \{e\} \times H$ *— нормальные подгруппы в* $G \times H$*;*
3. $\pi_1 \circ i_1 = \mathrm{id}_G$, $\pi_2 \circ i_2 = \mathrm{id}_H$*;* $\pi_1 \circ i_2 = 0$, $\pi_2 \circ i_1 = 0$*;*

Доказательство. 1. Очевидно.

2. $\mathrm{Im}(i_1)$ состоит в точности из элементов вида (g, e), а $\mathrm{Ker}(\pi_2)$ состоит из элементов (g, h) таких, что $h = e$; и то, и другое совпадает с $G \times \{e\} = \{(g, e) \in G \times H \mid g \in G\}$. Нормальность следует из предложения 10.4.5. Оставшееся аналогично.

3. $\pi_1(i_1(g)) = \pi_1((g, e)) = g$, $\pi_2(i_1(g)) = \pi_2((g, e)) = e$. Оставшееся аналогично.

□

Таким образом, отображения i_1, i_2 устанавливают изоморфизмы $G \cong G \times \{e\}$ и $H \cong \{e\} \times H$ между группами G, H и подгруппами в $G \times H$. Естественно поинтересоваться, когда верно обратное: когда в данной группе F можно найти две подгруппы G, H такие, что F изоморфно прямому произведению $G \times H$, и подгруппы G, H получаются посредством вложений i_1, i_2 для этого прямого произведения? Ответ дает следующая теорема.

Теорема 10.8.3. *Пусть F — группа. Пусть $G \leqslant F$, $H \leqslant F$ — две подгруппы в F. Обозначим через $j_1 \colon G \to F$, $j_2 \colon H \to F$ соответствующие вложения. Предположим, что выполнены следующие условия:*

1. $G \cap H = \{e\}$ *(пересечение этих подгрупп тривиально);*
2. $GH = F$ *(любой элемент x группы F можно записать в виде $x = gh$ для некоторых $g \in G$, $h \in H$);*
3. $gh = hg$ *для всех $g \in G$, $h \in H$ (подгруппы G и H коммутируют).*

Тогда группа F изоморфна прямому произведению G и H; более того, существует такой изоморфизм $\varphi \colon F \to G \times H$, что композиция

$$\pi_1 \circ \varphi \circ j_1 \colon G \to F \to G \times H \to G$$

является тождественным отображением на G, а композиция

$$\pi_2 \circ \varphi \circ j_2 \colon H \to F \to G \times H \to H$$

является тождественным отображением на H.

Доказательство. Построим изоморфизм φ. Возьмем $x \in F$ и запишем его (пользуясь свойством 2) в виде $x = gh$, где $g \in G$ и $h \in H$. Заметим, что такое представление единственно: если $x = g'h'$ для $g' \in G$, $h' \in H$, то $gh = g'h'$, откуда $g'^{-1}g = h'h^{-1}$; в левой части стоит элемент G, а в правой — элемент H, значит (по свойству 1) $g'^{-1}g = e = h'h^{-1}$, откуда $g = g'$ и $h = h'$. Поэтому мы можем положить $\varphi(x) = (g, h)$.

Проверим, что φ — гомоморфизм групп. Возьмем $y \in F$ и запишем его в виде $y = g'h'$, где $g', h' \in H$. Тогда $xy = (gh)(g'h') = g(hg')h' = (gg')(hh')$ (по свойству 3. По определению φ теперь $\varphi(xy) = (gg', hh')$, в то время как $\varphi(x) = (g, h)$, $\varphi(y) = (g', h')$, и, стало быть, $\varphi(x)\varphi(y) = (g, h)(g', h') = (gg', hh')$.

Для доказательства инъективности φ достаточно проверить тривиальность его ядра (лемма 10.4.8). Но если $\varphi(x) = (e, e)$, то $x = ee = e$. Для всех пар $(g, h) \in G \times H$ найдется $x = gh \in F$ такой, что $\varphi(x) = (g, h)$, поэтому φ сюръективен. Наконец, $\pi_1(\varphi(j_1(g))) = \pi_1(\varphi(g)) = \pi_1((g, e)) = g$ и $\pi_2(\varphi(j_2(h))) = \pi_2(\varphi(h)) = \pi_2((e, h)) = h$. □

10.9 Симметрическая группа

Литература: [F], гл. X, § 5, п. 4; [K1], гл. 1, § 8, п. 2, гл. 4, § 2, п. 3; [Bog], гл. 1, § 4.

Сейчас мы вернемся к изучению группы S_n.

Определение 10.9.1. Перестановка $\pi \in S_n$ называется **циклом длины** k, если для некоторых различных $i_1, \ldots, i_k \in \{1, \ldots, n\}$ выполнено $\pi(i_1) = i_2$, $\pi(i_2) = i_3$, $\ldots$, $\pi(i_{k-1}) = i_k$, $\pi(i_k) = i_1$, и для всех $j \in \{1, \ldots, n\} \backslash \{i_1, \ldots, i_k\}$ выполнено $\pi(j) = j$. Такой цикл мы будем обозначать так: $(i_1\ i_2\ \ldots i_k)$. При этом множество $\{i_1, \ldots, i_k\} \subseteq \{1, \ldots, n\}$ называется **носителем** цикла π. Два цикла $\pi, \rho \in S_n$ называются **независимыми**, если их носители не пересекаются. Заметим, что циклы длины 1 не очень полезно рассматривать: это тождественная перестановка.

Замечание 10.9.2. Заметим, что цикл длины k можно записать k различными способами: $(i_1\ i_2\ \ldots\ i_{k-1}\ i_k) = (i_2\ i_3\ \ldots\ i_k\ i_1) = \cdots = (i_k\ i_1\ \ldots\ i_{k-2}\ i_{k-1})$.

Лемма 10.9.3. *Независимые циклы коммутируют: если $\pi, \rho \in S_n$ — независимые циклы, то $\pi\rho = \rho\pi$.*

Доказательство. Непосредственное вычисление. □

Определение 10.9.4. Пусть $\pi \in S_n$. Множество $\mathrm{Fix}(\pi) = \{i \in \{1, \ldots, n\} \mid \pi(i) = i\}$ называется **множеством неподвижных точек** перестановки π, а его элементы — **неподвижными точками** π.

Теорема 10.9.5. *Любую перестановку $\pi \in S_n$ можно представить в виде произведения независимых циклов, носители которых не пересекаются с $\mathrm{Fix}(\pi)$.*

Доказательство. Будем вести индукцию по числу $i \in \{1, \ldots, n\}$ таких, что $\pi(i) \neq i$, то есть, по $n - \mathrm{Fix}(\pi)$. Если это число равно 0, то перестановка π тождественна и, таким образом, есть произведение пустого множества циклов. Это база индукции. Докажем переход. Пусть теперь множество $I = \{i \in \{1, \ldots, n\} \mid \pi(i) \neq i\}$ непусто; например, $i_1 \in I$. Рассмотрим последовательность $i_1, \pi(i_1), \pi^2(i_1), \ldots$. По предположению $i_1 \neq \pi(i_1)$. Рассмотрим первый элемент этой последовательности, совпадающий с каким-то из ранее встретившихся: такой найдется, поскольку все элементы этой последовательности лежат в конечном множестве $\{1, \ldots, n\}$. Пусть это $\pi^k(i_1) = \pi^l(i_1)$ при $k > l$. Если $l > 0$, ты применяя к этому равенству π^{-1}, получаем $\pi^{k-1}(i_1) = \pi^{l-1}(i_1)$, что противоречит предположению о минимальности k. Значит, $l = 0$ и $\pi^k(i_1) = i_1$. Кроме того, опять же в силу минимальности k, все элементы $i_1, \pi(i_1), \pi^2(i_1), \ldots, \pi^{k-1}(i_1)$ различны. Обозначим $i_2 = \pi(i_1), i_3 = \pi^2(i_1), \ldots, i_k = \pi^{k-1}(i_1)$ и рассмотрим цикл $\sigma = (i_1\ i_2\ \ldots\ i_k)$. Мы знаем, что $\pi(i_1) = i_2$, $\pi(i_2) = i_3$, $\ldots$, $\pi(i_{k-1}) = i_k$ и $\pi(i_k) = i_1$, поэтому произведение $\pi' = \sigma^{-1} \circ \pi$ обладает следующим свойством: $\pi'(i_1) = i_1$, $\pi'(i_2) = i_2$, $\ldots$, $\pi'(i_k) = i_k$, и $\pi'(j) = \pi(j)$ для всех $j \in \{1, \ldots, n\} \backslash \{i_1, \ldots, i_k\}$.

Это значит, что к π' можно применить предположение индукции: действительно, $\mathrm{Fix}(\pi') = \mathrm{Fix}(\pi) \cup \{i_1, \ldots, i_k\}$, поэтому мощность множества $\{i \in \{1, \ldots, n\} \mid \pi'(i) \neq i$ на k меньше, чем мощность аналогичного множества для π. По предположению индукции π' можно записать в виде произведения независимых циклов, носители которых не пересекаются с $\mathrm{Fix}(\pi')$: $\pi' = \tau_1 \ldots \tau_s$. После этого остается записать $\pi = \sigma\pi' = \sigma\tau_1 \ldots \tau_s$ и заметить, что носитель цикла σ — это множество $\{i_1, \ldots, i_k\}$, не пересекающееся с $\mathrm{Fix}(\pi) = \mathrm{Fix}(\pi') \backslash \{i_1, \ldots, i_k\}$. □

Определение 10.9.6. Запись элемента $\pi \in S_n$ в виде, указанном в теореме, называется **цикленной записью перестановки** π.

Пример 10.9.7. Цикленные записи нетождественных перестановок из S_3 выглядят так: $(1\ 2)$, $(1\ 3)$, $(2\ 3)$, $(1\ 2\ 3)$, $(1\ 3\ 2)$. Цикленная запись тождественной перестановки пуста. В S_4 имеются три перестановки, в цикленной записи которых более одного цикла: $(1\ 2)(3\ 4)$, $(1\ 3)(2\ 4)$, $(1\ 4)(2\ 3)$.

Замечание 10.9.8. Как мы видели выше (замечание 10.9.2), запись цикла в виде $(i_1\ i_2\ \dots\ i_k)$ не вполне однозначна: на первое место можно поставить любой элемент из $i_1, \dots, i_k$. Кроме того, в произведении нескольких независимых циклов их можно переставлять местами произвольным образом (независимые циклы коммутируют). Несложно понять, что в остальном циклическая запись перестановки единственна. Действительно, каждое число от 1 до n либо не встречается ни в одном из циклов (и тогда это неподвижная точка), либо встречается ровно в одном цикле (поскольку циклы независимы), и тогда его образ однозначно определен. Часто для удобства в каждом цикле $(i_1\ i_2\ \dots\ i_k)$ на первое место ставят минимальный элемент из $i_1, \dots, i_k$, а все циклы в цикленной записи располагают в порядке возрастания первых элементов этих циклов.

Цикленная запись полезна, среди прочего, для визуализации сопряжения перестановки.

Лемма 10.9.9. *Пусть* $\pi \in S_n$, $i_1, \dots, i_k$ — *различные элементы* $\{1, \dots, n\}$. *Тогда*

$$^{\pi}(i_1\ i_2\ \dots\ i_k) = (\pi(i_1)\ \pi(i_2)\ \dots\ \pi(i_k)).$$

Таким образом, сопряженный элемент к циклу длины k *также является циклом длины* k.

Доказательство. Пусть $\pi' = {}^{\pi}(i_1\ i_2\ \dots\ i_k)$. Применяя π' к $\pi(i_s)$, получаем $\pi'(\pi(i_s)) = (\pi \circ (i_1\ i_2\ \dots\ i_k))(i_s) = \pi(i_{s+1})$ при $s < k$ и $\pi(i_1)$ при $s = k$. Если же $j \in \{1, \dots, n\}$ не совпадает ни с одним из $\pi(i_1), \dots, \pi(i_k)$, то $\pi^{-1}(j)$ не совпадает ни с одним из $i_1, \dots, i_k$, поэтому $\pi'(j) = (\pi \circ (i_1\ i_2\ \dots\ i_k))(\pi^{-1}(j)) = \pi(\pi^{-1}(j)) = j$. Значит, элементы $\pi(i_1), \dots, \pi(i_k)$ под действием π' сдвигаются по циклу (в указанном порядке), а остальные остаются на месте. □

Определение 10.9.10. Пусть $\pi \in S_n$. Набор длин циклов в цикленной записи π (с учетом кратностей) называется **цикленным типом** перестановки π. Так, к примеру, цикленный тип перестановки $(1\ 2\ 3)$ равен $\{3\}$, а перестановки $(1\ 2)(3\ 4)$ — $\{2, 2\}$.

Теорема 10.9.11. *Цикленные типы двух сопряженных перестановок одинаковы. Обратно, если у двух перестановок цикленные типы совпадают, то они сопряжены.*

Доказательство. Если $\pi, \rho \in S_n$ и $\rho = \rho_1\rho_2 \dots \rho_s$ — разложение перестановки ρ в произведение независимых циклов, то ${}^{\pi}\rho = \pi\rho\pi^{-1} = \pi\rho_1\rho_2 \dots \rho_s\pi^{-1} = \pi\rho_1\pi^{-1}\pi\rho_2\pi^{-1} \dots \pi\rho_s\pi^{-1} = {}^{\pi}\rho_1 \cdot {}^{\pi}\rho_2 \cdot \dots \cdot {}^{\pi}\rho_s$. Поскольку при сопряжении цикла получается цикл той же длины, первая часть теоремы доказана.

Пусть теперь $\rho = \rho_1\rho_2\dots\rho_s$ и $\tau = \tau_1\tau_2\dots\tau_t$ — разложения перестановок из S_n в произведения независимых циклов с одинаковым цикленным типом. Это означает, что $s = t$ и после перестановки сомножителей можно считать, что циклы ρ_i и τ_i имеют одинаковую длину для всех $i = 1,\dots,s$. Укажем перестановку $\pi \in S_n$ такую, что $\tau = {}^\pi\rho$. Пусть цикл ρ_1 имеет вид $\rho_1 = (i_1\ \ i_2\ \ \dots\ \ i_k)$, а цикл τ_1 имеет вид $\tau_1 = (j_1\ \ j_2\ \ \dots\ \ j_k)$. Положим $\pi(i_1) = j_1$, $\pi(i_2) = j_2$, $\dots$, $\pi(i_k) = j_k$. Совершим такую же процедуру с циклами ρ_2 и τ_2, $\dots$, ρ_s и τ_s. Заметим, что все элементы, входящие в записи циклов $\rho_1,\rho_2,\dots,\rho_s$ попарно различны, так что противоречия не возникнет. Кроме того, все элементы, входящие в записи циклов $\tau_1,\tau_2,\dots,\tau_s$ попарно различны, так что пока что π принимает различные значения, которых столько же, сколько всего элементов в циклах $\rho_1,\rho_2\dots,\rho_s$. Для элементов $j \in \{1,\dots,n\}$, которые не входят ни в один из циклов $\rho_1,\rho_2,\dots,\rho_s$, положим $\pi(j)$ равным произвольным различным элементам, не входящим ни в один из циклов $\tau_1,\tau_2,\dots,\tau_s$. Это можно сделать, поскольку их поровну. Легко видеть, что мы получили биекцию $\pi \in S_n$ и в силу леммы 10.9.9 имеем ${}^\pi\rho_i = \tau_i$ для всех $i = 1,\dots,n$. Поэтому и ${}^\pi\rho = \tau$. □

Замечание 10.9.12. Из доказательства теоремы 10.9.11 видно, что искомая перестановка π, как правило, далеко не единственна.

Следующая теорема показывает, что изучение симметрических групп может быть важным шагом в изучении всех конечных групп.

Теорема 10.9.13 (Теорема Кэли). *Любая конечная группа* G *изоморфна некоторой подгруппе группы* S_n *для некоторого натурального* n.

Доказательство. Положим $n = |G|$. Занумеруем элементы группы G числами от 1 до n: $G = \{g_1,\dots,g_n\}$. Сопоставим каждому элементу $g \in G$ перестановку $\pi_g \in S_n$ следующим образом: для $i = 1,\dots,n$ посмотрим на элемент gg_i в группе G. Этот элемент должен иметь некоторый номер; его и возьмем в качестве $\pi_g(i)$. Таким образом, $gg_i = g_{\pi_g(i)}$ для всех i. Прежде всего, нужно показать, что π_g действительно является перестановкой. Инъективность π_g показать легко: если $\pi_g(i) = \pi_g(j)$, то $gg_i = gg_j$, откуда $g_i = g_j$ и $i = j$. Биективность теперь следует из того, что π_g действует на конечном множестве $\{1,\dots,n\}$ (принцип Дирихле).

Мы построили по каждому элементу $g \in G$ перестановку $\pi_g \in S_n$; покажем теперь, что соответствие $\pi\colon g \mapsto \pi_g$ является гомоморфизмом групп. Необходимо показать, что $\pi_{gg'} = \pi_g \circ \pi'_g$. Но для каждого $i = 1,\dots,n$ имеем $(gg')g_i = g_{\pi_{gg'}(i)}$; с другой стороны, $g(g'g_i) = gg_{\pi_{g'}(i)} = g_{\pi_g(\pi_{g'}(i))}$. Поэтому $\pi_{gg'}(i) = \pi_g(\pi_{g'}(i))$ для всех i, что и требовалось.

Наконец, гомоморфизм π инъективен, поскольку из $\pi_g = \pi_h$ следует $gg_1 = g_{\pi_g(1)} = g_{\pi_h(1)} = hg_1$ и, после сокращения на g_1, $g = h$. Мы построили инъективный гомоморфизм $\pi\colon G \to S_n$; его образ $\mathrm{Im}(\pi)$ по теореме о гомоморфизме 10.5.3 изоморфен фактору G по ядру гомоморфизма π, которое тривиально. Поэтому группа $\mathrm{Im}(\pi)$ изоморфна G и является подгруппой в S_n. □

10.10 Диэдральная группа

Литература: [К3], гл. 1, § 4, п. 5.

Рассмотрим на эвклидовой плоскости правильный n-угольник с вершинами $A_1, \dots, A_n$ и центром в начале координат (точке O). Множество всех поворотов плоскости, переводящих этот n-угольник в себя, образует группу (см. пример 10.1.3 (8)). Нетрудно понять, что это циклическая группа: в качестве образующей можно взять поворот с центром в O на угол $2\pi/n$ в положительном направлении (whatever this means). Обозначим этот поворот через x. Любой поворот, переводящий n-угольник в себя, должен переводить вершины в вершины: пусть он переводит A_1 в A_k. Тогда A_2 переходит в A_{k+1}, и так далее (если считать, что вершины занумерованы в положительном направлении, и номера понимаются по модулю n, то есть, $A_{n+1} = A_1$, $A_{n+2} = A_2$, $\dots$). Таким образом, этот поворот совпадает с x^k.

Рассмотрим теперь множество *всех движений* плоскости, переводящих наш правильный n-угольник в себя. Это тоже группа; обозначим ее через D_n. Она содержит в качестве подгруппы, порожденной элементом x, циклическую группу порядка n. Кроме того, в ней содержатся некоторые осевые симметрии: их описание зависит от четности n. Для нечетного n ось каждой симметрии проходит через вершину и середину противоположной ей стороны (например, через вершину A_1 и середину стороны $A_{\frac{n+1}{2}}A_{\frac{n+3}{2}}$): таких симметрий n. Для четного n имеется $n/2$ симметрий относительно прямых, соединяющих противоположные вершины (например, $A_1A_{\frac{n}{2}+1}$), и $n/2$ симметрий относительно прямых, соединяющих середины противоположных сторон (например, середину стороны A_1A_2 с серединой стороны $A_{\frac{n}{2}+1}A_{\frac{n}{2}+2}$). В любом случае, всего осевых симметрий ровно n, и можно показать, что они вместе с n поворотами исчерпывают все элементы группы D_n. Таким образом, $|D_n| = 2n$.

Для подробного изучения группы D_n мы будем пользоваться ее *матричным представлением*. А именно, заметим, что все описанные повороты и симметрии сохраняют точку O. Движение эвклидовой плоскости, сохраняющее точку O, является, среди прочего, линейным отображением соответствующего двумерного векторного пространства. Поэтому после выбора ортогонального базиса можно отождествить элементы группы D_n с их матрицами в этом базисе. Нетрудно понять, что

$$x = \begin{pmatrix} \cos(2\pi/n) & \sin(2\pi/n) \\ -\sin(2\pi/n) & \cos(2\pi/n) \end{pmatrix},$$

и поэтому

$$x^k = \begin{pmatrix} \cos(2\pi k/n) & \sin(2\pi k/n) \\ -\sin(2\pi k/n) & \cos(2\pi k/n) \end{pmatrix}.$$

Удобно считать, что вершины нашего многоугольника — это в точности корни степени n из единицы (см. замечание 3.5.3): $1, \varepsilon, \varepsilon^2, \dots, \varepsilon^{n-1}$. Тогда одна из осевых симметрий, лежащих в D_n — это просто комплексное сопряжение; обозначим эту симметрию через y:

$$y = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

Группа D_n также должна содержать элементы вида yx^k для $k = 1, \dots, n-1$:

$$yx^k = \begin{pmatrix} \cos(2\pi k/n) & \sin(2\pi k/n) \\ \sin(2\pi k/n) & -\cos(2\pi k/n) \end{pmatrix}.$$

Теперь можно забыть про школьную геометрию и определить группу D_n как множество, состоящее из матриц x^k и yx^k, где $k = 0, \ldots, n-1$.

Теорема 10.10.1. *Множество* $D_n = \{x^k \mid 0 \leqslant k \leqslant n-1\} \cup \{yx^k \mid 0 \leqslant k \leqslant n-1\}$ *(матрицы* x, y *указаны выше) является группой относительно обычного умножения матриц (и, таким образом, подгруппой в* $\mathrm{GL}(2, \mathbb{R})$*). Группа* D_n *порождена двумя элементами* x *и* y*;* $\mathrm{ord}_{D_n}(x) = n$, $\mathrm{ord}_{D_n}(y) = 2$*. Подгруппа* $\langle x \rangle \leqslant D_n$ *циклическая порядка* n*; она нормальна в* D_n*.*

Доказательство. Прямое вычисление показывает, что $x^n = 1$ и $y^2 = 1$; более того, порядок x равен n. Показатель степени x теперь можно воспринимать по модулю n: $x^m = x^{m \bmod n} \in D_n$. Кроме того, $yxy = x^{-1}$, откуда $xy = yx^{-1}$ и, итерируя, получаем $x^k y = yx^{-k}$. Поэтому $x^k \cdot x^l = x^{k+l}$, $yx^k \cdot x^l = yx^{k+l}$, $x^k \cdot yx^l = yx^{-k}x^l = yx^{l-k}$, $yx^k \cdot yx^l = yyx^{-k}x^l = x^{l-k}$. Наконец, отсюда следует, что $(x^k)^{-1} = x^{-k}$ и $(yx^k)^{-1} = yx^k$. Мы получили, что умножение и взятие обратного не выводит нас за пределы множества D_n; поэтому $D_n \leqslant \mathrm{GL}(2, \mathbb{R})$. В частности, D_n является группой. По определению каждый элемент D_n записан в виде произведения некоторого количества элементов x и y, поэтому $D_n = \langle x, y \rangle$. Из того, что $\mathrm{ord}_{D_n}(x) = n$, следует, что $\langle x \rangle$ — циклическая порядка n. Наконец, $yx^l \cdot x^k \cdot (yx^l)^{-1} = yx^l \cdot x^k \cdot yx^l = yx^l \cdot yx^{l-k} = x^{l-k-l} = x^{-k} \in \langle x \rangle$, поэтому $\langle x \rangle \trianglelefteqslant D_n$ (впрочем, нормальность следует и из примера 10.3.11 (3): $\langle x \rangle$ имеет индекс 2 в D_n). □

Замечание 10.10.2. Обозначим $\langle y \rangle = G$, $\langle x \rangle = H$. Тогда $D_n = GH$: любой элемент D_n можно записать (и даже единственным образом) в виде gh, где $g \in G$, $h \in H$. Кроме того, $G \cap H = \{e\}$. Более того, группа D_n/H состоит из двух элементов, потому она циклическая (теорема 10.7.8) и изоморфна G. Однако, D_n не является прямым произведением G и H (при $n > 2$): не хватает условия 3 из теоремы 10.8.3. Еще один аргумент: подгруппа $G = \langle y \rangle$ не нормальна в D_n ($xyx^{-1} = yx^{-2} \notin \langle y \rangle$) а сомножители должны быть нормальны в прямом произведении (предложение 10.8.2, пункт 2).

11 Полилинейная алгебра

11.1 Полилинейные отображения

Литература: [КМ], ч. 2, § 2, п. 1; ч. 4, § 1, пп. 1–2.

Пусть k — поле, $V_1, \dots, V_m, U$ — векторные пространства над k. Отображение $f\colon V_1 \times \dots \times V_m \to U$ называется **полилинейным**, если оно линейно по каждому аргументу при фиксированных значениях остальных. Иными словами, f **аддитивно** по каждому аргументу:

$$f(\dots, v_i' + v_i'', \dots) = f(\dots, v_i', \dots) + f(\dots, v_i'', \dots).$$

Кроме того, отображение f **однородно степени 1** по каждому аргументу (также при фиксированных остальных):

$$f(\dots, \lambda v_i, \dots) = \lambda f(\dots, v_i, \dots).$$

Приведем примеры полилинейных отображений, которые мы встречали раньше:

- Скалярное произведение: билинейная форма $B\colon V \times V \to R$ является полилинейным отображением по самому определению (см. определение 9.1.1).
- Определитель: пусть $V = k^n$ — пространство столбцов высоты n. Можно рассмотреть отображение
$$\det\colon k^n \times \dots \times k^n \to k, \quad (v_1, \dots, v_n) \mapsto \det(v_1, \dots, v_n),$$
сопоставляющий набору столбцов определитель матрицы, составленной из этих столбцов. Это отображение полилинейно (см. раздел 5.7).

Оказывается, что полилинейные отображения из $V_1 \times \dots V_m$ в U в точности соответствуют *линейным* отображениям из некоторого нового объекта (тензорного произведения пространств $V_1, \dots, V_m$) в U.

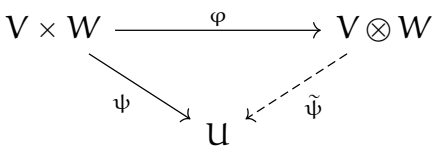
11.2 Тензорное произведение двух пространств

Литература: [F], гл. XIV, § 4, пп. 1, 2; [K2], гл. 6, § 1, п. 5; [КМ], ч. 4, § 1, пп. 2–5.

Определение 11.2.1. Пусть V, W — векторные пространства над полем k. **Тензорным произведением** пространств V и W называется векторное пространство $V \otimes W$ вместе с билинейным отображением $\varphi\colon V \times W \to V \otimes W$, удовлетворяющие следующему *универсальному свойству*:

- для любого векторного пространства U и любого билинейного отображения $\psi\colon V \times W \to U$ существует единственное линейное отображение $\tilde{\psi}\colon V \otimes W \to U$ такое, что $\psi = \tilde{\psi} \circ \varphi$.

Универсальное свойство можно изобразить следующей диаграммой:



Теорема 11.2.2. *Тензорное произведение любых векторных пространств V, W над полем k существует и единственно с точностью до канонического изоморфизма. Последнее означает, что если $\overline{\varphi}\colon V \times W \to V\overline{\otimes}W$ — еще одно тензорное произведение в смысле определения 11.2.1, то существует единственный изоморфизм векторных пространств $\alpha\colon V \otimes W \to V\overline{\otimes}W$ такой, что $\overline{\varphi} = \alpha \circ \varphi$:*

$$\begin{array}{ccc} V \times W & \xrightarrow{\quad\varphi\quad} & V \otimes W \\ & {\scriptstyle\overline{\varphi}}\searrow \quad \swarrow{\scriptstyle\alpha} & \\ & V\overline{\otimes}W & \end{array}$$

Доказательство. Сначала докажем единственность. Итак, пусть $\varphi\colon V \times W \to V \otimes W$ и $\overline{\varphi}\colon V \times W \to V\overline{\otimes}W$ — два тензорных произведения пространств V и W. Рассмотрим следующую диаграмму:

$$\begin{array}{ccc} V \times W & \xrightarrow{\quad\varphi\quad} & V \otimes W \\ & {\scriptstyle\overline{\varphi}}\searrow & \\ & V\overline{\otimes}W & \end{array}$$

Поскольку $V \otimes W$ является тензорным произведением V и W, можно подставить в универсальное свойство $U = V\overline{\otimes}W$ и $\psi = \overline{\varphi}$. Значит, существует единственное линейное отображение $\alpha\colon V \otimes W \to V\overline{\otimes}W$, для которого $\overline{\varphi} = \alpha \circ \varphi$. Осталось доказать, что α является изоморфизмом. Для этого мы построим отображение, обратное к α. Рассмотрим диаграмму

$$\begin{array}{ccc} V \times W & \xrightarrow{\quad\overline{\varphi}\quad} & V\overline{\otimes}W \\ & {\scriptstyle\varphi}\searrow & \\ & V \otimes W & \end{array}$$

Поскольку $V\overline{\otimes}W$ также является тензорным произведением V и W, можно подставить в универсальное свойство $U = V \otimes W$ и $\psi = \varphi$. Значит, существует единственное линейное отображение $\beta\colon V\overline{\otimes}W \to V \otimes W$ такое, что $\varphi = \beta \circ \overline{\varphi}$. Покажем, что β является обратным к α. Рассмотрим диаграмму

$$\begin{array}{ccc} V \times W & \xrightarrow{\quad\varphi\quad} & V \otimes W \\ & {\scriptstyle\varphi}\searrow & \\ & V \otimes W & \end{array}$$

Из универсального свойства для $V \otimes W$ следует, что существует единственное линейное отображение $V \otimes W \to V \otimes W$, композиция которого с φ равна φ. Но мы знаем два таких отображения: одно из них тождественное, $\mathrm{id}_{V\otimes W}$, а другое равно композиции $\beta \circ \alpha$. Действительно, $(\beta \circ \alpha) \circ \varphi = \beta \circ \overline{\varphi} = \varphi$. Из единственности в универсальном свойстве следует, что эти отображения должны совпадать. Поэтому $\beta \circ \alpha = \mathrm{id}_{V\otimes W}$. Аналогичное соображение для $V\overline{\otimes}W$ показывает, что $\alpha \circ \beta = \mathrm{id}_{V\overline{\otimes}W}$.

Для доказательства существования тензорного произведения мы приведем явную конструкцию. Рассмотрим вспомогательное векторное пространство L, базис которого состоит из

всевозможных выражений вида «$v \otimes w$» для всех векторов $v \in V$, $w \in W$. Иными словами, L — это множество всех [конечных] формальных линейных комбинаций выражений вида «$v \otimes w$» (с коэффициентами из k) с очевидными операциями суммы и умножения на скаляры.

Несложно определить отображение $f\colon V \times W \to L$: положим $f(v, w) = «v \otimes w»$. Однако, это отображение не является билинейным: например, $f(v_1 + v_2, w) = «(v_1 + v_2) \otimes w»$, в то время как $f(v_1, w) + f(v_2, w) = «v_1 \otimes w» + «v_2 \otimes w»$. В нашем пространстве $«(v_1 + v_2) \otimes w» \neq «v_1 \otimes w» + «v_2 \otimes w»$, поскольку равенство означало бы наличие линейной комбинации между базисными элементами. Кроме того, $f(\lambda v, w) = «(\lambda v) \otimes w»$, но $\lambda f(v, w) = \lambda «v \otimes w»$. Для того, чтобы исправить это, мы профакторизуем по всем таким соотношениям, и в полученном фактор-пространстве нужные выражения совпадут. А именно, обозначим через R линейную оболочку в L следующих векторов:

$$\begin{aligned}
&«(v_1 + v_2) \otimes w» - «v_1 \otimes w» - «v_2 \otimes w»,\\
&«(\lambda v) \otimes w» - \lambda «v \otimes w»,\\
&«v \otimes (w_1 + w_2)» - «v \otimes w_1» - «v \otimes w_2»,\\
&«v \otimes (\lambda w)» - \lambda «v \otimes w»
\end{aligned}$$

для всех $v_1, v_2, v, w_1, w_2, w \in V$ и $\lambda \in k$. Рассмотрим фактор-пространство L/R и покажем, что оно удовлетворяет определению тензорного произведения V и W. Нам еще нужно построить билинейное отображение $\varphi\colon V \times W \to L/R$; для этого рассмотрим композицию f и канонической проекции $\pi\colon L \to L/R$. Проверим, что φ билинейно. Например, $\varphi(v_1 + v_2, w) - \varphi(v_1, w) - \varphi(v_2, w) = \pi(«(v_1 + v_2) \otimes w») - \pi(«v_1 \otimes w») - \pi(«v_2 \otimes w») = \pi(«(v_1 + v_2) \otimes w» - «v_1 \otimes w» - «v_2 \otimes w») = 0$, поскольку выражение в скобках лежит в R. Аналогично проверяется однородность и линейность по второму аргументу.

Наконец, проверим универсальное свойство. Пусть $\psi\colon V \times W \to U$ — билинейное отображение. По универсальному свойству базиса (теорема 7.1.9) существует единственное линейное отображение $\psi'\colon L \to U$ такое, что $\psi = \psi' \circ f$. Для того, чтобы это отображение «пропустить» через фактор-пространство L/R, достаточно проверить, что отображение ψ' переводит каждый элемент R в 0 (в этом случае отображение $L/R \to U$, $x + R \mapsto \psi'(x)$ корректно определено). Но для этого достаточно проверить, что ψ' переводит каждый элемент из нашей системы, порождающей пространство R, в 0. Это очевидно в силу билинейности ψ; например,

$$\begin{aligned}
\psi'(«(v_1 + v_2) \otimes w» - «v_1 \otimes w» - «v_2 \otimes w») &= \psi'(f(v_1 + v_2, w) - f(v_1, w) - f(v_2, w))\\
&= \psi'(f(v_1 + v_2, w)) - \psi'(f(v_1, w)) - \psi'(f(v_2, w))\\
&= \psi(v_1 + v_2, w) - \psi(v_1, w) - \psi(v_2, w)\\
&= 0.
\end{aligned}$$

Таким образом, мы построили отображение $\widetilde{\psi}\colon L/R = V \otimes W \to U$, для которого $\widetilde{\psi} \circ \varphi = \psi$. Для доказательства единственности осталось заметить, что элементы вида $\varphi(v, w)$ для $u \in V$, $w \in W$ являются образами в L/R базисных элементов пространства L. Поэтому такие элементы порождают $U \otimes V$. Значит, линейное отображение $\widetilde{\psi}\colon V \otimes W \to U$ полностью определяется своими значениями на таких элементах: $\widetilde{\psi}(\varphi(v, w)) = \psi(v, w)$. □

Итак, мы построили векторное пространство $V \otimes W$ вместе с билинейным отображением $\varphi\colon V \times W \to V \otimes W$. Слово «универсальность» в названии универсального свойства означает, что билинейное отображение φ универсально среди всех билинейных отображений из $V \times W$ в следующем смысле: любое билинейное отображение из $V \times W$ пропускается через φ (является композицией φ и некоторого линейного отображения).

Элементы пространства $V \otimes W$ называются **тензорами**. Образ пары (v, w) под действием φ мы будем обозначать через $v \otimes w \in V \otimes W$ и называть **разложимым тензором**. Из определения немедленно следует, что $(v_1 + v_2) \otimes w = v_1 \otimes w + v_2 \otimes w$, $v \otimes (w_1 + w_2) = v \otimes w_1 + v \otimes w_2$, $(\lambda v) \otimes w = \lambda(v \otimes w) = u \otimes (\lambda v)$. Заметим, однако, что (как правило) не любой тензор является разложимым. В то же время, множество всех разложимых тензоров является системой образующих пространства $V \otimes W$, поскольку это образы базисных элементов пространства L в нашей конструкции. В частности, любой тензор является *суммой* конечного числа разложимых. Поэтому, например, для задания линейного отображения из $V \otimes W$ достаточно задать его на разложимых тензорах (на самом деле, это еще одна переформулировка универсального свойства). Точнее, если мы сопоставили каждому разложимому тензору $v \otimes w \in V \otimes W$ некоторый элемент пространства U *билинейным образом*, то однозначно определено линейное отображение $V \otimes W \to U$.

Отметим, что приведенная в доказательстве теоремы 11.2.2 конструкция совершенно чудовищна: даже если пространства V и W конечномерны, по пути к $V \otimes W$ мы строим пространство L, которое, как правило, бесконечномерно: даже если $\dim(V) = \dim(W) = 1$ и $k = \mathbb{R}$, базис пространства L имеет мощность континуума. На самом деле, тензорное произведение конечномерных пространств конечномерно; если в пространствах V и W выбраны базисы, то и в $V \otimes W$ естественным образом возникает базис.

Предложение 11.2.3. *Пусть V, W — векторные пространства над полем k, и пусть $\mathcal{B} = \{e_1, \ldots, e_m\}$ — базис V, $\mathcal{C} = \{f_1, \ldots, f_n\}$ — базис W. Тогда элементы вида $e_i \otimes f_j$, $1 \leqslant i \leqslant m$, $1 \leqslant j \leqslant n$, образуют базис пространства $V \otimes W$.*

Доказательство. Рассмотрим пространство X размерности mn, базис которого состоит из элементов вида $e_i \otimes f_j$. Сейчас мы определим билинейное отображение $V \otimes W \to X$ и проверим, что X вместе с этим отображением удовлетворяет универсальному свойству тензорного произведения.

Для определения φ сначала положим $\varphi(e_i, f_j) = e_i \otimes f_j$. Для двух произвольных векторов $v = \sum_i \lambda_i e_i \in V$ и $w = \sum_j \mu_j f_j \in W$ теперь определим $\varphi(u, v)$ так, чтобы φ было билинейным. Раскрывая скобки, получаем, что $\varphi(u, v) = \sum_{i,j} \lambda_i \mu_j e_i \otimes f_j$. Очевидно, что построенное отображение $\varphi\colon U \times V \to X$ билинейно.

Пусть теперь U — еще одно векторное пространство над k, и пусть $\psi\colon V \times W \to U$ — билинейное отображение. Так как векторы $e_i \otimes f_j$ образуют базис пространства X, для определения линейного отображения $\widetilde{\psi}\colon X \to U$ мы можем задать его значения на этих векторых произвольным образом; полученное линейное отображение определяется этим однозначно (теорема 7.1.9). Поэтому положим $\widetilde{\psi}(e_i \otimes f_j) = \psi(e_i, f_j)$ и продолжим $\widetilde{\psi}$ до линейного отображения $X \to U$. Композиция $\widetilde{\psi} \circ \varphi$ билинейна и совпадает с ψ на парах (e_i, f_j), поэтому $\widetilde{\psi} \circ \varphi = \psi$.

Вместе с тем, любое отображение, композиция которого с φ равна ψ, должно на базисных векторах $\varphi(e_i, f_j)$ принимать значения $\psi(e_i, f_j)$, поэтому такое отображение единственно. $\square$

Определение 11.2.4. Базис из предложения 11.2.3 называется **тензорным базисом** пространства $U \otimes V$. Обычно мы упорядочиваем его следующим (*лексикографическим*) образом: $e_1 \otimes f_1$, $e_1 \otimes f_2$, …, $e_1 \otimes f_n$, …, $e_m \otimes f_1$, $e_m \otimes f_2$, …, $e_m \otimes f_n$.

Следствие 11.2.5. *Если пространства V, W над полем k конечномерны, то $V \otimes W$ конечномерно и $\dim(V \otimes W) = \dim(V) \cdot \dim(W)$.*

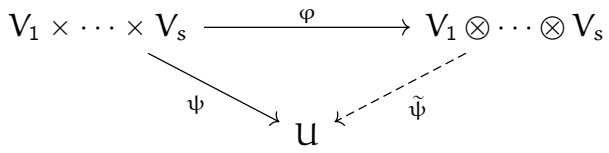
Замечание 11.2.6. Сравните формулу для размерности тензорного произведения с формулой для прямой суммы: $\dim(V \oplus W) = \dim(V) + \dim(W)$. Это свидетельство того, что тензорное произведение и прямая сумма — аналоги умножения и сложения для векторных пространств.

11.3 Тензорное произведение нескольких пространств

Литература: [F], гл. XIV, § 4, п. 3; [KM], ч. 4, § 1, пп. 2–5; § 2, пп. 1–3.

Мы можем теперь попытаться определить тензорное произведение *трех* пространств U, V, W формулой $U \otimes V \otimes W = (U \otimes V) \otimes W$. Однако, такое определение нарушает симметрию между U, V и W (почему не $U \otimes (V \otimes W)$?). Поэтому мы просто повторим универсальное определение тензорного произведения, изменив его соответствующим образом.

Пусть $V_1, \dots, V_s$ — векторные пространства над полем k. Тогда их **тензорным произведением** называется векторное пространство $V_1 \otimes \dots \otimes V_s$ над k вместе с полилинейным отображением $\varphi\colon V_1 \times \dots \times V_s \to V_1 \otimes \dots \otimes V_s$ таким, что для любого полилинейного отображения $\psi\colon V_1 \times \dots \times V_s \to U$ в некоторое векторное пространство W существует единственное линейное отображение $\tilde{\psi}\colon V_1 \otimes \dots \otimes V_s \to U$ такое, что $\psi = \tilde{\psi} \circ \varphi$:



Теорема 11.3.1. *Тензорное произведение любого конечного числа векторных пространств $V_1, \dots, V_s$ существует и единственно с точностью до канонического изоморфизма.*

Доказательство. Доказательство этой теоремы совершенно такое же, как в случае двух пространств (теорема 11.2.2). А именно, рассмотрим векторное пространство L с базисом, состоящим из элементов «$v_1 \otimes \dots \otimes v_s$», где $v_1, \dots, v_s$ пробегают всевозможные наборы элементов пространств $V_1, \dots, V_s$, соответственно. Имеется естественное отображение множеств $V_1 \times \dots \times V_s \to L$, переводящее набор $(v_1, \dots, v_s)$ в базисный элемент «$v_1 \otimes \dots \otimes v_s$». Чтобы сделать это отображение полилинейным, профакторизуем L по линейной оболочке R следующих элементов:

$$\begin{aligned}
&«\dots \otimes v_i + v_i' \otimes \dots» - «\dots \otimes v_i \otimes \dots» - «\dots \otimes v_i' \otimes \dots»;\\
&«\dots \otimes \lambda v_i \otimes \dots» - \lambda«\dots \otimes v_i \otimes \dots».
\end{aligned}$$

Теперь сквозное отображение $\varphi\colon V_1 \times \cdots \times V_s \to L \to L/R$ полилинейно. Проверим, что оно универсально: пусть $\psi\colon V_1 \times \cdots \times V_s \to U$ — некоторое полилинейное отображение. Сопоставление «$v_1 \otimes \cdots \otimes v_s$» $\mapsto \psi(v_1, \ldots, v_s)$ задает линейное отображение $L \to U$, и элементы, порождающие R, переходят в 0 в силу полилинейности ψ. Поэтому оно пропускается через фактор-пространство и мы получаем линейное отображение $L/R \to U$. Таким образом, мы можем положить $V_1 \otimes \cdots \otimes V_s = L/R$. Единственность тензорного произведения доказывается буквально так же, как и в случае двух пространств. □

Замечание 11.3.2. Как и в случае двух пространств, образ набора $(v_1, \ldots, v_s) \in V_1 \times \cdots \times V_s$ в пространстве $V_1 \otimes \cdots \otimes V_s$ обозначается через $v_1 \otimes \cdots \otimes v_s$ и называется **разложимым тензором**; для задания линейного отображения из $V_1 \otimes \cdots \otimes V_s$ в U достаточно определить его на разложимых тензорах билинейным образом. Проиллюстрируем это на примере доказательства следующей теоремы.

Предложение 11.3.3. *Тензорное произведение векторных пространств ассоциативно и коммутативно с точностью до канонических изоморфизмов: а именно, для любых трех векторных пространств U, V, W имеют место канонические изоморфизмы $(U \otimes V) \otimes W \cong U \otimes V \otimes W \cong U \otimes (V \otimes W)$ и $U \otimes V \cong V \otimes U$.*

Доказательство. Определим отображение $U \otimes V \otimes W \to (U \otimes V) \otimes W$ на разложимых тензорах формулой $u \otimes v \otimes w \mapsto (u \otimes v) \otimes w$. Эта формула задает линейные отображения, и той же формулой, прочитанной справа налево, задается отображение в обратную сторону. Очевидно, что композиция этих отображений $U \otimes V \otimes W \to (U \otimes V) \otimes W \to U \otimes V \otimes W$ тождественна на разложимых тензорах, и потому тождественна на всем пространстве. Аналогично доказывается изоморфизм $U \otimes V \otimes W \cong U \otimes (V \otimes W)$. Для задания отображения $U \otimes V \to V \otimes U$ отправим $u \otimes v$ в $v \otimes u$; доказательство завершается так же. □

Предложение 11.3.4. *Пусть $V_1, \ldots, V_s$ — векторные пространства над полем k размерностей $n_1, \ldots, n_s$; $\mathcal{B}_j = \{e_1^j, \ldots, e_{n_j}^j\}$ — базис V_i для каждого $j = 1, \ldots, s$. Тогда элементы вида $e_{i_1}^1 \otimes \cdots \otimes e_{i_s}^s$, где $1 \leqslant i_k \leqslant n_k$ для всех $k = 1, \ldots, s$, образуют базис пространства $V_1 \otimes \cdots \otimes V_s$.*

Доказательство. Мы можем повторить доказательство предложения 11.2.3. А именно, рассмотрим векторное пространство W над k, базисом которого являются формальные символы вида $e_{i_1}^1 \otimes \cdots \otimes e_{i_s}^s$. Определим полилинейное отображение $\varphi\colon V_1 \times \cdots \times V_s \to W$ следующим образом: набор базисных векторов $(e_{i_1}^1, \ldots, e_{i_s}^s) \in V_1 \times \cdots \times V_s$ отправим в базисный элемент $e_{i_1}^1 \otimes \cdots \otimes e_{i_s}^s$, а дальше продолжим по полилинейности. А именно, если $(v_1, \ldots, v_s) \in V_1 \times \cdots \times V_s$ — набор векторов, разложим каждый v_j по базису $\mathcal{B}_j$. Получим равенства вида $v_j = \sum_{i_j=1}^{n_j} e_{i_j}^j a_{i_j,j}$.

Положим

$$\begin{aligned}\varphi(v_1,\dots,v_s) &= \varphi\Big(\sum_{i_1=1}^{n_1} e^1_{i_1} a_{i_1,1},\dots,\sum_{i_s=1}^{n_s} e^s_{i_s} a_{i_s,s}\Big)\\ &= \sum_{i_1=1}^{n_1}\dots\sum_{i_s=1}^{n_s} a_{i_1,1}\dots a_{i_s,s}\varphi(e^1_{i_1},\dots,e^s_{i_s})\\ &= \sum_{i_1=1}^{n_1}\dots\sum_{i_s=1}^{n_s} a_{i_1,1}\dots a_{i_s,s} e^1_{i_1}\otimes\dots\otimes e^s_{i_s}.\end{aligned}$$

Очевидно, что это отображение полилинейно; покажем, что пространство W вместе с φ удовлетворяет универсальному свойству из определения тензорного произведения. Пусть U — произвольное векторное пространство над k, и $\psi\colon V_1\times\dots\times V_s\to U$ — полилинейное отображение. Покажем, что оно представляется в виде композиции φ и некоторого линейного отображения $\widetilde{\psi}$. Для задания $\widetilde{\psi}\colon W\to U$ достаточно задать его (произвольным образом) на базисе, то есть, на элементах вида $e^1_{i_1}\otimes\dots\otimes e^s_{i_s}$. Это можно сделать единственным образом: положим $\widetilde{\psi}(e^1_{i_1}\otimes\dots\otimes e^s_{i_s}) = \psi(e^1_{i_1},\dots,e^s_{i_s})$. Композиция $\widetilde{\psi}\circ\varphi$, разумеется, является полилинейным отображением и совпадает с ψ на наборах вида $(e^1_{i_1},\dots,e^s_{i_s})$, и цепочка равенств выше показывает, что значение полилинейного отображения на произвольном наборе $(v_1,\dots,v_s)$ выражается через его значения на наборах такого вида. Поэтому $\widetilde{\psi}\circ\varphi$ совпадает с ψ. □

11.4 Двойственное пространство

Литература: [vdW], гл. IV, § 21; [KM], ч. 1, § 1, п. 9.

Пусть V — векторное пространство над полем k. Рассмотрим k как [одномерное] векторное пространство над k. Тогда множество $\mathrm{Hom}(V,k)$ линейных отображений из V в k (*линейных функций* на V) само является векторным пространством над k (см. раздел 7.2). Операции на нем вполне естественны: сложение функций и умножение функций на скаляры. Это пространство мы будем обозначать через $V^* = \mathrm{Hom}(V,k)$ и называть **пространством, двойственным к** V

Пусть теперь V — *конечномерное* векторное пространство над k и $\mathcal{B} = (e_1,\dots,e_n)$ — базис V. По универсальному свойству базиса (теорема 7.1.9) для задания элемента $\varphi\in V^* = \mathrm{Hom}(V,k)$ достаточно задать (произвольным образом) элементы $\varphi(e_1),\dots,\varphi(e_n)\in k$.

Предложение 11.4.1. *Пусть V — векторное пространство над k с базисом $\mathcal{B} = (e_1,\dots,e_n)$. Обозначим через e^*_i функцию $V\to k$, равную 1 на базисном векторе e_i и 0 на всех остальных базисных векторах. Таким образом, $e^*_i(e_i) = 1$ и $e^*_i(e_j) = 0$ при всех $j\neq i$. Тогда $(e^*_1,\dots,e^*_n)$ — базис пространства V^*.*

Доказательство. Пусть $\varphi\colon V\to k$ — произвольный элемент пространства V^*. Мы знаем (теорема 7.1.9), что задать φ — это то же самое, что задать значения $\varphi(e_1),\dots,\varphi(e_n)\in k$. Рассмотрим функцию $\varphi(e_1)e^*_1+\dots+\varphi(e_n)e^*_n$. Покажем, что она совпадает с φ. Действительно, для базисного вектора e_i получаем $(\varphi(e_1)e^*_1+\dots+\varphi(e_n)e^*_n)(e_i) = \varphi(e_1)e^*_1(e_i)+\dots+\varphi(e_1)e^*_n(e_i) =$

$\varphi(e_i)e_i^*(e_i) = \varphi(e_i)$. Значит, функции $\varphi(e_1)e_1^* + \cdots + \varphi(e_n)e_n^*$ и φ совпадают на базисных векторах, а потому совпадают везде. Значит, мы представили функцию φ как линейную комбинацию функций e_i^*. Осталось показать, что функции e_i^* линейно независимы.

Действительно, предположим, что $c_1e_1^* + \cdots + c_ne_n^* = 0$ — нетривиальная линейная комбинация. Это означает, что $c_i \neq 0$ при некотором i. Но тогда и $(c_1e_1^* + \cdots + c_ne_n^*)(e_i) = 0$, а левая часть равна $c_1e_1^*(e_i) + \cdots + c_ne_n^*(e_i) = c_i \neq 0$ — противоречие. □

Таким образом, в конечномерном случае пространства V и V^* имеют одинаковую размерность. Из этого следует, что они изоморфны (теорема 7.5.4). Например, имеется изоморфизм $V \to V^*$, отправляющий e_i в φ_i при $i = 1, \ldots, n$, если $e_1, \ldots, e_n$ — базис V. Однако, этот изоморфизм не является каноническим, то есть, существенно зависит от выбора базиса. В то же время, *дважды двойственное* пространство $V^{**} = \mathrm{Hom}(V^*, k)$ *канонически* изоморфно V.

Предложение 11.4.2. *Рассмотрим отображение $V \to V^{**}$, сопоставляющее вектору $v \in V$ функцию $v^{**}\colon V^* \to k$, заданную равенством $v^{**}(\varphi) = \varphi(v)$ для всех $\varphi \in V^*$. Если пространство V конечномерно, то указанное отображение является изоморфизмом.*

Доказательство. Нетрудно проверить, что v^{**} является линейным отображением $V^* \to k$. Действительно, если $\varphi, \psi \in V^*$, $\lambda \in k$, то $v^{**}(\varphi+\psi) = (\varphi+\psi)(v) = \varphi(v)+\psi(v) = v^{**}(\varphi)+v^{**}(\psi)$ и $v^{**}(\lambda\varphi) = (\lambda\varphi)(v) = \lambda \cdot \varphi(v) = \lambda \cdot v^{**}(\varphi)$.

Таким образом, $v^{**} \in V^{**}$ для всех $v \in V$. Покажем, что сопоставление $v \mapsto v^{**}$ линейно зависит от v. Необходимо проверить, что $(v+w)^{**} = v^{**}+w^{**}$ и $(\lambda v)^{**} = \lambda v^{**}$. Чтобы проверить совпадение двух отображений $V^* \to k$, достаточно проверить, что результаты их применения к произвольному элементу $\varphi \in V^*$ совпадают: $(v + w)^{**}(\varphi) = \varphi(v + w) = \varphi(v) + \varphi(w) = v^{**}(\varphi) + w^{**}(\psi)$, $(\lambda v)^{**}(\varphi) = \varphi(\lambda v) = \lambda \cdot \varphi(v) = \lambda \cdot v^{**}(\varphi)$.

Мы получили линейное отображение $V \to V^{**}$. Покажем, что оно инъективно. Для этого достаточно проверить, что его ядро тривиально. Пусть вектор $v \in V$ таков, что $v^{**} = 0$. Это означает, что $v^{**}(\varphi) = 0$ для всех $\varphi \in V^*$, то есть, что $\varphi(v) = 0$ для всех $\varphi\colon V \to k$. Покажем, что из этого следует, что $v = 0$. Действительно, если $v \neq 0$, то вектор v можно дополнить до базиса $(v, e_1, e_2, \ldots)$ пространства v. Определим функцию $\varphi_v \in V^*$ равенствами $\varphi_v(v) = 1$, $\varphi_v(e_i) = 0$ для всех i. По универсальному свойству базиса этого достаточно для корректного определения линейного отображения $\varphi_v\colon V \to k$. По предположению $\varphi_v(v) = 0$, в то время как мы положили $\varphi_v(v) = 1$ — противоречие.

Наконец, воспользуемся конечномерностью: мы знаем, что $\dim(V^{**}) = \dim(V^*) = \dim(V)$, и у нас есть инъективное отображение $V \to V^{**}$. По теореме о гомоморфизме 7.3.8 из этого следует, что наше отображение сюръективно и, стало быть, является изоморфизмом векторных пространств. □

11.5 Канонические изоморфизмы

Литература: [КМ], ч. 4, § 2, пп. 4–6.

Теорема 11.5.1 (Выражение Hom через $\otimes$). *Для любых конечномерных векторных пространств U, V над k имеет место канонический изоморфизм*

$$U \otimes V \cong \mathrm{Hom}(U^*, V).$$

Доказательство. Определим отображение $\eta\colon U \otimes V \to \mathrm{Hom}(U^*, V)$, отправив разложимый тензор $u \otimes v \in U \otimes V$ в отображение $U^* \to V$, $\varphi \mapsto \varphi(u)v$. Написанная формула билинейно зависит от u и от v, поэтому корректно определяет линейное отображение из тензорного произведения $U \otimes V$.

Покажем, что η — изоморфизм. Для этого выберем базис $(f_1, \ldots, f_m)$ в U и базис $(e_1, \ldots, e_n)$ в V. При этом $\{f_j \otimes e_i\}$ — базис в $U \otimes V$ (предложение 11.2.3). Вспомним, как строится базис пространства $\mathrm{Hom}(U^*, V)$. Заметим, что в пространстве U^* у нас есть базис $(\varphi_1, \ldots, \varphi_m)$, двойственный базису $(f_1, \ldots, f_m)$. Как мы знаем из теоремы 7.5.6, после выбора базисов в U^* и V пространство $\mathrm{Hom}(U^*, V)$ оказывается изоморфно пространству матриц $M(n, m, k)$, а в этом пространстве имеется стандартный базис из матричных единиц. Матричная единица E_{ij} соответствует отображению $U^* \to V$, которое φ_j переводит в e_i, а все остальные базисные векторы φ_h, $h \neq j$, отправляет в 0. Обозначим это отображение через a_{ij}.

Мы утверждаем, что отображение η переводит $f_j \otimes e_i$ в a_{ij}. Действительно, по нашему определению $f_j \otimes e_i$ переводится в отображение $U^* \to V$, $\varphi \mapsto \varphi(f_j)e_i$. Проверим, что это и есть a_{ij}. Действительно, $\varphi_j \mapsto \varphi_j(f_j)e_i = e_i$ и $\varphi_h \mapsto \varphi_h(f_j)e_i = 0$ при $h \neq j$.

Таким образом, отображение η переводит базис пространства $U \otimes V$ в базис пространства $\mathrm{Hom}(U^*, V)$, а потому биективно. □

Следствие 11.5.2. *Для любых конечномерных векторных пространств U, V над k имеет место канонический изоморфизм*

$$U^* \otimes V \cong \mathrm{Hom}(U, V).$$

Доказательство. Применим предыдущую теорему к U^* и V: $U^* \otimes V \cong \mathrm{Hom}((U^*)^*, V) \cong \mathrm{Hom}(U, V)$. □

Следствие 11.5.3. *Для любого конечномерного векторного пространства U над k имеет место канонический изоморфизм $U \otimes k \cong U$.*

Доказательство. По теореме 11.5.1 есть канонический изоморфизм $U \otimes k \cong \mathrm{Hom}(U^*, k)$; правая часть по определению равна $(U^*)^* \cong U$. □

Теорема 11.5.4 (Двойственность и $\otimes$). *Для любых конечномерных векторных пространств U, V над k имеет место канонический изоморфизм*

$$(U \otimes V)^* \cong U^* \otimes V^*.$$

Доказательство. Зададим отображение $U^* \otimes V^* \to (U \otimes V)^*$. Как всегда, достаточно определить его на разложимых тензорах $\varphi \otimes \psi \in U^* \otimes V^*$. Образом этого тензора должен быть элемент пространства $(U \otimes V)^*$, то есть, линейное отображение $U \otimes V \to k$, которое достаточно задать

на разложимых тензорах $u \otimes v \in U \otimes V$. Отправим такой тензор в $\varphi(u)\psi(v) \in k$. Очевидно, что написанное выражение билинейно зависит от (u, v), потому определяет элемент пространства $(U \otimes V)^*$. С другой стороны, этот элемент билинейно зависит от (φ, ψ). Итак, мы построили линейное отображение $\eta\colon U^* \otimes V^* \to (U \otimes V)^*$: отправляющее $\varphi \otimes \psi$ в линейное отображение $u \otimes v \mapsto \varphi(u)\psi(v)$.

Покажем, что построенное отображение является изоморфизмом. Для этого выберем базис $(f_1, \ldots, f_m)$ в пространстве U и базис $(e_1, \ldots, e_n)$ в пространстве V. Тогда в пространствах U^* и V^* возникают двойственные базисы: $(f_1^*, \ldots, f_m^*)$ и $(e_1^*, \ldots, e_n^*)$, соответственно. Поэтому в пространстве $U^* \otimes V^*$ естественно взять тензорное произведение этих двойственных базисов $(f_j^* \otimes e_i^*)$. С другой стороны, в пространстве $(U \otimes V)^*$ естественно выбрать базис, двойственный к тензорному произведению исходных базисов U и V: $(f_j \otimes e_i)^*$.

Покажем, что при нашем линейном отображении η базисный элемент $f_j^* \otimes e_i^*$ переходит в базисный элемент $(f_j \otimes e_i)^*$. Действительно, по определению $\eta(f_j^* \otimes e_i^*)$ — это линейное отображение, отправляющее $u \otimes v$ в $f_j^*(u)e_i^*(v)$. Если мы подставим в него $u = f_j$ и $v = e_i$, то получим $f_j^*(f_j)e_i^*(e_i) = 1$; если же подставим любую другую пару $u = f_k$, $v = e_h$ (где $k \neq j$ или $h \neq i$), то получим $f_j^*(f_k)e_i^*(e_h) = 0$, поскольку хотя бы один сомножитель равен нулю. Значит, $\eta(f_j^* \otimes e_i^*)$ переводит базисный элемент $f_j \otimes e_i \in U \otimes V$ в 1, а все остальные базисные элементы в 0. Но $(f_j \otimes e_i)^*$ действует ровно так же на базисных элементах, поэтому $\eta(f_j^* \otimes e_i^*) = (f_j \otimes e_i)^*$, что и требовалось. Таким образом, η переводит базис в базис, и потому является изоморфизмом. □

Следствие 11.5.5. *Для любых конечномерных векторных пространств $U_1, \ldots, U_s$ над k имеет место канонический изоморфизм*

$$(U_1 \otimes \cdots \otimes U_s)^* \cong U_1^* \otimes \cdots \otimes U_s^*.$$

Доказательство. По индукции из теоремы 11.5.4 и предложения 11.3.3. □

Теорема 11.5.6 (Сопряженность $\otimes$ и Hom). *Для любых конечномерных векторных пространств U, V, W над k имеет место канонический изоморфизм*

$$\mathrm{Hom}(U \otimes V, W) \cong \mathrm{Hom}(U, \mathrm{Hom}(V, W)).$$

Доказательство. Заметим сначала, что размерности обеих частей равны $\dim(U) \cdot \dim(V) \cdot \dim(W)$. Рассмотрим произвольный элемент $\varphi\colon \mathrm{Hom}(U, \mathrm{Hom}(V, W))$. Он сопоставляет (линейным образом) каждому элементу $u \in U$ некоторое линейное отображение $\varphi_u\colon V \to W$, $v \mapsto \varphi_u(v)$. Построим теперь по этому элементу φ линейное отображение из $U \otimes V$ в W следующим образом: разложимый тензор $u \otimes v \in U \otimes V$ отправим в $\varphi_u(v) \in W$. Это сопоставление билинейно зависит от u и от v, (поскольку φ и φ_u линейны), и потому мы получили однозначно определенное линейное отображение $\eta(\varphi)\colon U \otimes V \to W$, то есть, элемент $\mathrm{Hom}(U \otimes V, W)$. При этом сопоставление $\varphi \mapsto \eta(\varphi)$ является, очевидно, линейным. Наконец, покажем, что η является инъекцией. Предположим, что $\eta(\varphi) = 0$, то есть, $\eta(\varphi)(u \otimes v) = 0$ для всех $u \in U$, $v \in V$. Но по нашему определению $\eta(\varphi)(u \otimes v) = \varphi_u(v)$; поэтому $\varphi_u(v) = 0$ при всех $u \in U$, $v \in V$, откуда $\varphi_u = 0$ при всех $u \in U$, откуда $\varphi = 0$. Теперь из инъективности η и совпадения размерностей следует, что η и сюръективно, а потому является изоморфизмом. □

На самом деле в доказательстве этой теоремы можно было, как и раньше, выбрать базисы в U, V, W, получить базисы во всех фигурирующих в формулировке пространствах, и честно проверить, что построенное отображение η переводит базис в базис. Еще один вариант доказательства теоремы 11.5.6 — воспользоваться уже доказанными изоморфизмами: $\mathrm{Hom}(U \otimes V, W) \cong (U \otimes V)^* \otimes W \cong (U^* \otimes V^*) \otimes W \cong U^* \otimes (V^* \otimes W) \cong U^* \otimes \mathrm{Hom}(V, W) \cong \mathrm{Hom}(U, \mathrm{Hom}(V, W))$

11.6 Тензорное произведение линейных отображений

Литература: [К2], гл. 6, § 1, пп. 2, 5; [КМ], ч. 4, § 2, п. 7.

Пусть $\varphi\colon U \to V$, $\psi\colon W \to Z$ — линейные отображения. Сейчас мы определим их **тензорное произведение** $\varphi \otimes \psi$, которое будет линейным отображением из $U \otimes W$ в $V \otimes Z$. Сопоставим разложимому тензору $u \otimes w \in U \otimes W$ разложимый тензор $\varphi(u) \otimes \psi(w) \in V \otimes Z$. Нетрудно видеть, что это сопоставление ведет себя билинейно по u и по w, и потому задает корректно определенное линейное отображение

$$\varphi \otimes \psi\colon U \otimes W \to V \otimes Z.$$

Покажем, что это определение обладает естественными свойствами.

Теорема 11.6.1. *Тензорное произведение линейных отображение обладает следующими свойствами:*

1. $(\varphi'\varphi) \otimes (\psi'\psi) = (\varphi' \otimes \psi')(\varphi \otimes \psi)$;
2. $\mathrm{id}_U \otimes \mathrm{id}_V = \mathrm{id}_{U \otimes V}$;
3. $(\varphi + \varphi') \otimes \psi = \varphi \otimes \psi + \varphi' \otimes \psi$;
4. $\varphi \otimes (\psi + \psi') = \varphi \otimes \psi + \varphi \otimes \psi'$;
5. $(\lambda\varphi) \otimes \psi = \lambda(\varphi \otimes \psi) = \varphi \otimes (\lambda\psi)$.

Доказательство. Мы проверим самое сложное свойство — первое. Пусть $U \xrightarrow{\varphi} V \xrightarrow{\varphi'} V'$, $W \xrightarrow{\psi} Z \xrightarrow{\psi'} Z'$ — линейные отображения. Выберем векторы $u \in U$, $w \in W$ и применим $(\varphi'\varphi) \otimes (\psi'\psi)$ к разложимому тензору $u \otimes w$. По определению получаем

$$((\varphi'\varphi) \otimes (\psi'\psi))(u \otimes w) = (\varphi'\varphi)(u) \otimes (\psi'\psi)(w) = \varphi'(\varphi(u)) \otimes \psi'(\psi(w)).$$

С другой стороны,

$$(\varphi' \otimes \psi')(\varphi \otimes \psi)(u \otimes w) = (\varphi' \otimes \psi')(\varphi(u) \otimes \psi(w)) = \varphi'(\varphi(u)) \otimes \psi'(\psi(w)).$$

Значит, два указанных отображения совпадают на всех разложимых тензорах, а потому равны. □

Теорема 11.6.2. *Для любых конечномерных векторных пространств* U, V, W, Z *над* k *имеет место канонический изоморфизм*

$$\operatorname{Hom}(U \otimes W, V \otimes Z) \cong \operatorname{Hom}(U, V) \otimes \operatorname{Hom}(W, Z).$$

Доказательство. Мы построили отображение $\operatorname{Hom}(U, V) \times \operatorname{Hom}(W, Z) \to \operatorname{Hom}(U \otimes W, V \otimes Z)$, $(\varphi, \psi) \mapsto \varphi \otimes \psi$. По теореме 11.6.1 это сопоставление билинейно, поэтому определяет линейное отображение $\operatorname{Hom}(U, V) \otimes \operatorname{Hom}(W, Z) \to \operatorname{Hom}(U \otimes W, V \otimes Z)$, и обычные рассуждения (например, выбор базисов во всех указанных пространствах) убеждают нас, что получился изоморфизм. Еще один способ доказательства — воспользоваться уже доказанными изоморфизмами:

$$\operatorname{Hom}(U \otimes W, V \otimes Z) \cong (U \otimes W)^* \otimes (V \otimes Z) \cong (U^* \otimes V) \otimes (W^* \otimes Z) \cong \operatorname{Hom}(U, V) \otimes \operatorname{Hom}(W, Z).$$

□

Выясним, как выглядит матрица тензорного произведения линейных отображений. Пусть вообще $x \in M(l, m, k)$, $y \in M(n, p, k)$ — две произвольные матрицы над полем k. Определим **кронекерово произведение** матриц x и y как матрицу $x \otimes y \in M(lm, np, k)$, которую проще всего представлять себе блочной матрицей

$$x \otimes y = \begin{pmatrix} x_{11}y & \dots & x_{1m}y \\ \vdots & \ddots & \vdots \\ x_{l1}y & \dots & x_{lm}y \end{pmatrix}.$$

Обратите внимание, что кронекерово произведение матриц мы обозначаем тем же значком $\otimes$, что и тензорное произведение. Это не случайно: заметим пока, что кронекерово произведение обладает многими обычными свойствами тензорного произведения.

Предложение 11.6.3 (Свойства кронекерова произведения).

1. Ассоциативность*:* $(x \otimes y) \otimes z = x \otimes (y \otimes z)$ *(после забывания блочных структур).*
2. Дистрибутивность относительно сложения*:* $(x+y)\otimes z = x\otimes z + y\otimes z$, $x\otimes(y+z) = x\otimes y + x\otimes z$.
3. Однородность*:* $(\alpha x) \otimes y = \alpha(x \otimes y) = x \otimes (\alpha y)$.
4. Взаимная дистрибутивность кронекерова произведения и умножения*:* $(xy) \otimes (uv) = (x \otimes u)(y \otimes v)$.

Доказательство. Все эти свойства легко проверяются прямым вычислением. □

Наконец, мы готовы показать, что матрица тензорного произведения линейных отображений является кронекеровым произведением матриц этих отображений. Для простоты мы ограничимся случаем линейных операторов (то есть, квадратных матриц). Рассмотрим линейные операторы $\varphi\colon U \to U$, $\psi\colon V \to V$ на конечномерных пространствах U, V. Как обычно, после выбора базисов $(e_1, \dots, e_m)$ в U и $(f_1, \dots, f_n)$ в V мы можем считать, что $U = k^m$,

$V = k^n$ — пространства столбцов. В этом случае векторы $u \in U$, $v \in V$ истолковываются как столбцы высоты m и n, соответственно, а линейный оператор — как умножение на квадратную матрицу: если a, b — матрицы операторов φ, ψ в выбранных базисах, получаем линейные отображения

$$\varphi\colon U \to U, u \mapsto au,$$

где $a \in M(m, k)$, и

$$\psi\colon V \to V, v \mapsto bv,$$

где $b \in M(n, k)$.

В пространстве $U \otimes V$ имеется тензорный базис $(e_i \otimes f_j)$, в котором mn элементов. Он позволяет отождествить $U \otimes V$ с k^{mn}. При нашем упорядочивании тензорного базиса (см. определение 11.2.4) это отождествление выглядит следующим образом. Пусть $u = \sum_i u_i e_i$, $v = \sum_j v_j f_j$. Тогда $u \otimes v = (\sum_i u_i e_i) \otimes (\sum_j v_j f_j) = \sum_{i,j} u_i v_j (e_i \otimes f_j)$. Это означает, что

$$\begin{pmatrix} u_1 \\ \dots \\ u_m \end{pmatrix} \otimes \begin{pmatrix} v_1 \\ \dots \\ v_n \end{pmatrix} = \begin{pmatrix} u_1 v_1 \\ \dots \\ u_1 v_n \\ u_2 v_1 \\ \dots \\ u_m v_1 \\ \dots \\ u_m v_n \end{pmatrix}.$$

Теорема 11.6.4. *Если матрица оператора φ в базисе (e_i) равна a, а матрица оператора ψ в базисе (f_j) равна b, то матрица оператора $\varphi \otimes \psi$ в тензорном базисе $(e_i \otimes f_j)$ равна кронекеровому произведению $a \times b$.*

Доказательство. Пусть $u \in U$, $v \in V$ — произвольные векторы. По определению тензорное произведение отображений φ и ψ действует на разложимый тензор $u \otimes v \in U \otimes V$ следующим образом: $(\varphi \otimes \psi)(u \otimes v) = \varphi(u) \otimes \psi(v)$. С другой стороны, кронекерово произведение $a \otimes b$ умножается на столбец $u \otimes v$ следующим образом: $(a \otimes b)(u \otimes v) = (au \otimes bv)$ — здесь мы воспользовались свойством 4 из предложения 11.6.3. Но при наших отождествлениях $au = \varphi(u)$, $bv = \psi(v)$. Поэтому отображение $\varphi \otimes \psi$ совпадает с умножением на матрицу $a \otimes b$ на разложимых тензорах, а значит и везде. □

11.7 Тензорные пространства

Литература: [F], гл. XIV, § 4, п. 4; [K2], гл. 6, § 1, п. 1; [vdW], гл. IV, § 24; [KM], ч. 4, § 3, пп. 1–2.

Пусть V — конечномерное векторное пространство над полем k, и $V^* = \mathrm{Hom}(V, k)$ — двойственное к нему. В ближайших параграфах мы будем изучать векторные пространства

$$T^p_q(V) = \underbrace{V \otimes \dots \otimes V}_{p \text{ раз}} \otimes \underbrace{V^* \otimes \dots \otimes V^*}_{q \text{ раз}}.$$

Пространство $T^p_q(V)$ традиционно называется пространством q раз ковариантных и p раз контравариантных тензоров, или просто **тензорным пространством** (если из контекста понятно, о каких значениях p, q идет речь). Элементы тензорных пространств называются **тензорами** над V. Если $x \in T^p_q(V)$, то пара (p, q) называется **типом** тензора x, p называется его **контравариантной валентностью**, а q — его **ковариантной валентностью**. Сумма $p+q$ называется **полной валентностью**. Если $p = 0$, тензор x называется **чисто ковариантным**, а если $q = 0$ — **чисто контравариантным**.

На самом деле, нам уже встречались тензоры небольшой валентности:

- При $p = q = 0$ удобно считать, что $T^0_0(V) = k$; тензоры типа $(0, 0)$ — это просто скаляры.
- $T^1_0(V) = V$ — векторы;
- $T^0_1(V) = V^*$ — ковекторы;
- $T^2_0(V) = V \otimes V = (V^* \otimes V^*)^* = \mathrm{Hom}(V^* \otimes V^*, k)$. Напомним, что (по определению тензорного произведения) линейные отображения из $V^* \otimes V^*$ в k — это то же самое, что *билинейные* отображения из $V^* \times V^*$ в k. Поэтому тензоры типа $(2, 0)$ можно интерпретировать как билинейные формы на V^*.
- $T^1_1(V) = V \otimes V^* = \mathrm{Hom}(V, V)$ — линейные операторы на V.
- $T^0_2(V) = V^* \otimes V^* = (V \otimes V)^* = \mathrm{Hom}(V \otimes V, k)$. Как и в случае тензоров типа $(2, 0)$, заметим, что линейные отображения из $V \otimes V$ в k — это в точности билинейные отображения из $V \times V$ в k. Поэтому тензоры типа $(0, 2)$ можно интерпретировать как билинейные формы на V.
- $T^1_2(V) = V \otimes V^* \otimes V^* = (V \otimes V)^* \otimes V = \mathrm{Hom}(V \otimes V, V)$; то есть, тензоры типа $(1, 2)$ — это билинейные отображения из $V \times V$ в V; при желании можно это интерпретировать как задание умножения на векторах, дистрибутивного относительно суммы.

11.8 Тензоры в классических обозначениях

Литература: [F], гл. XIV, § 1; [K2], гл. 6, § 1, пп. 3, 4; [KM], ч. 4, § 4, пп. 1–4.

В прикладной математике и инженерных науках все встречающиеся тензоры (тензор деформации, тензор электромагнитного поля, тензор инерции, тензор Эйнштейна...) возникают почти исключительно в координатной записи. Напомним, что если в пространстве V выбран базис $\mathcal{E} = (e_1, \ldots, e_n)$, то в двойственном пространстве возникает двойственный базис $(e^*_1, \ldots, e^*_n)$. Для того, чтобы приблизить наши обозначения к традиционным, мы будем обозначать двойственный базис через $(e^1, \ldots, e^n)$. Каждый вектор $v \in V$ можно разложить по базису $\mathcal{E}$:

$$v = \sum e_i v^i = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix},$$

а каждый ковектор $\varphi \in V^*$ — по двойственному базису:

$$\varphi = \sum \varphi_i e^i = \begin{pmatrix} \varphi_1 & \dots & \varphi_n \end{pmatrix} \begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix}.$$

При этом в тензорном пространстве T^p_q (для произвольных p, q) возникает тензорный базис, состоящий из векторов вида $e_{i_1} \otimes \dots \otimes e_{i_p} \otimes e^{j_1} \otimes \dots \otimes ej_q$, где $1 \leqslant i_1, \dots, i_p, j_1, \dots, j_q \leqslant n$. Таким образом, каждый тензор $x \in T^p_q(V)$ можно единственным образом записать в виде

$$x = \sum_{\substack{i_1,\dots,i_p \\ j_1,\dots,j_q}} x^{i_1\dots i_p}_{j_1\dots j_q} e_{i_1} \otimes \dots \otimes e_{i_p} \otimes e^{j_1} \otimes e^{j_q},$$

где $x^{i_1\dots i_p}_{j_1\dots j_q} \in k$ — координаты тензора в этом базисе. Традиционно тензор задавался явным перечислением своих координат. При этом, поскольку этот набор зависит от выбора базиса, приходится указывать, как же преобразуются координаты тензора при другом выборе базиса.

Для этого выберем в V другой базис $\mathcal{F} = (f_1, \dots, f_n)$, который будет называться *новым* (в отличие от *старого* базиса $\mathcal{E} = (e_1, \dots, e_n)$). Напомним, что мы изучали, как связаны координаты векторов в этих базисах, с помощью [обратимой] матрицы перехода $C = (\mathcal{E} \rightsquigarrow \mathcal{F})$ (см. определение 7.9.1):

$$\begin{pmatrix} f_1 & \dots & f_n \end{pmatrix} = \begin{pmatrix} e_1 & \dots & e_n \end{pmatrix} \cdot C.$$

Вспомним, как преобразуются координаты вектора $v = \sum_i e_i v^i$ при замене базиса:

$$v = \begin{pmatrix} e_1 & \dots & e_n \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} e_1 & \dots & e_n \end{pmatrix} \cdot C \cdot C^{-1} \cdot \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} f_1 & \dots & f_n \end{pmatrix} \cdot C^{-1} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}.$$

Таким образом, при переходе в новый базис столбец координат вектора умножается на C^{-1}. Это означает (см. замечание 7.9.4), что координаты вектора преобразуются *контравариантным образом*; именно поэтому число p в определении тензорного пространства $T^p_q(V)$ называется контравариантной валентностью. В то же время координаты *ковектора* преобразуются *ковариантным образом*. Действительно, по определению двойственного базиса

$$e^i(e_j) = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}.$$

Это означает, что

$$\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} \cdot \begin{pmatrix} e_1 & \dots & e_n \end{pmatrix} = \begin{pmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{pmatrix} = E.$$

и аналогично для базиса $\mathcal{F}$. Домножим последнее равенство на C^{-1} слева и на C справа:

$$C^{-1}\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} \cdot \begin{pmatrix} e_1 & \dots & e_n \end{pmatrix} C = C^{-1}EC = E.$$

В левой части стоит $C^{-1}\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} \cdot \begin{pmatrix} f_1 & \dots & f_n \end{pmatrix}$, поэтому

$$C^{-1}\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} = \begin{pmatrix} f^1 \\ \vdots \\ f^n \end{pmatrix}.$$

Это и означает, что двойственный базис преобразуется с помощью матрицы C^{-1}, а потому координаты ковекторов преобразуются с помощью матрицы $(C^{-1})^{-1} = C$. Это несложно проверить и непосредственно: если $\varphi = \sum \varphi_i e^i$, то

$$\varphi = \begin{pmatrix} \varphi_1 & \dots & \varphi_n \end{pmatrix}\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} = \begin{pmatrix} \varphi_1 & \dots & \varphi_n \end{pmatrix} \cdot C \cdot C^{-1} \cdot \begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} = \begin{pmatrix} \varphi_1 & \dots & \varphi_n \end{pmatrix} C \cdot \begin{pmatrix} f^1 \\ \vdots \\ f^n \end{pmatrix}.$$

У нас все готово к тому, чтобы выяснить, как меняются координаты произвольного тензора при замене базиса. Пусть

$$x = \sum_{\substack{i_1,\dots,i_p \\ j_1,\dots,j_q}} y^{i_1\dots i_p}_{j_1\dots j_q} f_{i_1} \otimes \dots \otimes f_{i_p} \otimes f^{j_1} \otimes \dots \otimes f^{j_q}$$

— выражение того же тензора x в новом тензорном базисе. Мы хотим выразить $\left(y^{i_1\dots i_p}_{j_1\dots j_q}\right)$ через $\left(x^{i_1\dots i_p}_{j_1\dots j_q}\right)$. В следующей теореме удобно элемент матрицы C, стоящий на пересечении i-й строки и j-го столбца записывать как C^i_j, а не C_{ij}.

Теорема 11.8.1. *Пусть $C = (C^i_j)$ — матрица перехода от старого базиса к новому, $\widetilde{C} = (\widetilde{C}^i_j) = C^{-1}$ — обратная к ней. Тогда координаты тензора $x \in T^p_q(V)$ в новом тензорном базисе следующим образом выражаются через его координаты в старом тензорном базисе:*

$$y^{i_1\dots i_p}_{j_1\dots j_q} = \sum_{\substack{h_1,\dots,h_p \\ k_1,\dots,k_q}} \widetilde{C}^{i_1}_{h_1} \dots \widetilde{C}^{i_p}_{h_p} C^{k_1}_{j_1} \dots C^{k_q}_{j_q} x^{h_1\dots h_p}_{k_1\dots k_q}$$

Доказательство. Достаточно доказать эту формулу для разложимых тензоров, а в этом случае нужно применить формулы преобразования координат векторов и ковекторов в каждом из сомножителей. □

Иными словами, координаты тензора преобразуются контравариантно (при помощи матрицы C^{-1}) по контравариантным сомножителям, и ковариантно (при помощи матрицы C) по ковариантным сомножителям.

Предметный указатель

аддитивное отображение, 217
аддитивность
  определителя, 87
  производной, 53
аксиома выбора, 11, 109
алгебраическое дополнение, 94
алгорифм Эвклида, 59
аргумент, 39
  главное значение, 39
ассоциативность, 14
  обобщенная, 15
  в группе, 81, 197
ассоциированность
  целых чисел, 17
  многочленов, 48
база индукции, 13
базис
  ортогональный, 171
  ортонормированный, 171
  относительный, 130
биекция, 9
билинейная форма, 165
блочная матрица, 80
блочная структура, 80
цикл, 212
цикленная запись перестановки, 213
четная перестановка, 84
число инверсий перестановки, 84
делимость
  целых чисел, 17
  многочленов, 48
делитель
  наибольший общий
    нескольких чисел, 21
  наибольший общий
    целых чисел, 18
  общий, 18
делитель нуля, 30
  нетривиальный, 30
  тривиальный, 30
детерминант, 86
длина вектора, 168
дробь, 61
единица, 14
  левая, 14
  правая, 14
  в кольце, 28
единичный элемент
  в группе, 81, 197
эндоморфизм
  векторных пространств, 113
фактор-множество, 13
форма
  эрмитова, 167
  неотрицательно определенная, 166, 167
  положительно определенная, 166, 167
  полуторалинейная, 166
формальное равенство многочленов, 51
формулы Крамера, 95
функциональное равенство многочленов, 51
функция Эйлера, 32
гомоморфизм
  групп, 204
  тривиальный, 204
  векторных пространств, 113
график, 11
группа, 81, 197
  абелева, 197
  циклическая, 208
  кольца, аддитивная, 197
  коммутативная, 197
  обратимых элементов кольца, 198
  перестановок, 81, 197
  полная линейная, 198
  поля, мультипликативная, 197
  симметрическая, 197

специальная линейная, 198
углов, 198, 207
знакопеременная, 199
характеристика поля, 55
индекс подгруппы, 208
индукционный переход, 13
инъекция, 9
интерполяционная задача, 56
интерполяционный многочлен
Лагранжа, 57
Ньютона, 57
инверсия, 84
изометрия, 191
каноническая проекция, 8, 13
каноническая запись многочлена, 47
каноническое разложение, 24
класс эквивалентности, 13
класс вычетов, 27
коэффициенты матрицы, 71
кольцо, 28
классов вычетов, 29
квадратных матриц, 75
нулевое, 29
коммутативность, 14
комплексное число, 37
алгебраическая форма записи, 37
экспоненциальная форма, 44
тригонометрическая форма, 39
композиция, 8
корень
первообразный, 42
степени n, 42
корень многочлена, 50
кратный, 52
кратности m, 52
простой, 52
кососимметричность определителя, 88
кронекерово произведение, 228
линейная комбинация, 104
линейная независимость
над подпространством, 130
линейное представление НОД, 19
многочленов, 59
линейность
определителя, 88
матрица, 71
единичная, 74
эрмитова, 170
квадратная, 72
обратимая, 75
окаймленная единичная, 78
оператора, 136
ортогональная, 173
перехода, 131
присоединенная, 94
ранга 1, 126
расширенная, 67
симметрическая, 170
системы линейных уравнений, 67
ступенчатая, 70
транспонированная, 72
унитарная, 173
взаимная, 94
матричная единица, 75
минор
дополнительный, 93
мнимая единица, 37
мнимая часть, 37
многочлен, 45
неприводимый, 60
модуль, 38
наибольший общий делитель, 18
многочленов, 58
нечетная перестановка, 84
нейтральный элемент, 14
левый, 14
правый, 14
неподвижные точки перестановки, 212
независимые циклы, 212
носитель цикла, 212
нуль
в кольце, 28

область целостности, 30
область определения, 8
область значений, 8
обратимый элемент, 15
  слева, 15
  справа, 15
обратимое отображение, 10
  двусторонне, 10
  слева, 10
  справа, 10
обратный элемент, 15
  левый, 14
  правый, 14
  в группе, 81, 197
обратное отображение, 10
  левое, 10
  правое, 10
образ, 8
общий делитель
  многочленов, 58
однородное отображение, 217
ограничение, 8
операция
  ассоциативная, 14
  бинарная, 14
  коммутативная, 14
оператор, 136
  кососимметрический, 189
  линейный, 136
  неотрицательно определенный, 193
  нормальный, 183
  ортогональный, 189
  положительно определенный, 193
  самосопряженный, 189
  сохраняет скалярное произведение, 191
  унитарный, 189
определитель, 86
ортогональные векторы, 165
ортогональное дополнение, 176
основная теорема алгебры, 51
отношение, 12
  бинарное, 12
  эквивалентности, 12
  рефлексивное, 12
  симметричное, 12
  транзитивное, 12
отображение, 8, 12
перестановка, 81
подгруппа, 199
  нормальная, 203
  порожденная подмножеством, 200
  тривиальная, 199
подпространство, 100
поле, 29
  алгебраически замкнутое, 51
  частных, 61
  рациональных функций, 62
полилинейное отображение, 217
полиномиальная функция, 49
порождающая система
  над подпространством, 130
порождающее множество, 201
порядок
  элемента в группе, 208
  квадратной матрицы, 72
позиция элемента в матрице, 72
правильная дробь, 63
принцип математической индукции, 13
производная, 53
прообраз, 8
простейшая дробь, 63
простое число, 23
пространство
  эвклидово, 166
  унитарное, 167
противоположный элемент, 28
прямое произведение
  групп, 210
ранг, 127
  линейного отображения, 134
ранг матрицы
  строчный, 125

ранг матрицы
    столбцовый, 125
    тензорный, 126
разложение определителя
    по столбцу, 94
    по строке, 94
размерность, 110
решение системы линейных уравнений, 67
система линейных уравнений, 67
система линейных уравнений
    совместная, 127
система образующих
    над подпространством, 130
система порождающих, 201
скаляр, 98
соотношения ортогональности, 94
сопряжение, 37
    матрицы, 136
    в группе, 203
сопряженное отображение, 179
сравнение по модулю
    подпространства, 128
сравнимость по модулю, 26
степень многочлена, 47
свободные переменные, 71
сюръекция, 9
табличная запись перестановки, 82
тензор, 220, 230
    чисто контравариантный, 230
    чисто ковариантный, 230
    разложимый, 220, 222
тензорный базис, 221
тензорное произведение, 217
    линейных отображений, 227
    нескольких пространств, 221
тензорное пространство, 230
тип тензора, 230
тождественная перестановка, 81, 197
тождественное отображение, 8
тождество Лейбница, 53
транспонирование, 72
транспозиция, 83
    элементарная, 83
угол между векторами, 169
умножение перестановок, 81
валентность
    контравариантная, 230
    ковариантная, 230
    полная, 230
ведущие элементы, 71
вектор, 98
векторное пространство
    столбцов матрицы, 125
векторное пространство, 98
    бесконечномерное, 110
    двойственное, 223
    строк матрицы, 125
вещественная часть, 37
взаимная простота, 20
зависимые переменные, 71
значение многочлена, 49
знак перестановки, 84