Алгебра и теория чисел*

Александр Лузгарев

Содержание

*Конспект лекций для механиков, 2012–2013 учебный год; предварительная версия

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 $\{s \in S \mid s$ 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 $A \setminus B$ состоит из элементов A, не лежащих в B: $A \setminus B = \{x \in A \mid x \notin B\}$. Иначе говоря, $x \in A \setminus B$ 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: $A \Delta B = (A \cup B) \setminus (A \cap B)$.

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 **ограничениеограничение** 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.

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 1.3.4. 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).

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. 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 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.

Принципа математической индукции равносилен следующему принципу полном 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.

Ассоциативная операция удовлетворяет более сильному условию — обобщенной ассоциативности: неформально говоря, результат выполнения ассоциативной операции не зависит от порядка расстановки скобок.

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 **делится на** y (или, что y **делит** x) если существует такое целое число k, что $x = yk$. Обозначение: $x \vdots y$, $y|x$.

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

1. $x \vdots x$, $x \vdots 1$, $x \vdots (-x)$, $x \vdots (-1)$;

2. *если* $x \vdots y$ *и* $y \vdots z$, *то* $x \vdots z$ *(отношение делимости транзитивно);*

3. *если* $x \vdots z$ *и* $y \vdots z$, *то* $x + y \vdots z$;

4. *если* $x \vdots z$, *то* $xy \vdots z$;

5. *если* $z \neq 0$, *то* $xz \vdots yz$ *равносильно* $x \vdots y$;

6. $0 \vdots x$; *если* $x \vdots 0$, *то* $x = 0$.

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

2. $x = yk$, $y = zl$, поэтому $x = zlk$;

3. $x = zk$, $y = zl$, поэтому $x + y = z(k + l)$;

4. $x = zk$, поэтому $xy = zky$;

5. если $x = yk$, то $xz = yzk$; обратно, если $xz = yzk$, то $(x - yk)z = 0$, откуда $x - yk = 0$, то есть, $x = yk$;

6. $0 = x \cdot 0$; если $x = 0 \cdot k$, то $x = 0$.

□

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

Заметим, что это означает, что $x = yk$ и $y = xl$, 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 2.1.4 (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 $a \vdots d$ и $b \vdots d$.

Определение 2.2.2. Пусть $a, b \in \mathbb{Z}$. Натуральное 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 $d \vdots d'$.

Обозначение: $d = \gcd(a, b)$.

Легко видеть, что $\gcd(0, a) = |a|$; в частности, $\gcd(0, 0) = 0$. Заметим, что НОД (если он существует) единственен: действительно, если d и d' — два наибольших общих делителя a и b, то из определения следует, что $d \vdots d'$ и $d' \vdots d$, откуда $d = \pm d'$. Они натуральны, поэтому $d = d'$.

Некоторые авторы называют наибольшим общим делителем не натуральное, а *целое* число с этими свойствами. При этом наибольший общий делитель становится не единственным, а определенным с точностью до знака. Для целых чисел мы будем считать, что $\gcd$ является натуральным числом; позже, при изучении многочленов, мы перейдем на другую точку зрения и откажемся от единственности $\gcd$.

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: $a \vdots d'$ и $b \vdots d'$. Тогда по свойствам делимости $au_0 \vdots d'$, $bv_0 \vdots d'$ и $d = au_0 + bv_0 \vdots d'$, что и требовалось. □

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

Практический способ для нахождения наибольшего общего делителя — алгорифм Эвклида.

Пусть $a, b \in \mathbb{N}$. Наша цель — найти $\gcd(a, b)$. Если одно из чисел a, b равно 0, цель достигнута. Иначе пусть для определенности $a > b > 0$. 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 $b > r_0$: $b = r_0q_1 + r_1$ и так далее: $r_0 = r_1q_2 + r_2$... заметим, что максимальное число в паре всегда уменьшается; значит, процесс когда-то остановится (остаток станет равен 0). Мы утверждаем, что последний ненулевой остаток равен $\gcd(a, b)$.

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, $a \vdots d$, $b \vdots d$, откуда $r = a - bq$ делится на d, и по определению НОД $d' \vdots d$. Кроме того, $b \vdots d'$, $r \vdots d'$, откуда $a = bq + r$ делится на d', и по определению НОД $d \vdots d'$. Получили, что $d \vdots d'$ и $d' \vdots d$, но это натуральные числа; отсюда $d = d'$. □

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* $y \vdots x$.

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 $y \vdots x$ по определению. Обратно, пусть $y \vdots x$, тогда $|x|$ — общий делитель x и y, и если d' — какой-то общий делитель x, y, то, в частности, $|x| \vdots d'$. 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 $\gcd(x, y) = d$; $x \vdots d$, $y \vdots d$, откуда $zx \vdots zd$ и $zy \vdots zd$; поэтому $\gcd(zx, zy) \vdots zd$. Обратно, $zx \vdots z$, $zy \vdots z$, поэтому

$\gcd(zx, zy) \vdots z$. Запишем $\gcd(zx, zy) = zc$ для некоторого c. Значит, $zx \vdots zc$, $zy \vdots zc$, откуда $x \vdots c$ и $y \vdots c$ (поскольку $z \neq 0$). Поэтому $d = \gcd(x, y) \vdots c$, откуда $zd \vdots zc$, то есть, $zd \vdots \gcd(zx, zy)$.

□

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). *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 $ab \vdots c$ и $a \perp c$, то $b \vdots c$.*

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

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 $au_0 \vdots d$, $bv_0 \vdots d$, откуда $1 = au_0 + bv_0 \vdots d$ и $d = 1$.

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

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

□

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 $c \vdots a$, то $x = ac$, y — любое; иначе решений нет.

Обозначим $d = \gcd(a, b)$. Заметим, что $a \vdots d$, $b \vdots d$, поэтому выражение $ax + by$ должно делиться на d при всех x, y. Значит, если c не делится на d, то решений нет. Пусть теперь $c \vdots d$. Запишем $a = da'$, $b = db'$, $c = dc'$; тогда наше уравнение 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, $a(x - x_0) \vdots b$; но $a \perp b$, поэтому $x - x_0 \vdots b$. 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 $t \in \mathbb{Z}$.

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, каждое a_i делится на d);
2. если d' — общий делитель $a_1, \ldots, a_n$, то $d \vdots d'$.

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 c *делится на* $\gcd(a_1, \ldots, a_n)$.

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:

$$\gcd(a_1, a_2)y_1 + a_3y_2 + \cdots + a_ny_{n-1} = c.$$

Это линейное диофантово уравнение от $n-1$ неизвестных $y_1, \ldots, y_{n-1}$. По предположению индукции оно разрешимо тогда и только тогда, когда $c \vdots \gcd(\gcd(a_1, a_2), a_3, \ldots, a_n) = \gcd(a_1, a_2, a_3, \ldots, a_n) = d$. У нас $c \vdots d$, поэтому новое уравнение имеет решение $(y_1, \ldots, y_{n-1})$. Построим теперь решение нашего первоначального уравнения. Посмотрим на 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 $x_3 = y_2, \ldots, x_n = y_{n-1}$. Тогда

$$\begin{aligned} a_1x_1 + a_2x_2 + a_3x_3 + \cdots + a_nx_n &= \gcd(a_1, a_2)y_1 + a_3x_3 + \cdots + a_nx_n \\ &= \gcd(a_1, a_2)y_1 + a_3y_2 + \cdots + a_ny_{n-1} \\ &= c, \end{aligned}$$

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.

Предложение 2.5.2 (Свойства простых чисел). *1. если целое число n не делится на простое число p, то n и p взаимно просты;*

2. если ab делится на p, то a делится на p или b делится на p.

3. всякое целое число, большее 1, делится по крайней мере на одно простое

4. простых чисел бесконечно много.

5. *если* p_1 *и* p_2 *— два различных простых числа, то они взаимно просты.*

6. *если произведение нескольких целых чисел делится на простое число* p*, то хотя бы одно из них делится на* p*.*

Доказательство. 1. Пусть n не делится на p и $d = \gcd(n, p)$. При этом $p \vdots d$, поэтому d либо ассоциировано с p, либо ассоциировано с 1. Заметим, что n делится на d, поэтому если d ассоциировано с p, то n делится на p — противоречие. Значит, d ассоциировано с 1, откуда $n \perp p$.

2. Пусть ab делится на p и a не делится на p. По предыдущему свойству $a \perp p$, и по свойству взаимно простых чисел получаем, что $b \vdots p$.

3. Пусть $n > 1$. Если n простое, доказывать нечего. Если же n не простое, то $n = m_1 n_1$ 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 $n_1 = m_2 n_2$, и так далее. Заметим, что $n > n_1 > n_2 > \ldots$, 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.

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

5. Пусть p_1 и p_2 не взаимно просты; тогда по пункту (1) имеем $p_1 \vdots p_2$ и $p_2 \vdots p_1$, то есть, они ассоциированы.

6. Индукция по n; база — пункт (2). $(a_1 a_2)a_3 \ldots a_n \vdots p$, поэтому, либо $a_1 a_2$, либо какое-то из a_i, $i > 2$, делится на p; если $a_1 a_2$ делится на p, то либо a_1, либо a_2 делится на p.

□

Теорема 2.5.3 (Основная теорема арифметики). *Каждое натуральное число, большее единицы, 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: теперь $n > 1$, и мы знаем, что $n = p_1 n_1$ для некоторого простого p. 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 6 простых чисел (2.5.2) один из сомножителей

$p_1, \ldots, p_k$ делится на q_1. Пусть это p_i: $p_i \vdots q_1$. Но по свойству 5 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.

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 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 делится на n, можно записать $m = nr$. Пусть $r = q_1 \cdots q_t$ — какое-то разложение r на простые сомножители. Тогда равенство $m = nr$ превращается в равенство

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

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 n встретилось $p_i^{l_i}$ для $l_i > 0$, то справа в равенстве 1 простой сомножитель p_i встретился как минимум l_i раз; значит, и слева он должен встретиться как минимум l_i раз. Однако слева показатель при l_i равен k_i. Значит, $k_i \geqslant l_i$. Если же $l_i = 0$ для какого-то i, то неравенство $k_i \geqslant l_i$ выполнено автоматически. Мы доказали следующее предложение:

Предложение 2.5.6. *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* m *делится на* n, *то* $k_i \geqslant l_i$ *для всех* $i = 1, \ldots, s$.

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

Предложение 2.5.7. *Пусть* $n = p_1^{k_1} \cdots p_s^{k_s}$ — *каноническое разложение числа* n. *Тогда количество всех натуральных делителей* n *равно* $(1 + k_1) \cdots (1 + k_s)$.

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

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 $m \vdots d$. Аналогично, $n \vdots d$. 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 $m \vdots d'$, 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 $d' \vdots d$, 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 $a - b$ делится на m. Обозначение: $a \equiv b \pmod{m}$, $a \equiv_m b$.

Следующие наблюдения верны для произвольного натурального $m > 0$.

Предложение 2.6.2 (Свойства сравнений). *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. $a - a \vdots 0$.

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

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

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

5. Пусть $n \in \mathbb{Z}$. Поделим n на m с остатком: $n = mq + r$, где $0 \leqslant r \leqslant m - 1$; тогда $n - r = mq \vdots m$, 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 $r_1 - r_2 \vdots m$. Но $|r_1 - r_2| \leqslant m - 1$, поэтому $r_1 = r_2$.

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

7. Если $a \perp m$, то $1 = au_0 + mv_0$ для некоторых целых u_0, v_0, откуда $au_0 - 1 = -mv_0 \vdots m$ и $au_0 \equiv 1 \pmod{m}$. Обратно, если $ax_0 \equiv 1 \pmod{m}$ для некоторого x_0, то $ax_0 - 1 \vdots m$, 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 Китайская теорема об остатках

Литература: [V], гл. IV, § 3.

Теорема 2.7.1 (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 $x - x' \vdots mn$, 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.

2.8 Классы вычетов, действия над ними

Литература: [F], гл. I, § 2, пп. 2, 3, § 3, п. 2; [K1], гл. 4, § 3, пп. 1, 2, 4; [vdW], гл. 3, § 11; [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 2.8.1. 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 2.8.2. 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 $x'y' \equiv xy \pmod m$.

Предложение 2.8.3 (Свойства операций на $\mathbb{Z}/m\mathbb{Z}$). *1.* $a + (b + c) = (a + b) + c$ *для любых* $a, b, c \in \mathbb{Z}/m\mathbb{Z}$ *(ассоциативность сложения);*

2. $\overline{0} + a = a + \overline{0} = a$ *для любого* $a \in \mathbb{Z}/m\mathbb{Z}$ *(то есть,* $\overline{0}$ *— нейтральный элемент по сложению);*

3. для любого $a \in \mathbb{Z}/m\mathbb{Z}$ *существует* $a' \in \mathbb{Z}/m\mathbb{Z}$ *такой, что* $a + a' = \overline{0}$ *(то есть,* a' *— обратный по сложению к элементу* a*; такой элемент обычно обозначается через* $-a$*);*

4. $a + b = b + a$ *для любых* $a, b \in \mathbb{Z}/m\mathbb{Z}$ *(коммутативность сложения);*

5. $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ *для любых* $a, b, c \in \mathbb{Z}/m\mathbb{Z}$ *(ассоциативность умножения);*

6. $\overline{1} \cdot a = a \cdot \overline{1} = a$ *для любого* $a \in \mathbb{Z}/m\mathbb{Z}$ *(то есть,* $\overline{1}$ *— нейтральный элемент по умножению);*

7. $a \cdot b = b \cdot a$ *для любых* $a, b \in \mathbb{Z}/m\mathbb{Z}$ *(коммутативность умножения);*

8. $a \cdot (b + c) = a \cdot b + a \cdot c$ *и* $(b + c) \cdot a = b \cdot a + c \cdot a$ *для любых* $a, b, c \in \mathbb{Z}/m\mathbb{Z}$ *(дистрибутивность умножения относительно сложения).*

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 лишь, что если $a = \overline{x}$, то в свойстве (3) нужно взять $a' = \overline{-x}$. □

Предложение 2.8.3 фактически означает, что операции, введенные нами на множестве $\mathbb{Z}/m\mathbb{Z}$, обладают хорошими свойствами, похожими на свойства обычных целых чисел. В этом случае алгебраисты говорят, что множество $\mathbb{Z}/m\mathbb{Z}$ является *кольцом* относительно этих операций.

Определение 2.8.4. Множество R с двумя бинарными операциями $+$ и $\cdot$ называется **коммутативным ассоциативным кольцом с единицей**, если выполняются следующие свойства (**аксиомы кольца**):

1. $a + (b + c) = (a + b) + c$ для любых $a, b, c \in \mathbb{R}$ (ассоциативность сложения);

2. существует элемент $0 \in R$ такой, что $0 + a = a + 0 = a$ для любого $a \in R$ (то есть, 0 — нейтральный элемент по сложению);

3. для любого $a \in R$ существует $a' \in R$ такой, что $a + a' = 0$ (то есть, a' — обратный по сложению к элементу a; такой элемент обычно обозначается через $-a$);

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

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

6. существует элемент $1 \in R$ такой, что $1 \cdot a = a \cdot 1 = a$ для любого $a \in R$ (то есть, 1 — нейтральный элемент по умножению);

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

8. $a \cdot (b + c) = a \cdot b + a \cdot c$ и $(b + c) \cdot a = b \cdot a + c \cdot a$ для любых $a, b, c \in \mathbb{R}$ (дистрибутивность умножения относительно сложения).

Замечание 2.8.5. На самом деле, R называется *кольцом*, если выполнены аксиомы 1–4 и 8; при выполнении аксиомы 5 говорят об *ассоциативном* кольце, при выполнении аксиомы 6 — о *кольце с единицей*, при выполнении аксиомы 7 — о *коммутативном кольце*. Но мы пока что интересуемся в основном структурами, в которых выполнены все аксиомы 1–8, поэтому будем сокращать термин «ассоциативное коммутативное кольцо с единицей» до одного слова «кольцо», и при необходимости произносить словосочетания «неассоциативное кольцо», «кольцо без единицы» и «некоммутативное кольцо» для структур, в которых не обязательно выполняются соответствующие аксиомы 5, 6, 7.

Примеры 2.8.6. Множества целых чисел $\mathbb{Z}$, рациональных чисел $\mathbb{Q}$, вещественных чисел $\mathbb{R}$ (со стандартными операциями сложения и умножения) являются кольцами. По предложению 2.8.3 множество $\mathbb{Z}/m\mathbb{Z}$ с операциями, введенными выше, является кольцом.

Обратите внимание, что аксиомы 1–4 говорят для сложения почти то же, что аксиомы 5–7 говорят для умножения. Отличие состоит в том, что для умножения мы не требуем наличия обратного элемента по умножению к каждому. Интуиция из приведенных выше примеров показывает, что вряд ли в кольце есть обратный элемент по умножению к элементу 0: и действительно, нетрудно показать, что в произвольном кольце R выполнено $a \cdot 0 = 0$ для любого $a \in R$, поэтому, если $0^\bullet$ таков, что $0 \cdot 0^\bullet = 1$, то $a \cdot 0 \cdot 0^\bullet = 0 \cdot 0^\bullet$, откуда $a = 1$, и R состоит из одного элемента. Тем не менее, интересно рассматривать структуры, в которых обратный элемент по умножению есть у всех *ненулевых* элементов.

Определение 2.8.7. [Ассоциативное коммутативное] кольцо [с единицей] R называется **полем**, если оно удовлетворяет следующему условию (**аксиоме поля**):

9. в R больше одного элемента, и для любого $a \in R$, $a \neq 0$, существует элемент $a^\bullet \in R$ такой, что $a \cdot a^\bullet = a^\bullet \cdot a = 1$ (то есть, $a^\bullet$ — обратный по умножению к элементу a; такой элемент обычно обозначается через a^{-1}).

Условие «в R больше одного элемента» в аксиоме поля избавляет нас от вырожденного случая *нулевого кольца*, в котором $0 = 1$.

Определение 2.8.8. Пусть R — кольцо. Элемент $a \in R$ называется **обратимым**, если у него есть обратный по умножению, то есть, если найдется $a^\bullet \in R$ такой, что $a \cdot a^\bullet = a^\bullet a = 1$. Множество всех обратимых элементов кольца R обозначается через R^*.

То есть, поле — это кольцо, в котором все ненулевые элементы обратимы.

Примеры 2.8.9. Множество $\mathbb{Z}$ не является полем: в нем только два обратимых элемента, 1 и -1. Множества $\mathbb{Q}$ и $\mathbb{R}$ являются полями.

Возникает вопрос: при каких m кольцо $\mathbb{Z}/m\mathbb{Z}$ является полем? Опишем все обратимые элементы в $\mathbb{Z}/m\mathbb{Z}$.

Предложение 2.8.10. *Пусть $m > 0$ — натуральное число, $x \in \mathbb{Z}$. Класс $\overline{x}$ обратим в $\mathbb{Z}/m\mathbb{Z}$ тогда и только тогда, когда $x \perp m$.*

Доказательство. Заметим, что $\overline{y}$ является обратным к $\overline{x} \Leftrightarrow \overline{x} \cdot \overline{y} = \overline{1}$ Leftrightarrow $\overline{xy} = \overline{1}$ $\Leftrightarrow xy \equiv 1 \pmod{m} \Leftrightarrow xy - 1 = mk$ для некоторого k. Поэтому olx обратим в $\mathbb{Z}/m\mathbb{Z}$ тогда и только тогда, когда уравнение $xy - mk = 1$ разрешимо в целых числах, то есть, по свойству 2 взаимной простоты (2.3.3), когда $x \perp m$. □

Предложение 2.8.11. *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 2.8.10 класс $\overline{x}$ обратим. Обратно, если m не простое, можно записать $m = kl$ для некоторых натуральных k, l, причем k отлично от m и 1. Заметим, что $\gcd(k, m) = k$ по свойству 1 НОД (2.3.1), поэтому k не взаимно просто с m. По предложению 2.8.10 из этого следует, что $\overline{k}$ необратим в $\mathbb{Z}/m\mathbb{Z}$, и $\overline{k} \neq \overline{0}$, поскольку k не делится на m. Значит, $\mathbb{Z}/m\mathbb{Z}$ не является полем. □

2.9 Теорема Вильсона

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

Теорема 2.9.1 (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 $\mathbb{Z}/p\mathbb{Z}$:

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

В произведении справа выписаны все ненулевые элементы $\mathbb{Z}/p\mathbb{Z}$. По предложению 2.8.11 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 $x^2 - 1 \vdots p$, то есть, $(x-1)(x+1) \vdots p$, и по свойству 2 простых чисел (2.5.2) из этого следует, что $x \pm 1 \vdots p$, 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 $1, \ldots, p-1$ встретится k. Если все-таки $(p-1)! \equiv -1 \pmod{p}$, то $(p-1)! + 1 \vdots p$, 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. □

2.10 Функция Эйлера

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

Определение 2.10.1. 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 $\mathbb{N} \setminus \{0\} \to \mathbb{N}$, которая называется **функцией Эйлера**.

Пример 2.10.2. Прямое вычисление показывает, что $\varphi(1) = 1$, $\varphi(2) = 1$, $\varphi(3) = 2$.

Предложение 2.10.3. *Пусть $n \in \mathbb{N}$, $n > 0$. Тогда $\varphi(n)$ равно количеству обратимых элементов кольца $\mathbb{Z}/n\mathbb{Z}$. В обозначениях определения 2.8.8, $\varphi(n) = |(\mathbb{Z}/n\mathbb{Z})^*|$.*

Доказательство. Пусть $0 \leqslant x < n$; по предложению 2.8.10 $x \perp n$ тогда и только тогда, когда $\overline{x}$ обратим. □

Замечание 2.10.4. Теперь можно посчитать $\varphi(p)$ для простого p: по предложению 2.8.11 кольцо $\mathbb{Z}/p\mathbb{Z}$ является полем, то есть, $(\mathbb{Z}/p\mathbb{Z})^* = (\mathbb{Z}/p\mathbb{Z}) \setminus \{\overline{0}\}$, 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 2.10.5. *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 $x - x' \vdots mn$, поэтому $x - x' \vdots m$ и $x - x' \vdots n$. Значит, $[x]_m = [x']_m$ и $[x]_n = [x']_n$. По китайской теореме об остатках (2.7.1) 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 2.10.5 отображении обратимые классы переходят в пары обратимых классов.

Предложение 2.10.6. *Пусть m, n, f таковы, как в формулировке теоремы 2.10.5, $\overline{x} \in \mathbb{Z}/mn\mathbb{Z}$, $f(\overline{x}) = (a, b)$. 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 2.10.7 (Мультипликативность функции Эйлера). *Если $m, n \geqslant 1$ — натуральные числа и $m \perp n$, то $\varphi(mn) = \varphi(m)\varphi(n)$.*

Доказательство. По предложению 2.10.3, $\varphi(mn) = |(\mathbb{Z}/mn\mathbb{Z})^*|$ 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 2.10.6 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 2.10.8. *Если $n = p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_s^{k_s}$ — каноническое разложение натурального числа n, то $\varphi(n) = p_1^{k_1-1}(p_1 - 1) \cdot p_2^{k_2-1}(p_2 - 1) \cdot \ldots \cdot p_s^{k_s-1}(p_s - 1)$.*

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 2.10.7 и замечание 2.10.4, получаем $\varphi(n) = \varphi(p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_s^{k_s}) = \varphi(p_1^{k_1}) \cdot \varphi(p_2^{k_2}) \cdot \ldots \cdot \varphi(p_s^{k_s}) = p_1^{k_1-1}(p_1 - 1) \cdot p_2^{k_2-1}(p_2 - 1) \cdot \ldots \cdot p_s^{k_s-1}(p_s - 1)$, что и требовалось. □

2.11 Малая теорема Ферма и теорема Эйлера

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

Лемма 2.11.1. *Пусть R — кольцо (не обязательно коммутативное). Произведение обратимых элементов R обратимо. В частности, произведение обратимых классов вычетов в $\mathbb{Z}/m\mathbb{Z}$ обратимо.*

Доказательство. Предположим, что $x_1, x_2, \ldots, x_n \in R$ — обратимые элементы R. Тогда существуют $x_1^{-1}, x_2^{-1}, \ldots, x_n^{-1} \in R$ такие, что $x_i x_i^{-1} = x_i^{-1} x_i = 1$ в R. Покажем, что элемент $x_n^{-1} \ldots x_2^{-1} x_1^{-1}$ является обратным к $x_1 x_2 \ldots x_n$. Действительно, $(x_n^{-1} \ldots x_2^{-1} x_1^{-1})(x_1 x_2 \ldots x_n) = x_n^{-1} \ldots x_2^{-1}(x_1^{-1} x_1) x_2 \ldots x_n = x_n^{-1} \ldots (x_2^{-1} x_2) \ldots x_n = \cdots = x_n^{-1} x_n = 1$ и, совершенно аналогично, $(x_1 x_2 \ldots x_n)(x_n^{-1} \ldots x_2^{-1} x_1^{-1}) = 1$. □

Теорема 2.11.2 (Теорема Эйлера). *Пусть n — натуральное число, $a \in \mathbb{Z}$ и $a \perp n$. Тогда $a^{\varphi(n)} \equiv 1 \pmod{n}$.*

Доказательство. Пусть $x_1, x_2, \ldots, x_k$ — все обратимые элементы кольца $\mathbb{Z}/n\mathbb{Z}$. По предложению 2.10.3 их ровно $\varphi(n)$, то есть, $k = \varphi(n)$. Пусть $\overline{a}$ — класс числа a в кольце $\mathbb{Z}/n\mathbb{Z}$. По предложению 2.8.10 элемент $\overline{a}$ обратим. Рассмотрим элементы $\overline{a}x_1, \overline{a}x_2, \ldots, \overline{a}x_k$. По лемме 2.11.1 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

$\overline{a}x_1, \overline{a}x_2, \ldots, \overline{a}x_k$ различны; иными словами, это элементы $x_1, x_2, \ldots, x_k$, только, возможно, в другом порядке. Но тогда произведения этих двух наборов элементов совпадают. Значит,

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

По лемме 2.11.1 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 2.11.3 (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 2.10.4 $\varphi(p) = p - 1$. Осталось применить теорему Эйлера для $n = p$. □

2.12 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 $ed = 1 + k\varphi(n)$.

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 $c^d \equiv (m^e)^d \equiv m^{ed} \equiv m^{1+k\varphi(n)} \equiv m \cdot (m^{\varphi(n)})^k \equiv m \pmod{n}$. Последнее сравнение выполнено по теореме Эйлера; для ее применения необходимо, чтобы m было взаимно простым с n. Этого можно добиться, поскольку n является произведением двух простых и вероятность того, что m имеет общий делитель с n, чрезвычайно мала. 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 2.8.4.

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, $\mathrm{Re}\, 1 = \mathrm{Re}(\frac{z}{z+w}) + \mathrm{Re}(\frac{w}{z+w})$. Заметим, что $\mathrm{Re}(z) \leqslant |z|$ для любого комплексного числа z, поэтому $\mathrm{Re}\, 1 \leqslant |\frac{z}{z+w}| + |\frac{w}{z+w}|$. 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 $\mathrm{Arg}(z)$ Обратно, по модулю r и аргументу φ комплексное число z однозначно восстанавливается: $z = a + bi$, $a = r\cos(\varphi)$, $b = r\sin(\varphi)$.

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 $\varphi + 2\pi k = \varphi + 2\pi l + 2\pi tn$ 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. 2.8.2), 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, $km \vdots n$. Но k и n взаимно просты; по свойству 3 взаимной простоты (2.3.3) теперь $m \vdots n$ — 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 (2.10.1). □

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; то есть, для комплексного числа $z = r(\cos(\varphi)+i\sin(\varphi))$ получаем $z = r \cdot e^{i\varphi}$, и 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: $\ln(z_1z_2) = \ln(z_1) + \ln(z_2)$. 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

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: $(a+b)_k = a_k + b_k$, $(ab)_k = \sum_{i=0}^{k} a_i b_{k-i}$. 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 (2.8.4). Сложение в $R[x]$ происходит покомпонентно, поэтому первые четыре аксиомы, отражающие свой-

ства 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 $a \in R[x]$ мы обозначаем через a^n многочлен $\underbrace{a \cdot \dots \cdot a}_{n}$; если $n = 0$, положим $a^0 = 1 \in R[x]$.

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**.

4.2 Области целостности

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

Определение 4.2.1. Пусть R — кольцо. Элемент $x \in R$ называется **делителем нуля** в R, если существует элемент $b \in R$, $b \neq 0$ такой, что $ab = 0$.

Замечание 4.2.2. Кольцо, состоящее из одного элемента 0 (с тривиальными операциями сложения и умножения $0 + 0 = 0 \cdot 0 = 0$) называется **нулевым** и часто обозначается через 0. В ненулевом кольце R всегда есть по крайней мере один делитель нуля: это элемент 0.

Определение 4.2.3. Ненулевое кольцо R называется **областью целостности**, если из того, что $ab = 0$ для некоторых $a, b \in R$, следует, что $a = 0$ или $b = 0$.

Замечание 4.2.4. Таким образом, кольцо R является областью целостности тогда и только тогда, когда оно ненулевое в нем нет ненулевых делителей нуля. Очевидно, что кольцо $\mathbb{Z}$ является областью целостностью. Кроме того, любое поле является областью целостности

(если $xy = 0$ в поле R и $x \neq 0$, то после домножения на x^{-1} получаем $y = 0$). Кольцо классов вычетов $\mathbb{Z}/m\mathbb{Z}$ при простом m является полем и, следовательно, областью целостности. С другой стороны, если m не простое, $m = kl$, то $\overline{k} \cdot \overline{l} = \overline{m} = \overline{0}$ — делители нуля в $\mathbb{Z}/m\mathbb{Z}$.

Теорема 4.2.5. *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 $f = a_0 + a_1x + \cdots + a_mx^m$, $g = b_0 + b_1x + \cdots + b_nx^n$. По определению степени имеем $a_m \neq 0$ и $b_n \neq 0$. Нетрудно видеть, что $fg = a_0b_0 + \cdots + a_mb_nx^{m+n}$ и $a_mb_n \neq 0$, поскольку R — область целостности. □

Замечание 4.2.6. Заметим, что теорема верна и для случая $f = 0$ или $g = 0$ за счет нашего соглашения $\deg(0) = -\infty$.

Следствие 4.2.7. *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 4.2.8. *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 $f \in R[x]^*$. □

4.3 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 будем рассматривать лишь многочлены над полем, и обозначать это поле через k.

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

Определение 4.3.1. Пусть $f, g \in k[x]$. Говорят, что многочлен f **делится на** многочлен g, если $f = gh$ для некоторого $h \in k[x]$. Обозначение: $f \vdots g$.

Предложение 4.3.2 (Свойства делимости в кольце многочленов). *Пусть $f, g, h \in k[x]$. Тогда*

1. *$f \vdots f$ и $f \vdots 1$;*
2. *если $f \vdots h$, $g \vdots h$, то $f + g \vdots h$;*

3. если $f \vdots h$, *то* $fg \vdots h$;

4. если $f \vdots g$, $g \vdots h$, *то* $f \vdots h$.

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 $f = gp$, $g = hq$, то $f = hpq$.

□

Определение 4.3.3. Два элемента $f, g \in k[x]$ называются **ассоциированными**, если $f \vdots g$ и $g \vdots f$.

Предложение 4.3.4. *Ассоциированность является отношением эквивалентности.*

Доказательство. Очевидно. □

Предложение 4.3.5. $f, g \in k[x]$ *ассоциированы тогда и только тогда, когда* $f = cg$ *для некоторой обратимой (то есть, ненулевой) константы* $c \in k^*$.

Доказательство. Если $f = cg$ для $c \in k^*$, то $f \vdots g$ и $g = c^{-1}f$, поэтому $g \vdots f$. Обратно, из $f \vdots g$ следует, что $f = gh$, а из $g \vdots f$ следует, что $g = fp$. Поэтому $f = gh = fph$, откуда $f(1 - ph) = 0$. Если $f = 0$, то и $g = 0$, и доказывать нечего. Иначе (поскольку $k[x]$ — бласть целостности) получаем $1 = ph$, откуда $p \in k[x]^* = k^*$. Значит, p — ненулевая константа, что и требовалось доказать. □

Теорема 4.3.6. *О делении с остатком в кольце многочленов Пусть* k *— поле,* $f, g \in k[x]$, $g \neq 0$. *Существуют единственные многочлены* $h, r \in k[x]$ *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 $m \geqslant n$. Рассмотрим многочлен $f_0 = f - g \cdot \frac{a_m}{b_n} x^{m-n}$. Степень g равна n, степень $\frac{a_m}{b_n} x^{m-n}$ равна $m - n$, поэтому степень $g \cdot \frac{a_m}{b_n} x^{m-n}$ равна m, как и степень f. Значит, степень f_0 не превосходит m. Посмотрим на коэффициент f_0 при x^m. Он равен разности коэффициентов f и $g \cdot \frac{a_m}{b_n} x^{m-n}$ при x^m, то есть, $a_m - b_n \cdot \frac{a_m}{b_n} = 0$. 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 $f = f_0 + g \cdot \frac{a_m}{b_n} x^{m-n} = gh_0 + r_0 + g \cdot \frac{a_m}{b_n} x^{m-n} = g(h_0 + \frac{a_m}{b_n} x^{m-n}) + r_0$. Возьмем $h = h_0 + \frac{a_m}{b_n} x^{m-n}$ 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 $r = \widetilde{r}$. □

4.4 Многочлен как функция

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

Определение 4.4.1. Пусть $f = a_0 + a_1x + \cdots + a_nx^n \in k[x]$, $c \in k$. **Значением** многочлена f в точке c называется $f(c) = a_0 + a_1c + \cdots + a_nc^n = \sum_{i=0}^{\infty} a_ic^i \in k$.

Замечание 4.4.2. Таким образом, с каждым многочленом $f \in k[x]$ связано отображение $\widetilde{f}\colon k \to k$, $c \mapsto f(c)$.

Предложение 4.4.3. *Для любых* $f, g \in k[x]$, $c \in k$, *выполнено*

1. $(f + g)(c) = f(c) + g(c)$;
2. $(fg)(c) = f(c) \cdot g(c)$;
3. *если* $f = r \in k$, *то* $f(c) = r$

Доказательство. Пусть $f = \sum_{i=0}^{\infty} a_ix^i$, $g = \sum_{i=0}^{\infty} b_ix^i$.

1. $f + g = \sum_{i=0}^{\infty}(a_i + b_i)x^i$, поэтому $(f + g)(c) = \sum_{i=0}^{\infty}(a_i + b_i)c^i = \sum_{i=0}^{\infty}(a_ic^i) + \sum_{i=0}^{\infty}(b_ic^i) = f(c) + g(c)$.
2. $fg = \sum_{m=0}^{\infty}\sum_{i+j=m}^{\infty}(a_ib_jx^m)$, поэтому $f(c)g(c) = (\sum_{i=0}^{\infty} a_ic^i)(\sum_{j=0}^{\infty} b_jc^j) = \sum_{i,j=0}^{\infty}(a_ib_jc^{i+j}) = \sum_{m=0}^{\infty}\sum_{i+j=m}(a_ib_jc^m) = (fg)(c)$.
3. $f(c) = r + 0 \cdot c + \cdots = r$.

□

Определение 4.4.4. Пусть $f \in k[x]$, $c \in k$. Говорят, что c является **корнем** многочлена f, если $f(c) = 0$.

Теорема 4.4.5 (Теорема Безу). *Пусть* $f \in k[x]$, $c \in k$. *Многочлен* f *делится на* $(x - c)$ *тогда и только тогда, когда* c *является корнем* f.

Доказательство. Пусть $f = (x - c)g$, тогда $f(c) = ((x - c)g)(c) = 0 \cdot g(c) = 0$. Обратно, пусть c — корень f. Поделим f на $(x - c)$ с остатком: $f = (x - c)h + r$. Заметим, что $\deg(r) < \deg(x - c) = 1$, поэтому $r \in k$ — константа. Подставим c в обе части этого равенства: $f(c) = ((x - c)h + r)(c) = ((x - c)h)(c) + r(c) = 0 \cdot h(c) + r = r$. По предположению $f(c) = 0$, поэтому $r = 0$ и $f = (x - c)h$. □

Предложение 4.4.6. *Пусть* $f \in k[x]$, $f \neq 0$. *Тогда* f *можно записать в виде* $f = (x - c_1)\dots(x - c_m)h$, *где* $c_1, \dots, c_m \in k$ — *все корни* f *(возможно, с повторениями), а* $h \in k[x]$ — *многочлен, у которого нет корней в поле* k.

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 4.4.5) $f = (x - c)f_1$, $\deg(f_1) < \deg(f)$, 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.

Очевидно, что каждый c_i, $i = 1, \ldots, m$, является корнем f. Обратно, если c — некоторый корень f, то $0 = f(c) = (c - c_1) \ldots (c - c_m)h(c)$. При этом $h(c) \neq 0$, поскольку у h нет корней, значит (поскольку $k[x]$ — область целостности), одна из скобок вида $(c - c_i)$ равна 0, поэтому c содержится среди $c_1, \ldots, c_m$. □

Следствие 4.4.7. *Число различных корней ненулевого многочлена над полем не превосходит его степени.*

Доказательство. Посмотрим на разложение из предложения 4.4.6. Все корни c многочлена $f \in k[x]$ 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 4.6.3) мы уточним это следствие с помощью понятия *кратности* корня.

Определение 4.4.8. Пусть $f, g \in k[x]$ — многочлены над полем. Говорят, что многочлен f **функционально равен** многочлену g, если $f(c) = g(c)$ для любого $c \in k$. Иными словами, многочлены функционально равны, если задаваемые ими функции равны: $\widetilde{f} = \widetilde{g}$ (см. замечание 4.4.2). Обычное равенство многочленов при этом иногда называют **формальным равенством**: многочлены f и g формально равны, если $f = g$.

Пример 4.4.9. Пусть $k = \mathbb{Z}/2\mathbb{Z} = \{\overline{0}, \overline{1}\}$. 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 $k = \mathbb{Z}/p\mathbb{Z}$: достаточно взять $f = x^p - x$ и вспомнить малую теорему Ферма (следствие 2.11.3).

Замечание 4.4.10. Очевидно, что из формального равенства многочленов следует функциональное: если $f = g$, то $f(c) = g(c)$ для любого $c \in k$.

Теорема 4.4.11. *Если поле k бесконечно, то из функционального равенства многочленов над k следует их формальное равенство.*

Доказательство. Пусть $f, g \in k[x]$ и $f(c) = g(c)$ для всех $c \in k$. Посмотрим на разность $h = f - g \in k[x]$. Для любого $c \in k$ выполнено $h(c) = f(c) - g(c) = 0$, поэтому c — корень h. Если h ненулевой, то по следствию 4.4.7 число корней h не превосходит его степени; с другой стороны, как мы только что видели, любой элемент бесконечного поля является корнем h — противоречие. Значит, $h = 0$, поэтому и $f = g$. □

4.5 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.

Определение 4.5.1. Поле k называется **алгебраически замкнутым**, если у любого многочлена $f \in k[x]$ степени выше нулевой имеется корень в k.

Пример 4.5.2. 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 4.5.3 (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 4.4.6 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 4.5.4 (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 $\lambda \in \mathbb{C} \setminus \mathbb{R}$, 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. □

4.6 Кратные корни и производная

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

Определение 4.6.1. Пусть $f \in k[x]$, $c \in k$. 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 4.6.2. *Пусть* $f \in k[x]$, $c \in k$, $m \geqslant 1$. *Элемент* c *является корнем* f *кратности* m *тогда и только тогда, когда* f *можно представить в виде* $f = (x-c)^m \cdot g$ *так, что* $g(c) \neq 0$.

Доказательство. Если c — корень f кратности m, то $f = (x-c)^m \cdot g$ для некоторого $g \in k[x]$. 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 $k[x]$ — область целостности, получаем $g-(x-c)h = 0$, откуда $g = (x-c)h$ и $g(c) = 0$ — противоречие. □

Замечание 4.6.3. Таким образом, если в выражении для многочлена f из следствия 4.4.6 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 4.6.4. Пусть $f \in k[x]$, $f = \sum_{s=0}^{\infty} a_s x^s$. **Производным многочленом** от многочлена f (или его **производной**) называется многочлен $f' = \sum_{s=1}^{\infty} s a_s x^{s-1}$.

Замечание 4.6.5. Напомним, что для элемента $c \in k$ и натурального числа s можно положить $sc = \underbrace{c + \cdots + c}_{s} = \underbrace{(1 + \cdots + 1)}_{s} \cdot c \in k$.

Предложение 4.6.6 (Свойства производной). *Пусть* $f, g \in k[x]$, $c \in k$, $m \geqslant 1$.

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 $(g^m)' = (g \cdot g^{m-1})' = g' \cdot g^{m-1} + g \cdot (g^{m-1})' = g^{m-1}g' + g \cdot (m-1)g^{m-2}g' = mg^{m-1}g'$, что и требовалось.

□

Предложение 4.6.7 (Связь между корнями многочлена и его производной). *Пусть* $f \in k[x]$, $c \in k$. *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. □

Определение 4.6.8. 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 4.6.9. 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 4.6.10. *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 4.6.11. *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 4.6.2 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 4.6.2, 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. □

Теорема 4.6.12. *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 4.6.7 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 4.6.11 тогда c — корень f кратности m, что и требовалось доказать. □

4.7 Интерполяция

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

Определение 4.7.1. Пусть k — поле, $x_1, \dots, x_n \in k$ — некоторые попарно различные элементы k, и $y_1, \dots, y_n \in k$. **Интерполяционной задачей** (или **задачей интерполяции в n точках**) с данными $(x_1, \dots, x_n; y_1, \dots, y_n)$ мы будем называть задачу нахождения многочлена $f \in k[x]$ такого, что $f(x_i) = y_i$ для всех $i = 1, \dots, m$.

Теорема 4.7.2. *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 $(x - x_1) \dots (x - x_n)$.*

Доказательство. Пусть $f, g \in k[x]$ — два многочлена, являющихся решениями одной интерполяционной задачи с данными $(x_1, \dots, x_n; y_1, \dots, y_n)$. Это означает, что $f(x_i) = y_i = g(x_i)$ для всех $i = 1, \dots, n$. 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 $x_1, \dots, x_n$. По предложению 4.4.6 из этого следует, что h делится на $(x - x_1) \dots (x - x_n)$. 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 4.7.3. 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 $(x_1, \dots, x_n; y_1, \dots, y_n)$ — фиксированная интерполяционная задача. Обозначим

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

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 $x_1, \dots, \widehat{x_i}, \dots, x_n$.

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 $x_1, \dots, x_n$.

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 последовательности

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

4.8 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.

Определение 4.8.1. Пусть $f, g \in k[x]$. Многочлен d называется **общим делителем** многочленов f и g, если $f \vdots d$ и $g \vdots d$.

Определение 4.8.2. 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 h — общий делитель f и g, то $d \vdots h$.

Замечание 4.8.3. Сразу же заметим, что если d и d' — два наибольших общих делителя многочленов f и g, то по определению имеем $d \vdots d'$ и $d' \vdots d$; 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 там мы выбирали из них натуральный и именно его называли 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 4.8.4. *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 4.8.5. Представление из теоремы 4.8.4 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 4.8.6. *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 $d \vdots h$. 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 4.8.4:

Предложение 4.8.7. *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 4.8.4 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 4.8.8. 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 4.8.9. *Пусть* $f, g, p \in k[x]$ *и* p *неприводим. Если* $fg \vdots p$, *то* $f \vdots p$ *или* $g \vdots p$.

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 4.8.10. *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; единственное изменение состоит в том, что если $p, q \in k[x]$ неприводимы и p делится на q, то p *ассоциировано* с q (a не равно q). □

4.9 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 4.2.3). 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 $R \times (R \setminus \{0\})$ 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 4.9.1. *Это отношение эквивалентности на* $R \times (R \setminus \{0\})$.

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 $R \times (R \setminus \{0\})$ 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 $R \times (R \setminus \{0\})$. 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 4.9.2. *Пусть R — область целостности. Множество $\mathrm{Frac}(R)$ с введеннъыми выше операциями является полем.*

Определение 4.9.3. $\mathrm{Frac}(R)$ 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 $s \in R \setminus \{0\}$.

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 $s \in R \setminus \{0\}$.

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

4.10 Поле рациональных функций

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

Определение 4.10.1. 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 4.10.2. 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 4.10.3. Рациональная функция $\frac{f}{g} \in k(x)$ называется **правильной**, если $\deg(f) < \deg(g)$

Лемма 4.10.4. *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 4.10.5. *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. □

Предложение 4.10.6. *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. Если же $f + \psi = \widetilde{f} + \widetilde{\psi}$, то $f - \widetilde{f} = psi - \psi$. В левой части этого равенства стоит многочлен, в правой — правильная дробь; достаточно доказать, что если многочлен равен правильной дроби, то он нулевой. Пусть теперь $f = \psi$, где $\psi = \frac{g}{h}$ — правильная дробь ($f, g, h \in k(x)$). Тогда $fh = g$ и из неравенства $\deg(g) < \deg(h)$ следует, что $\deg(f) = \deg(g) - \deg(h) < 0$, поэтому $f = 0$. 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 4.10.7. 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 4.10.8. *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 4.10.6 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 4.10.5, правильной дробью, и из единственности в предложении 4.10.6, $c = 0$. □

Лемма 4.10.9. *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 4.10.10. *Пусть* $\frac{f}{g} \in k(x)$ *— правильная дробь,* $g = p_1^{m_1} \dots p_s^{m_s}$ *— каноническое разложение* g *на неприводимые множители. Тогда* $\frac{f}{g}$ *можно представить в виде суммы простейших дробей, в знаменателях которых стоят* $p_1, p_1^2, \dots, p_1^{m_1}$, $p_2, p_2^2, \dots, p_2^{m_2}, \dots$, $p_s, p_s^2, \dots, p_s^{m_s}$. *Кроме того, такое представление единственно с точностью до порядка, в котором записаны слагаемые.*

Доказательство. По предложению 4.10.8 можно расщепить знаменатель правильной дроби на два взаимно простых сомножителя; применяя ее несколько раз, получаем, что $\frac{f}{g} = \frac{f_1}{p_1^{m_1}} + \dots + \frac{f_s}{p_s^{m_s}}$. Далее, по лемме 4.10.9, 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 $p_1, \dots, p_s$) 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 $\frac{a_1}{p_1} + \frac{a_2}{p_1^2} + \dots + \frac{a_{m_1}}{p_1^{m_1}}$, а в правой части — $\frac{b_1}{p_1} + \frac{b_2}{p_1^2} + \dots + \frac{b_{m_1}}{p_1^{m_1}}$. По нашему предположению, $a_n \neq b_n$ для некоторого n. Рассмотрим максимальное такое n. Тогда $a_{n+1} = b_{n+1}, \dots, a_{m_1} = b_{m_1}$, поэтому дроби $\frac{a_{n+1}}{p_1^{n+1}}, \dots, \frac{a_{n+1}}{p_1^{n+1}}$ в левой части равны соответственно дробям $\frac{b_{n+1}}{p_1^{n+1}}, \dots, \frac{b_{n+1}}{p_1^{n+1}}$ в правой части. Вычеркивая эти дроби, получаем равенство вида

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

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 4.10.11. *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 4.10.10 и теорем 4.5.3, 4.5.4. □

Теорема 4.10.10 не указывает явного алгоритма нахождения разложения правильной дроби в сумму простейших. Этот алгоритм можно извлечь из доказательства предложения 4.10.8 и леммы 4.10.9, но он несколько замысловат: например, в доказательстве 4.10.8 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 4.10.12. *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 4.10.10 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

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

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

Ненулевые элементы $a_{1,j_1}, a_{2,j_2}, \dots, a_{s,j_s}$ 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: $0x_1 + \dots + 0x_n = a_{s,j_s}$, 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 $x_{j_1}, \dots, x_{j_s}$. *Все остальные* переменные мы будем называть **свободными**, а переменные $x_{j_1}, \dots, x_{j_s}$ — **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 $s+1, \dots, m$ фактически имеют вид $0 = 0$, поэтому не влияют на множество решений системы, и их можно выбросить. Последнее уравнение имеет вид $a_{s,j_s}x_{j_s} + \dots = b_s$, 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: $a_{s-1,j_{s-1}}x_{j_{s-1}} + \dots = b_{s-1}$; 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. $(AB)^T=B^TA^T$.

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

□

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

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

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 $(0 \ \cdots \ 0 \ 1 \ 0 \ \cdots \ 0)$ (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 $(0 \ \cdot \ 0 \ 1 \ 0 \ \cdots \ 0)$ (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 $A = (a_{ij})$.

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

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 $(1, j_1), (2, j_2), \dots (r, j_r)$, где стоят единицы. После этого перестановкой столбцов можно добиться того, что эти единицы будут стоять в позициях $(1, 1), (2, 2), \dots (r, r)$. 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 $P_1, \dots, P_t, Q_1, \dots, Q_s$ такие, что*

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

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 $P_t P_{t-1} \dots P_1 A Q_1 \dots Q_{s-1} Q_s = \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix}$. 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 $A = P_1^{-1} \dots P_t^{-1} D Q_s^{-1} \dots Q_1^{-1}$. Положим теперь $P = P_1^{-1} \dots P_t^{-1}$, $Q = Q_s^{-1} \dots Q_1^{-1}$; матрицы P и Q обратимы, поскольку они являются произведениями обратимых матриц. Получим $A = PDQ$, что и требовалось. □

Заметим, что набор матриц $P_1, \dots, P_s, Q_1, \dots, Q_t$ 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 $A = P_s P_{s-1} \dots P_1 \begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} Q_1 \dots Q_{t-1} Q_t$, 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 2.11.1).

Обратно, предположим, что A обратима. Из равенства

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

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 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 5.5.1. 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 5.5.2. 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 5.5.1 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 $\{1,\dots,n\}$, образом элемента 2 — любой из оставшихся $n-1$, и так далее; всего получаем $n\cdot(n-1)\cdot\dots\cdot 1 = n!$ 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:

$$\rho\circ\pi\colon \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}\begin{pmatrix}1&2&3\\2&3&1\end{pmatrix} \to \begin{pmatrix}1&2&3\\1&3&2\end{pmatrix}\begin{pmatrix}1&2&3\\2&1&3\end{pmatrix} \to \begin{pmatrix}1&2&3\\2&1&3\end{pmatrix}$$

$$\pi\circ\rho\colon \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}\begin{pmatrix}1&2&3\\1&3&2\end{pmatrix} \to \begin{pmatrix}1&2&3\\2&3&1\end{pmatrix}\begin{pmatrix}1&2&3\\3&2&1\end{pmatrix} \to \begin{pmatrix}1&2&3\\3&2&1\end{pmatrix}$$

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 5.5.3. 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 5.5.4. *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: $\pi(1), \pi(2), \dots, \pi(k)$. В конце концов (дойдя до $k = n$) мы получим перестановку π путем домножения id на транспозиции, что и требовалось. □

Предложение 5.5.5. *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 5.5.6. 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 5.5.7. 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 5.5.8. 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 5.5.9. *Пусть $\pi \in S_n$, $\tau_{ij} \in S_n$ — транспозиция. Тогда* $\mathrm{sgn}(\pi) = -\mathrm{sgn}(\pi \circ \tau_{ij})$.

Доказательство. Посмотрим, как меняется число инверсий перестановки при домножении на *элементарную транспозицию*. Сравним перестановки

$$\pi = \begin{pmatrix} \ldots & i & i+1 & \ldots \\ \ldots & \pi(i) & \pi(i+1) & \ldots \end{pmatrix} \text{ и } \pi \circ \tau_{i,i+1} = \begin{pmatrix} \ldots & i & i+1 & \ldots \\ \ldots & \pi(i+1) & \pi(i) & \ldots \end{pmatrix}.$$

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: $\mathrm{inv}(\pi \circ \tau_{i,i+1}) = \mathrm{inv}(\pi) \pm 1$. Поэтому эти числа имеют разную четность.

Это означает, что при домножении на элементарную транспозицию перестановка меняет знак. По предложению 5.5.5 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 5.5.10. *Пусть* $\pi = \tau_1 \circ \cdots \circ \tau_s$*, где* $\tau_1, \ldots, \tau_s$ *— транспозиции. Тогда* $\mathrm{sgn}(\pi) = (-1)^s$.

Доказательство. Запишем $\pi = \mathrm{id} \circ \tau_1 \circ \cdots \circ \tau_s$ 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 5.5.11. *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 5.5.9 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 5.5.12. *Пусть* $\pi, \rho \in S_n$*; тогда* $\mathrm{sgn}(\pi \circ \rho) = \mathrm{sgn}(\pi) \cdot \mathrm{sgn}(\rho)$.

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 5.5.10 имеем $\mathrm{sgn}(\pi) = (-1)^s$ и $\mathrm{sgn}(\rho) = (-1)^t$. 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 $\mathrm{sgn}(\pi \circ \rho) = (-1)^{s+t} = (-1)^s \cdot (-1)^t = \mathrm{sgn}(\pi) \cdot \mathrm{sgn}(\rho)$. □

Следствие 5.5.13. *Пусть* $\pi \in S_n$*; тогда* $\mathrm{sgn}(\pi^{-1}) = \mathrm{sgn}(\pi)$.

Доказательство. Заметим, что $\pi \circ \pi^{-1} = \mathrm{id}$, поэтому $\mathrm{sgn}(\pi) \cdot \mathrm{sgn}(\pi^{-1}) = \mathrm{sgn}(\mathrm{id}) = 1$. □

5.6 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 5.6.1. Пусть $A = (a_{ij}) \in M(n, k)$ — квадратная матрица над полем k. Ее **определителем** (или **детерминантом**) называется следующий элемент поля k:

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \cdot a_{1,\pi(1)} \cdot a_{2,\pi(2)} \cdot \dots \cdot a_{n,\pi(n)} = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) \prod_{i=1}^{n} a_{i,\pi(i)}.$$

Мы будем также использовать обозначение $|A| = \det(A)$.

Примеры 5.6.2.

- 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 5.6.3. *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, $(a'_{i1}, a'_{i2}, \ldots, a'_{in})$ — некоторая строка. Рассмотрим матрицу A', полученную заменой i-ой строки матрицы A на строку $(a'_{i1}, a'_{i2}, \ldots, a'_{in})$, и матрицу A'', полученную заменой i-ой строки матрицы A на строку $(a_{i1} + a'_{i1}, a_{i2} + a'_{i2}, \ldots, a_{in} + a'_{in})$. Схематично мы будем изображать это так:

$$A = \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{in} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}, A' = \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ a'_{i1} & a'_{i2} & \ldots & a'_{in} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix},$$
$$A'' = \begin{pmatrix} \vdots & \vdots & \ddots & \vdots \\ a_{i1} + a'_{i1} & a_{i2} + a'_{i2} & \ldots & a_{in} + a'_{in} \\ \vdots & \vdots & \ddots & \vdots \end{pmatrix}.$$

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 $u = (a_{i1}, a_{i2}, \ldots, a_{in})$, $v = (a'_{i1}, a'_{i2}, \ldots, a'_{in})$ (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'':

$$\det(A'') = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots (a_{i,\pi(i)} + a'_{i,\pi(i)}) \ldots a_{n,\pi(n)}$$

(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 — $a'_{i,\pi(i)}$:

$$\begin{aligned} \det(A'') &= \sum_{\pi \in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a_{i,\pi(i)} \ldots a_{n,\pi(n)} + \operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a'_{i,\pi(i)}) \ldots a_{n,\pi(n)} \right) \\ &= \sum_{\pi \in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a_{i,\pi(i)} \ldots a_{n,\pi(n)} \right) + \sum_{\pi \in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a'_{i,\pi(i)} \ldots a_{n,\pi(n)} \right). \end{aligned}$$

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 $(a_{i1}, a_{i2}, \ldots, a_{in})$ на строку $(\lambda a_{i1}, \lambda a_{i2}, \ldots, \lambda a_{in})$. Обозначим полученную матрицу через A'. Тогда $\det(A') = \lambda \det(A)$. Действительно, определитель матрицы A' равен

$$\det(A') = \sum_{\pi \in S_n} \left(\operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots (\lambda a_{i,\pi(i)}) \ldots a_{n,\pi(n)} \right).$$

В каждом слагаемом полученной суммы присутствует множитель λ. После вынесения его за скобки получаем

$$\det(A') = \lambda \left(\sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a_{i,\pi(i)} \ldots a_{n,\pi(n)} \right) = \lambda \det(A).$$

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 $l = 1, \ldots, n$, 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 5.5.11 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 $A_n \to S_n \setminus A_n$, $\pi \mapsto f \circ \tau$, 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:

$$\det(A) = \sum_{\pi \in S_n} \operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a_{i,\pi(i)} \ldots a_{j,\pi(j)} \ldots a_{n,\pi(n)}.$$

Теперь при помощи биекции f разобьем все слагаемые на пары, поставив в одну пару слагаемые, соответствующие перестановкам $\pi \in A_n$ и $f(\pi) = \pi \circ \tau \in S_n \setminus A_n$:

$$\det(A) = \sum_{\pi \in A_n} \big(\operatorname{sgn}(\pi) a_{1,\pi(1)} \ldots a_{i,\pi(i)} \ldots a_{n,\pi(n)} + \operatorname{sgn}(\pi\circ\tau) a_{1,(\pi\circ\tau)(1)} \ldots a_{i,(\pi\circ\tau)(i)} \ldots a_{j,(\pi\circ\tau)(j)} \ldots a_{n,(\pi\circ\tau)(n)} \big).$$

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 5.6.4. *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 5.6.5. *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. □

5.7 Дальнейшие свойства определителя

Литература: [K1], гл. 3, § 2, п. 2; [vdW], гл. 4, § 19.

Теорема 5.7.1 (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 $1, \dots, m$ встречаются среди чисел $\pi(1), \dots, \pi(m)$. 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 $1, \dots, m$ и $m+1, \dots, n$, не переставляя числа из разных наборов. Поэтому каждая такая перестановка однозначно определяет две перестановки: на числах $1, \dots, m$ и на числах $m+1, \dots, n$. Обозначим первую из них через ρ, а вторую сдвинем на m влево (чтобы получить перестановку чисел $1, \dots, n-m$, 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

$$\left(\sum_{\rho \in S_m} \operatorname{sgn}(\rho) a_{1,\rho(1)} \dots a_{m,\rho(m)} \right) \cdot \left(\sum_{\sigma \in S_{n-m}} \operatorname{sgn}(\sigma) a_{m+1,m+\sigma(1)} \dots a_{n,m+\sigma(n-m)} \right).$$

При раскрытии скобок в этом произведении получим сумму слагаемых вида

$$\operatorname{sgn}(\rho)\operatorname{sgn}(\sigma)a_{1,\rho(1)}\dots a_{m,\rho(m)}a_{m+1,m+\sigma(1)}\dots a_{n,m+\sigma(n-m)}$$

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 $1,\dots,m$ и перестановкой σ (сдвинутой на m вправо) на числах $m+1,\dots,n$.

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 $1,\dots,m$ и числами из набора $m+1,\dots,n$. □

Следствие 5.7.2. *Определитель верхнетреугольной матрицы равен произведению ее диагональных элементов:*

$$\left|\begin{pmatrix} a_1 & * & * & \dots & * \\ 0 & a_2 & * & \dots & * \\ 0 & 0 & a_3 & \dots & * \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_n \end{pmatrix}\right| = a_1a_2\dots a_n.$$

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 5.7.3. *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 5.7.4. *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 5.7.2. □

Теорема 5.7.5 (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

$$\begin{gathered}
\det(T_{ij}(\lambda)C) = \det(T_{ij}(\lambda))\det(C),\\
\det(S_{ij}C) = \det(S_{ij})\det(C),\\
\det(D_i(\varepsilon)C) = \det(D_i(\varepsilon))\det(C),\\
\det(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} C) = \det(\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix})\det(C).
\end{gathered}$$

Действительно, первые три равенства следуют из теоремы 5.6.5 и предложения 5.7.4. При $r < n$ матрица $\begin{pmatrix} E_r & 0 \\ 0 & 0 \end{pmatrix} C$ имеет нулевую строку, поэтому ее определитель равен нулю (предложение 5.7.3), как и произведение определителей сомножителей (в силу предложения 5.7.4. При $r = n$ указанная матрица является единичной, поэтому результат следует из следствия 5.7.2.

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 $\det(AB) = \det(A)\det(B)$. □

5.8 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 5.7.1 (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**: $\det(A) = a_{i1}\widetilde{A}_{i1} + a_{i2}\widetilde{A}_{i2} + \dots + a_{in}\widetilde{A}_{in}$. Транспонируя полученный результат, мы получаем **разложение определителя по столбцу**: $\det(A) = a_{1i}\widetilde{A}_{1i} + a_{2i}\widetilde{A}_{2i} + \dots + a_{ni}\widetilde{A}_{ni}$.

Сформулируем чуть более общий результат.

Теорема 5.8.1 (Соотношения ортогональности). *Пусть* $A \in M(n,k)$ *и* $1 \leqslant i \leqslant n$. *Тогда*

$$a_{i1}\widetilde{A}_{j1} + a_{i2}\widetilde{A}_{j2} + \dots + a_{in}\widetilde{A}_{jn} = \begin{cases} \det(A), & \text{если } i = j; \\ 0, & \text{если } i \neq j. \end{cases}.$$

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 5.8.2. *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 5.8.3. *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,

$$X = A^{-1}B = \frac{1}{\det(A)} \operatorname{adj}(A)B = \frac{1}{\det(A)} \begin{pmatrix} \widetilde{A}_{11} & \widetilde{A}_{21} & \dots & \widetilde{A}_{n1} \\ \widetilde{A}_{12} & \widetilde{A}_{22} & \dots & \widetilde{A}_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{A}_{1n} & \widetilde{A}_{2n} & \dots & \widetilde{A}_{nn} \end{pmatrix} \cdot \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

$$= \frac{1}{\det(A)} \begin{pmatrix} b_1\widetilde{A}_{11} + b_2\widetilde{A}_{21} + \dots + b_n\widetilde{A}_{n1} \\ b_1\widetilde{A}_{12} + b_2\widetilde{A}_{22} + \dots + b_n\widetilde{A}_{n2} \\ \vdots \\ b_1\widetilde{A}_{1n} + b_2\widetilde{A}_{2n} + \dots + b_n\widetilde{A}_{nn} \end{pmatrix}.$$

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 $b_1\widetilde{A}_{1i} + b_2\widetilde{A}_{2i} + \dots + b_n\widetilde{A}_{ni}$ — то, что стоит в столбце X на позиции i (с точностью до множителя $1/\det(A)$. Сформулируем полученный результат в виде теоремы.

Теорема 5.8.4 (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 5.8.5 (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 5.8.6 (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. $(u+v)\lambda = u\lambda + v\lambda$ для любых $u, v \in V$, $\lambda \in k$ (*левая дистрибутивность*);
6. $u(\lambda+\mu) = u\lambda + u\mu$ для любых $u \in V$, $\lambda, \mu \in k$ (*правая дистрибутивность*);
7. $u \cdot (\lambda \cdot \mu) = (u \cdot \lambda) \cdot \mu$ (*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 5.5.1); 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.

Примеры 6.1.4.

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. 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 $\mathbb{R}$.

5. 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.

6. 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 6.1.5. *Пусть V — векторное пространство над k. Тогда*

1. *$v \cdot 0 = 0$ для любого вектора $v \in V$.*

2. *$v \cdot (-1) = -v$ 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. 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 $-v = (-v) + v + v \cdot (-1) = 0 + v \cdot (-1) = v \cdot (-1)$. □

Определение 6.1.6. Пусть V — векторное пространство над полем k. Непустое подмножество $U \subseteq V$ называется **подпространством**, если выполнены следующие условия:

1. если $u, v \in U$, то и $u + v \in U$;

2. если $u \in U$, $\lambda \in k$, то $u\lambda \in U$.

Тот факт, что U является подпространством V, мы будем обозначать так: $U \leqslant V$.

Замечание 6.1.7. Если $U \leqslant V$, то $0 \in U$ и $-u \in U$ для любого $u \in U$. Действительно, так как U непусто, можно выбрать некоторый $v \in U$ и тогда в силу второго условия из определения подпространства должно выполняться $0 = v \cdot 0 \in U$. Кроме того, для любого $u \in U$ выполнено $-u = u \cdot (-1) \in U$.

Примеры 6.1.8.
1. В любом пространстве V есть «тривиальные» подпространства $0 \leqslant V$ и $V \leqslant V$.
2. 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 $k[x]_{\leqslant n} \leqslant k[x]$.
3. Множество векторов, параллельных некоторой плоскости, является подпространством трехмерного пространства векторов.

Лемма 6.1.9. *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 $\lambda \in k$, то для каждого α выполнено $u\lambda \in U_\alpha$, откуда $u\lambda \in \bigcap_{\alpha \in A} U_\alpha$. □

6.2 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 6.2.1. Пусть V — векторное пространство над k, $v_1, \ldots, v_n \in V$ и $\lambda_1, \ldots, \lambda_n \in k$. Выражение вида $v_1\lambda_1 + \cdots + v_n\lambda_n$ называется **линейной комбинацией** элементов $v_1, \ldots, v_n$. Отметим, что иногда линейной комбинацией называется сама формальная сумма $v_1\lambda_1 + \cdots + v_n\lambda_n$, а иногда — ее значение (то есть, элемент V).

Определение 6.2.2. Подмножество $X \subseteq V$ называется **линейно зависимым**, если существуют $v_1, \ldots, v_n \in X$ и $\lambda_1, \ldots, \lambda_n \in k$ такие, что $v_1\lambda_1 + \cdots + v_n\lambda_n = 0$, причем не все коэффициенты $\lambda_1, \ldots, \lambda_n$ равны нулю. Иными словами, набор векторов линейно зависим, если некоторая *нетривиальная* линейная комбинация векторов из этого набора равна нулю. Здесь мы называем линейную комбинацию $v_1\lambda_1 + \cdots + v_n\lambda_n$ **тривиальной**, если все ее коэффициенты равны нулю: $\lambda_1 = \cdots = \lambda_n = 0$, и **нетривиальной** в остальных случаях (то есть, когда найдется коэффициент λ_i, не равный нулю).

Определение 6.2.3. Подмножество $X \subseteq V$ называется **линейно независимым**, если оно не является линейно зависимым. Расшифруем это определение: $X \subseteq V$ линейно независимо, если для любых $v_1, \ldots, v_n \in X$ и $\lambda_1, \ldots, \lambda_n \in k$ из того, что $v_1\lambda_1 + \cdots + v_n\lambda_n = 0$ следует, что $\lambda_1 = \cdots = \lambda_n = 0$. Иными словами, набор векторов линейно независим, если никакая нетривиальная линейная комбинация векторов из этого набора не равна нулю.

Примеры 6.2.4. 1. Если $0 \in X$, то X линейно зависимо: линейная комбинация из одного слагаемого $0 \cdot 1$ равна нулю, но ее единственный коэффициент 1 не равен нулю.

2. Одноэлементное множество $\{v\}$ линейно зависимо тогда и только тогда, когда $v = 0$: если $v = 0$, то $\{v\}$ линейно зависимо в силу предыдущего примера, а если $v \neq 0$, то рассмотрим некоторую линейную комбинацию $v\lambda = 0$; если $\lambda \neq 0$, то после умножения на λ^{-1} получаем $v = 0$ — противоречие. Значит, эта линейная комбинация тривиальна.

3. Два вектора на эвклидовой плоскости линейно зависимы тогда и только тогда, когда они коллинеарны; три вектора в пространстве линейно зависимы тогда и только тогда, когда они компланарны.

Отметим очевидные свойства линейной зависимости и независимости: если набор векторов линейно независим, то и любое его подмножество линейно независимо. Обратно, если набор векторов линейно зависим, то и любое содержащее его множество линейно зависимо. Сформулируем это в виде леммы для последующих ссылок.

Лемма 6.2.5. *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.*

Посмотрим, когда при добавлении к линейно независимой системе одного вектора теряется линейная независимость.

Лемма 6.2.6. *Пусть подмножество $X \subseteq V$ линейно независимо, $v \in V$ и множество $X \cup \{v\}$ линейно зависимо. Тогда v является линейной комбинацией векторов из X.*

Доказательство. Пусть $v_1\lambda_1 + \cdots + v_n\lambda_n + v\lambda = 0$ — нетривиальная линейная комбинация векторов из $X \cup \{v\}$, равная нулю; здесь $v_1, \ldots, v_n \in X$, $\lambda_1, \ldots, \lambda_n \in k$. Если $\lambda = 0$, то уже множество X было линейно зависимым. Значит, $\lambda \neq 0$, и мы можем записать $v = -v_1\lambda_1/\lambda - \cdots - v_n\lambda_n/\lambda$. □

Определение 6.2.7. Пусть $X \subseteq V$ — некоторое подмножество. Пересечение всех подпространств V, содержащих X, обозначается через $\langle X \rangle$ и называется **линейной оболочкой** множества X: $\langle X \rangle = \bigcap_{U \leqslant V, U \supseteq X} U$. Заметим, что это пересечение имеет смысл, поскольку есть по крайней мере одно подпространство V, содержащее X: это само пространство V.

Лемма 6.2.8. *Пусть $X \subseteq V$. Линейная оболочка X — это множество всех линейных комбинаций элементов из X:*

$$\langle X \rangle = \{v_1\lambda_1 + \cdots + v_n\lambda_n \mid v_1, \ldots, v_n \in X, \lambda_1, \ldots, \lambda_n \in k\}.$$

Доказательство. Обозначим множество в правой части через X'. Заметим, что если U — подпространство в V, содержащее X, то U содержит и все линейные комбинации $v_1\lambda_1 + \cdots + v_n\lambda_n$ элементов $v_1, \ldots, v_n$ из X; поэтому U содержит X'. Значит, X' содержится и в пересечении $\bigcap_{U \leqslant V, U \supseteq X} U = \langle X \rangle$.

Обратно, множество X' является подпространством V, поскольку сумма линейных комбинацией элементов из X является линейной комбинацией элементов из X, и произведение такой линейной комбинации на скаляр снова является такой линейной комбинацией. Поэтому X' является одним из подпространств U, по которым происходит пересечение, а значит, X' содержит $\langle X \rangle$. □

Определение 6.2.9. Подмножество $X \subseteq V$ называется **системой образующих** пространства V (или **порождающей системой** пространства V), если его линейная оболочка совпадает с V: $\langle X \rangle = V$. Иными словами, X — система образующих V, если любой вектор V является линейной комбинацией элементов из X.

Определение 6.2.10. Пространство V называется **конечномерным**, если в нем существует конечная система образующих.

Определение 6.2.11. Пусть V — векторное пространство над полем k. Линейно независимая система образующих V называется **базисом** пространства V.

Теорема 6.2.12. *Из любой конечной системы образующих пространства V можно выбрать базис.*

Доказательство. Пусть $X \subseteq V$ — конечная система образующих. Если она не является линейно независимой, то некоторый вектор $x \in X$ выражается через остальные. Рассмотрим множество $X \setminus \{x\}$ и покажем, что оно также является системой образующих. Действительно, любой элемент $v \in V$ можно представить как линейную комбинацию элементов из X: $v = x_1\lambda_1 + \cdots + x_n\lambda_n + x\lambda$, где $x_i \in X \setminus \{x\}$ и $\lambda_1, \ldots, \lambda_n, \lambda \in k$. При этом элемент x сам выражается через элементы X. Подставляя это выражение в равенство для v, получаем, что v является линейной комбинацией элементов из $X \setminus \{x\}$, что и требовалось.

Теперь проделаем с полученной системой ту же операцию: если она не является линейно независимой, то выбросим из нее вектор, который выражается через остальные. Этот процесс не может продолжаться бесконечно в силу конечности исходного множества X. Значит, на некотором шаге мы получим линейно независимую систему, то есть, базис. □

6.3 Базис

Литература: [F], гл. XII, § 1, п. 2; [K2], гл. 1, § 2, п. 1–2; [KM], ч. 1, § 2; [vdW], гл. 4, § 20.

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 6.3.1. *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, $u_1\lambda_1 + \cdots + u_n\lambda_n = v = u_1\mu_1 + \cdots + u_n\mu_n$ для некоторых $u_i \in \mathcal{B}$, $\lambda_i \in k$, то $u_1(\lambda_1 - \mu_1) + \cdots + u_n(\lambda_n - \mu_n) = 0$, и из линейной независимости $\mathcal{B}$ следует, что все коэффициенты в этой линейной комбинации равны 0, откуда $\lambda_i = \mu_i$ 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 $v_1\lambda_1 + \cdots + v_n\lambda_n = 0 = v_1 \cdot 0 + \cdots + v_n \cdot 0$. Мы получили два различных представления одного вектора $0 \in V$ (они различны, поскольку не все λ_i равны нулю) — противоречие. □

Теорема 6.3.2. *Подмножество $\mathcal{B} \subseteq V$ является базисом тогда и только тогда, когда $\mathcal{B}$ — минимальная система образующих, то есть, система образующих пространства V такая, что при удалении любого вектора из $\mathcal{B}$ она перестает быть системой образующих.*

Доказательство. Если $\mathcal{B}$ — базис, то $\mathcal{B}$ — система образующих. Докажем ее минимальность: возьмем $v \in \mathcal{B}$ и рассмотрим систему $\mathcal{B}\setminus\{v\}$. Если она все еще является системой образующих, то, в частности, вектор v должен выражаться в виде линейной комбинации векторов из $\mathcal{B}\setminus\{v\}$: $v = \sum_i v_i\lambda_i$. Но тогда $v - \sum_i v_i\lambda_i = 0$ — нетривиальная линейная зависимость между векторами из $\mathcal{B}$, что противоречит линейной независимости $\mathcal{B}$.

Обратно, пусть $\mathcal{B}$ — минимальная система образующих. Для того, чтобы доказать, что $\mathcal{B}$ является базисом, достаточно проверить линейную независимость. Предположим, что $\mathcal{B}$ линейно зависима, тогда по лемме 6.2.6 некоторый вектор $v \in \mathcal{B}$ выражается через остальные. Но тогда $\mathcal{B} \setminus \{v\}$ также является системой образующих: действительно, любой вектор из V является линейной комбинацией векторов из $\mathcal{B}$, то есть, линейной комбинацией векторов из $\mathcal{B} \setminus \{v\}$ и вектора v; в этой линейной комбинации можно заменить вектор v на его выражение через векторы из $\mathcal{B} \setminus \{v\}$ и получить линейную комбинацию векторов из $\mathcal{B} \setminus \{v\}$. Получаем противоречие с минимальностью, поэтому наше предположение о линейной зависимости $\mathcal{B}$ было неверным. □

Теорема 6.3.3. *Подмножество $\mathcal{B} \subseteq V$ является базисом тогда и только тогда, когда $\mathcal{B}$ — максимальная линейно независимая система, то есть, линейно независимая система векторов пространства V, которая при добавлении любого вектора перестает быть линейно независимой.*

Доказательство. Пусть $\mathcal{B}$ — базис; тогда это линейно независимая система. Добавим к ней вектор $v \in V \setminus \mathcal{B}$; так как $\mathcal{B}$ — система образующих, то вектор v выражается через элементы $\mathcal{B}$. Поэтому система $\mathcal{B} \cup \{v\}$ линейно зависима. Это доказывает максимальность линейно независимой системы $\mathcal{B}$.

Предположим теперь, что $\mathcal{B}$ — максимальная линейно независимая система. Нам нужно показать, что $\mathcal{B}$ является системой образующих. Заметим, что любой вектор из $\mathcal{B}$ тривиальным образом является линейной комбинацией элементов из $\mathcal{B}$. Возьмем теперь $v \in V \setminus \mathcal{B}$.

По предположению максимальности система $\mathcal{B} \cup \{v\}$ линейно зависима. По лемме 6.2.6 это означает, что v является линейной комбинацией элементов из $\mathcal{B}$. Стало быть, любой вектор из V является линейной комбинацией элементов из $\mathcal{B}$, что и требовалось доказать. □

Соберем три предыдущих теоремы в одну:

Теорема 6.3.4. *Пусть V — векторное пространство, $\mathcal{B} \subseteq V$. Равносильны:*

1. *$\mathcal{B}$ — базис;*
2. *любой вектор из V единственным образом представляется в виде линейной комбинации векторов из V;*
3. *$\mathcal{B}$ — минимальная система образующих;*
4. *$\mathcal{B}$ — максимальная линейно независимая система.*

6.4 Размерность

Литература: [F], гл. XII, § 1, п. 2; [K2], гл. 1, § 2, п. 1–2; [KM], ч. 1, § 2; [vdW], гл. 4, § 19.

Теорема 6.4.1 (О линейной зависимости линейных комбинаций). *Пусть V — векторное пространство над k, $u_1, \ldots, u_m, v_1, \ldots, v_n \in V$. Если $u_1, \ldots, u_m \in \langle v_1, \ldots, v_n \rangle$ и $m > n$, то $u_1, \ldots, u_m$ линейно зависимы.*

Доказательство. Заметим, что достаточно рассмотреть случай $m = n+1$; если мы докажем, что векторы $u_1, \ldots, u_{n+1}$ линейно зависимы, то, по лемме 6.2.5 и $u_1, \ldots, u_m$ линейно зависимы. Будем вести индукцию по n.

База: $n = 1$; тогда $u_1, u_2 \in \langle v_1 \rangle$. Можно записать $u_1 = v_1 a_1$, $u_2 = v_1 a_2$. Заметим, что если $a_1 = 0$, то $u_1 = 0$, поэтому векторы u_1, u_2 линейно зависимы (см. примеры 6.2.4). Если же $a_1 \neq 0$, то $u_1 a_2 - u_2 a_1 = 0$ — нетривиальная линейная зависимость между u_1 и u_2.

Переход: пусть теперь $u_1, \ldots, u_m \in \langle v_1, \ldots, v_n \rangle$ и для всех меньших n утверждение теоремы доказано. Запишем

$$\begin{aligned} u_1 &= v_1 a_{11} + \cdots + v_n a_{1n}, \\ &\vdots \quad\quad \vdots \\ u_m &= v_1 a_{m1} + \cdots + v_n a_{mn}. \end{aligned}$$

Посмотрим на последние коэффициенты в каждой строчке: $a_{1n}, \ldots, a_{mn}$. Если все они равны 0, то каждый из $u_1, \ldots, u_m$ является линейной комбинацией векторов $v_1, \ldots, v_{n-1}$, то есть, лежит в $\langle v_1, \ldots, v_{n-1} \rangle$. По предположению индукции тогда $u_1, \ldots, u_m$ линейно зависимы. Если же какой-то из коэффициентов при v_n не равен 0, можно считать (после перенумерации векторов $u_1, \ldots, u_m$), что $a_{mn} \neq 0$. Посмотрим на векторы $u_1' = u_1 - u_m \frac{a_{1n}}{a_{mn}}, \ldots, u_{m-1}' = u_{m-1} - u_m \frac{a_{m-1,n}}{a_{mn}}$. Благодаря удачному подбору коэффициентов, каждый из них является линейной комбинацией векторов $v_1, \ldots, v_{n-1}$, то есть, лежит в $\langle v_1, \ldots, v_{n-1}$, и их количество равно $m-1$. Поскольку $m > n$, то $m - 1 > n - 1$, и можно применить индукционное предположение.

Стало быть, векторы $u'_1, \dots, u'_{m-1}$ линейно зависимы. Запишем эту линейную зависимость: $u'_1\lambda_1 + \dots + u'_{m-1}\lambda_{m-1} = 0$, причем не все λ_i равны 0. Подставим сюда определения векторов u'_i и раскроем скобки: $0 = (u_1 - u_m\frac{a_{1n}}{a_{mn}})\lambda_1 + \dots + (u_{m-1} - u_m\frac{a_{m-1,n}}{a_{mn}})\lambda_{m-1} = u_1\lambda_1 + \dots + u_{m-1}\lambda_{m-1} + u_m(\dots)$. Получили линейную зависимость между $u_1, \dots, u_m$, не все коэффициенты которой равны 0, что и требовалось. □

Следствие 6.4.2. *Пусть V — конечномерное векторное пространство. В любых двух базисах V поровну элементов.*

Доказательство. Конечномерность V означает, что в V имеется конечная система образующих. По лемме 6.2.12 из этого следует, что в V есть конечный базис. Обозначим его через $\mathcal{B}$, а число элементов в нем через n. Пусть теперь $\mathcal{B}'$ — другой базис. Докажем, что $|\mathcal{B}| = |\mathcal{B}'|$.

Если $|\mathcal{B}'| > |\mathcal{B}|$, то можно выбрать $n+1$ элементов в $\mathcal{B}'$. Они лежат в $\langle\mathcal{B}\rangle = V$ и $|\mathcal{B}| = n$, поэтому по теореме о линейной зависимости линейных комбинаций они линейно зависимы. Но тогда и система $\mathcal{B}'$ линейно зависима, что противоречит тому, что $\mathcal{B}'$ — базис V.

Если же $|\mathcal{B}'| < |\mathcal{B}|$, то все векторы из $\mathcal{B}$ лежат в линейной оболочке $\langle\mathcal{B}'\rangle = V$ векторов из $\mathcal{B}'$, и их больше. Значит, по той же теореме, векторы $\mathcal{B}$ линейно зависимы, что противоречит тому, что $\mathcal{B}$ — базис V. □

Это следствие позволяет нам корректно определить *размерность* векторного пространства как количество элементов в [любом] базисе.

Определение 6.4.3. 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**.

Следствие 6.4.4. *Пусть V — конечномерное векторное пространство, $X, Y \subseteq V$, причем X — линейно независимая система, а Y — система образующих V. Тогда $|X| \leqslant \dim V \leqslant |Y|$.*

Доказательство. Пусть $\mathcal{B}$ — базис пространства V. Заметим, что $\langle Y\rangle = V = \langle\mathcal{B}\rangle$. Если $|Y| < |\mathcal{B}|$, то по теореме 6.4.1 векторы из $\mathcal{B}$ линейно зависимы — противоречие. Если $|X| > |\mathcal{B}|$, то по той же теореме векторы из $|X|$ линейно зависимы — противоречие. □

Предложение 6.4.5. *Любую линейно независимую систему в конечномерном пространстве V можно дополнить до базиса V.*

Доказательство. Пусть X — линейно независимая система в V; по следствию 6.4.4 она конечна. Если X не является системой образующих, то $\langle X\rangle \neq V$, поэтому найдется некоторый вектор $v \in V$, не лежащий в $\langle X\rangle$. Тогда система $X \cup \{v\}$ также линейно независима; действительно, если $X \cup \{v\}$ линейно зависима, то по лемме 6.2.6 v выражается через X, то есть, лежит в $\langle X\rangle$ — противоречие. Добавим к X вектор v и повторим процедуру: если полученная

система не является системой образующих, то можно добавить к ней еще один вектор, и так далее. На каждом шаге количество элементов в X увеличивается на 1, и при этом всегда $|X| \leqslant \dim V$ по следствию 6.4.4. Значит, когда-нибудь процесс прекратится и мы получим систему образующих, то есть, базис V. □

Предложение 6.4.6. *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 следствию 6.4.4 имеем $|\mathcal{B}'| \leqslant n$, что и требовалось.

Очевидно, что если $U = V$, то $\dim_k U = \dim_k V$. Пусть теперь $\dim_k U = \dim_k V = n$, и снова $\mathcal{B}'$ — некоторый базис U. По предложению 6.4.5 линейно независимую систему $\mathcal{B}'$ можно дополнить до базиса пространства V; пусть $\mathcal{B}' \cup Y$ — базис V. По следствию 6.4.2 имеем $|\mathcal{B}' \cup Y| = \dim_k V = \dim_k U = |\mathcal{B}'|$. Отсюда $Y = \varnothing$ и, следовательно, $\mathcal{B}'$ является базисом V. Поэтому $U = \langle \mathcal{B}' \rangle = V$. □

Определение 6.4.7. Пусть $\mathcal{B} = \{e_1, \ldots, e_n\}$ — *упорядоченный* базис конечномерного пространства V над полем k ($n = \dim_k V$). В силу теоремы 6.3.1 любой вектор $v \in V$ можно однозначным образом записать в виде $v = e_1a_1 + \cdots + e_na_n$. Коэффициенты $a_1, \ldots, a_n$ называются **координатами** вектора v в базисе $\mathcal{B}$. Мы будем записывать их в столбец

$$[v]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix},$$

называемый **координатным столбцом** вектора v в базисе $\mathcal{B}$.

Итак, при наличии фиксированного базиса $\mathcal{B}$ пространства V мы можем сопоставить каждому вектору $v \in V$ столбец $[v]_{\mathcal{B}} \in k^n$. Обратно, по столбцу $\begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \in k^n$ мы можем восстановить вектор $v \in V$: $v = e_1a_1 + \cdots + e_na_n$. Мы получили взаимно однозначное соответствие между элементами V и элементами k^n. Более того, эта биекция сохраняет операции:

Предложение 6.4.8. *Если $\mathcal{B}$ — базис пространства V, $u, v \in V$, $\lambda \in k$, то $[u+v]_{\mathcal{B}} = [u]_{\mathcal{B}} + [v]_{\mathcal{B}}$ и $[u\lambda]_{\mathcal{B}} = [u]_{\mathcal{B}}\lambda$.*

Доказательство. Пусть $[u]_{\mathcal{B}} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$, $[v]_{\mathcal{B}} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$. Это означает, что $u = e_1a_1 + \cdots + e_na_n$ и $v = e_1b_1 + \cdots + e_nb_n$. Складывая эти два равенства, получаем, что $u + v = e_1(a_1 + b_1) + \cdots +$

$e_n(a_n + b_n)$. Поэтому

$$[u+v]_{\mathcal{B}} = \begin{pmatrix} a_1 + b_1 \\ a_2 + b_2 \\ \vdots \\ a_n + b_n \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = [u]_{\mathcal{B}} + [v]_{\mathcal{B}}.$$

Кроме того, $u\lambda = e_1 a_1 \lambda + \cdots + e_n a_n \lambda$, поэтому

$$[u\lambda]_{\mathcal{B}} = \begin{pmatrix} a_1\lambda \\ a_2\lambda \\ \vdots \\ a_n\lambda \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \lambda = [u]_{\mathcal{B}}\lambda.$$

□

6.5 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 6.5.1. 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

$$\left\langle \begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \ldots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix} \right\rangle \leqslant k^m.$$

Линейная оболочка строк матрицы A называется **пространством строк матрицы** A; по определению оно является подпространством в ${}^n k$. Иными словами, это пространство

$$\left\langle \begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n} \end{pmatrix}, \ldots, \begin{pmatrix} a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix} \right\rangle \leqslant {}^n k.$$

Таким образом, пространство столбцов состоит из всевозможных линейных комбинаций столбцов матрицы A; аналогично и со строками.

Определение 6.5.2. 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 6.5.3. 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 6.5.4. *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 6.5.5. *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 $A = e_1 a_{1*} + e_2 a_{2*} + \cdots + e_m a_{m*}$, 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 6.4.4 получаем, что $d \leqslant r$.

Доказательство для столбцового ранга совершенно аналогично (или можно заметить, что тензорный ранг не меняется при транспонировании). □

Определение 6.5.6. 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 6.5.7. *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 6.5.5 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 6.5.8. *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 6.5.7. □

Теорема 6.5.9 (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 $u_1, \ldots, u_n$ — столбцы матрицы A. Система $AX = B$ имеет решение тогда и только тогда, когда существуют $x_1, \ldots, x_n \in k$ такие, что $u_1x_1 + \cdots + u_nx_n = B$. Это, в свою очередь равносильно тому, что B лежит в линейной оболочке векторов $u_1, \ldots, u_n$, то есть, тому, что $\langle u_1, \ldots, u_n \rangle = \langle u_1, \ldots, u_n, B \rangle$. Это равенство и означает совпадение [столбцовых] рангов матриц A и $(A|B)$.

Если же ранг равен количеству неизвестных n, то пространство $\langle u_1, \ldots, u_n \rangle$ имеет размерность n. При этом $\langle u_1, \ldots, u_n \rangle$ — его система образующих, и из нее можно выбрать базис, в котором должно быть n элементов. Значит, $u_1, \ldots, u_n$ образуют базис пространства столбцов матрицы A. Поэтому вектор B имеет единственное представление в виде $B = u_1x_1 + \cdots + u_nx_n$, что и означает единственность решения системы. □

Определение 6.5.10. Пусть $A \in M(m, n, k)$. **Минором** матрицы A называется квадратная матрица, полученная из A вычеркиванием некоторых строк и некоторых столбцов. Равносильно, это матрица, образованная пересечением некоторого количества строк матрицы A и того же количества столбцов матрицы A. Иногда минором называют не саму такую матрицу, а ее определитель.

Теорема 6.5.11. *Пусть $A \in M(m, n, k)$. Ранг матрицы A не превосходит r тогда и только тогда, когда любой минор порядка $r + 1$ матрицы A вырожден (то есть, его определитель равен нулю).*

Доказательство. Пусть $\mathrm{rk}(A) \leqslant r$, M — некоторый минор матрицы A порядка $r + 1$. Он составлен из $r + 1$ строк матрицы A. Эти строки можно рассматривать как элементы пространства строк, имеющего размерность не более r. Поэтому эти строки линейно зависимы. Но тогда и строки матрицы M линейно зависимы — между ними есть линейная зависимость с теми же коэффициентами, что и между строками матрицы A.

Обратно, если ранг матрицы A строго больше r, то он равен хотя бы $r+1$. Поэтому из строк матрицы A можно выбрать линейно независимую систему из $r + 1$ строки (по теореме 6.2.12 можно выбрать базис, а потом при необходимости уменьшить его). Матрица $(r + 1) \times n$, составленная из этих строк, имеет ранг $r + 1$. Поэтому из ее столбцов (совершенно аналогичным образом) можно выбрать систему из линейно независимых столбцов в количестве $r + 1$ штук. Полученный набор строк и столбцов определяет квадратный минор порядка $r+1$, ранг которого равен $r + 1$. По следствию 6.5.8 из этого следует его невырожденность. □

6.6 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 6.6.1. 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

$$\begin{pmatrix} v_1 & v_2 & \ldots & v_n \end{pmatrix} = \begin{pmatrix} u_1 & u_2 & \ldots & u_n \end{pmatrix} \cdot (\mathcal{B} \rightsquigarrow \mathcal{B}').$$

Предложение 6.6.2 (Свойства матрицы перехода). *Пусть* $\mathcal{B} = \{u_1, \ldots, u_n\}$, $\mathcal{B}' = \{v_1, \ldots, v_n\}$, $\mathcal{B}'' = \{w_1, \ldots, w_n\}$ — *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 $\{u_1, \ldots, u_n\}$ равен e_i, то есть, равен i-му столбцу единичной матрицы.

2. Мы знаем, что
$$(w_1, \ldots, w_n) = (u_1, \ldots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}'').$$
С другой стороны, $(w_1, \ldots, w_n) = (v_1, \ldots, v_n)(\mathcal{B}' \rightsquigarrow \mathcal{B}'') = (u_1, \ldots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}')(\mathcal{B}' \rightsquigarrow \mathcal{B}'')$. Поэтому
$$(u_1, \ldots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}'') = (u_1, \ldots, u_n)(\mathcal{B} \rightsquigarrow \mathcal{B}')(\mathcal{B}' \rightsquigarrow \mathcal{B}'').$$
Поскольку $(u_1, \ldots, u_n)$ является базисом, из равенства линейных комбинаций векторов $u_1, \ldots, u_n$ 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 6.6.3. *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 6.6.4. 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. □

7 Линейные отображения

7.1 Определения и примеры

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 U, V — векторные пространства над полем k. Отображение $\varphi\colon U \to V$ называется **линейным**, если выполняются следующие два свойства:

1. $\varphi(u_1 + u_2) = \varphi(u_1) + \varphi(u_2)$ для любых $u_1, u_2 \in U$;
2. $\varphi(u\lambda) = \varphi(u)\lambda$ для любых $\lambda \in k, u \in U$.

Линейное отображение часто называется **гомоморфизмом** векторных пространств; оно называется **эндоморфизмом**, если $U = V$.

Примеры 7.1.2.

1. Изоморфизм векторных пространств — это биективное линейное отображение.
2. Проекция векторов эвклидовой плоскости E на некоторую прямую L является линейным отображением $E \to L$; можно рассматривать ее и как линейное отображение $E \to E$.
3. Нулевое отображение $0\colon U \to V$, переводящее любой вектор $u \in U$ в $0 \in V$, является линейным отображением.
4. Тождественное отображение $\mathrm{id}\colon V \to V$ является линейным.
5. Умножение на любой элемент $\lambda \in k$ является эндоморфизмом $- \cdot \lambda\colon V \to V$.
6. Координатная проекция $\mathrm{pr}_i\colon k^n \to k$, сопоставляющая столбцу его i-ю компоненту, линейная. Вообще, если V — векторное пространство, а $\{e_1, \ldots, e_n\}$ — базис в нем, рассмотрим координатную функцию e_i^*, сопоставляющую каждому вектору $v \in V$ его i-ю координату в этом базисе (то есть, коэффициент при e_i в разложении v по этому базису). Из единственности разложения по базису следует, что e_i^* линейно.
7. Для любой матрицы $A \in M(m, n, k)$ отображение $k^n \to k^m$, $u \mapsto Au$, задаваемое умножением на матрицу A слева, является линейным: $A(u_1+u_2) = Au_1+Au_2$, $A(u\lambda) = A(u)\lambda$. На самом деле, чуть позже (следствие 7.8.2) мы узнаем, что никаких других линейных отображений между конечномерными векторными пространствами нет: любое линейное отображение после выбора базисов (и отождествления, таким образом, наших пространств с некоторыми пространствами столбцов) задается умножением на некоторую матрицу.

7.2 Фактор-пространство

Литература: [F], гл. XII, § 2, п. 5; [K2], гл. 1, § 2, п. 6; [KM], ч. 1, § 6.

Определение 7.2.1. 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. 2.8), 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 $(v + U)\lambda = v\lambda + U$ для любых $v, v_1, v_2 \in V$ и $\lambda \in k$. В других обозначениях, $\overline{v_1} + \overline{v_2} = \overline{v_1 + v_2}$ и $\overline{v} \cdot \lambda = \overline{v \cdot \lambda}$. 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 7.2.2. *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. $(\overline{v_1} + \overline{v_2}) \cdot \lambda = \overline{v_1 + v_2} \cdot \lambda = \overline{(v_1 + v_2) \cdot \lambda} = \overline{v_1\lambda + v_2\lambda} = \overline{v_1\lambda} + \overline{v_2\lambda} = \overline{v_1} \cdot \lambda + \overline{v_2} \cdot \lambda$.

6. $\overline{v}(\lambda + \mu) = \overline{v(\lambda + \mu)} = \overline{v\lambda + v\mu} = \overline{v\lambda} + \overline{v\mu} = \overline{v} \cdot \lambda + \overline{v} \cdot \mu$.

7. $\overline{v}(\lambda\mu) = \overline{v(\lambda\mu)} = \overline{(v\lambda)\mu} = \overline{v\lambda} \cdot \mu = (\overline{v} \cdot \lambda) \cdot \mu$.

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.

7.3 Ядро и образ линейного отображения

Литература: [F], гл. XII, § 4, п. 1; [K2], гл. 2, § 1, пп. 1, 3; [KM], ч. 1, § 3.

Пусть $\varphi\colon U \to V$ — линейное отображение. Мы будем называть **образом** φ его теоретико-множественный образ, то есть, $\operatorname{Im}(\varphi) = \{\varphi(u) \mid u \in U\}$. Более строго, $\operatorname{Im}(\varphi) = \{v \in V \mid v = \varphi(u)$ для некоторого $u \in U\}$. **Ядром** отображения φ называется прообраз $0 \in V$, то есть, $\operatorname{Ker}(\varphi) = \{u \in U \mid \varphi(u) = 0\}$.

Предложение 7.3.1. $\operatorname{Im}(\varphi) \leqslant V$, $\operatorname{Ker}(\varphi) \leqslant U$.

Доказательство. Пусть $v_1, v_2 \in \operatorname{Im}(\varphi)$, тогда $v_1 = \varphi(u_1)$, $v_2 = \varphi(u_2)$ для некоторых $u_1, u_2 \in U$. Но тогда $v_1 + v_2 = \varphi(u_1) + \varphi(u_2) = \varphi(u_1 + u_2)$, поэтому $v_1 + v_2 \in \operatorname{Im}(\varphi)$. Наконец, если $v \in \operatorname{Im}(\varphi)$ и $\lambda \in k$, то $v = \varphi(u)$ для $u \in U$ и $v\lambda = \varphi(u)\lambda = \varphi(u\lambda)$, поэтому и $v\lambda \in \operatorname{Im}(\varphi)$.

Теперь пусть $u_1, u_2 \in \operatorname{Ker}(\varphi)$. Это означает, что $\varphi(u_1) = \varphi(u_2) = 0$. Тогда $\varphi(u_1 + u_2) = \varphi(u_1) + \varphi(u_2) = 0 + 0 = 0$, поэтому и $u_1 + u_2 = 0$. Если же $u \in \operatorname{Ker}(\varphi)$ и $\lambda \in k$, то $\varphi(u) = 0$, откуда $\varphi(u\lambda) = \varphi(u)\lambda = 0\lambda = 0$, поэтому и $u\lambda \in \operatorname{Ker}(\varphi)$. □

Теорема 7.3.2. *Пусть $\varphi\colon U \to V$ — линейное отображение.*

1. *φ инъективно тогда и только тогда, когда $\operatorname{Ker}(\varphi) = 0$;*
2. *φ сюръективно тогда и только тогда, когда $\operatorname{Im}(\varphi) = V$;*
3. *φ является изоморфизмом тогда и только тогда, когда $\operatorname{Ker}(\varphi) = 0$ и $\operatorname{Im}(\varphi) = v$.*

Доказательство. 1. Если φ инъективно, то из $\varphi(u) = 0 = \varphi(0)$ следует, что $u = 0$. Обратно, если $\operatorname{Ker}(\varphi) = 0$ и $\varphi(u_1) = \varphi(u_2)$, то $\varphi(u_1 - u_2) = \varphi(u_1) - \varphi(u_2) = 0$, поэтому $u_1 - u_2 \in \operatorname{Ker}(\varphi) = 0$ и, стало быть, $u_1 = u_2$. Это доказывает инъективность φ.

2. Очевидно в силу того, что Im означает теоретико-множественный образ.

3. Простая комбинация первых двух частей.

□

Теорема 7.3.3 (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 $\lambda \in k$, то $f(\overline{u})\lambda = \varphi(u)\lambda$ и $f(\overline{u}\cdot\lambda) = f(\overline{u\lambda}) = \varphi(u\lambda) = \varphi(u)\lambda$.

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. □

7.4 Сумма подпространств

Литература: [F], гл. XII, § 2, пп. 2, 3; [K2], гл. 1, § 2, пп. 4, 5; [KM], ч. 1, § 5; [vdW], гл. XII, § 87.

Пусть U, V — подпространства векторного пространства W. **Суммой** подпространств U и V называется линейная оболочка их объединения: $U + V = \langle U \cup V\rangle$. Иными словами, это наименьшее подпространство в W, содержащее и U, и V. Из описания линейной оболочки (7.3.2) немедленно следует, что $U+V$ состоит из всевозможных попарных сумм элементов U и V: $U + V = \{u + v \mid u \in U, v \in V\}$.

Сейчас мы определим *внутреннюю* и *внешнюю* прямую сумму. Различие между этими понятиями состоит в том, что внешняя прямая сумма определяется для двух «отдельных» подпространств, а внутренняя сумма — это «разложение» пространства в сумму двух [хороших] подпространств.

Итак, пусть U и V — векторные пространства над k. Рассмотрим декартово произведение множеств U и V, то есть, множество всевозможных упорядоченных пар (u, v), где $u \in U$, $v \in V$. Наша цель — ввести на этом декартовом произведении структуру векторного пространства над k, пользуясь структурами векторного пространства на U и V. Для этого определим операции сложения пар и умножения пары на скаляр следующим образом: $(u_1, v_1) + (u_2, v_2) = (u_1 + u_2, v_1 + v_2)$, $(u, v)\cdot\lambda = (u\lambda, v\lambda)$. Несложно проверить (упражнение!), что декартово произведение U и V с так введенными операциями является векторным пространством над k. Мы будем обозначать его через $U \oplus V$ и называть **внешней прямой суммой** пространств U и V.

Теперь пусть U, V — два подпространства в W. Будем говорить, что W является **внутренней прямой суммой** подпространств U и V, если $U \cap V = 0$ и $U + V = W$ (здесь мы, как всегда, обозначаем подпространство $\{0\}$, состоящее лишь из нулевого вектора, через 0).

Теорема 7.4.1. *Пусть W — векторное пространство над k. Если W является внутренней прямой суммой подпространств U и V, то отображение $f\colon U \oplus V \to W$, $(u, v) \mapsto u+v$, устанавливает изоморфизм между внешней прямой суммой $U \oplus V$ и пространством W.*

Доказательство. Очевидно, что указанное отображение линейно. В силу теоремы 7.3.2 достаточно проверить, что его ядро тривиально, а образ совпадает с W. Поскольку $U + V = W$, любой элемент $w \in W$ можно представить в виде суммы $w = u + v$ для некоторых $u \in U$, $v \in V$. Но тогда $f((u, v)) = u+v = w$, поэтому $\mathrm{Im}(f) = W$. Наконец, если $f((u, v)) = u+v = 0$, то $u = -v$ — элемент, лежащий одновременно в подпространствах U и V. Из условия $U \cap V = 0$ следует, что $u = -v = 0$, поэтому $(u, v) = (0, 0)$ — нулевой элемент пространства $U \oplus V$. Поэтому $\mathrm{Ker}(f) = 0$. □

В силу этой теоремы мы не будем более делать различия между внутренней и внешней прямой суммой: запись $W = U \oplus V$ означает, что W является внешней прямой суммой U и V, но при этом в W имеются подпространства $\{(u, 0) \mid u \in U\}$ и $\{(0, v) \mid v \in V\}$, изоморфные U и V соответственно. Мы часто будем отождествлять указанные пары пространств.

Пусть теперь $W = U \oplus V$. В такой ситуации имеется несколько естественных линейных отображений:

$$\begin{aligned} &\mathrm{pr}_1\colon U \oplus V \to U, \quad (u, v) \mapsto u \\ &\mathrm{pr}_2\colon U \oplus V \to V, \quad (u, v) \mapsto v \\ &i_1\colon U \to U \oplus V, \quad u \mapsto (u, 0) \\ &i_2,\colon V \to U \oplus V, \quad v \mapsto (0, v) \end{aligned}$$

При этом нетрудно проверить, что $\mathrm{pr}_1 \circ i_1 = \mathrm{id}_U$, $\mathrm{pr}_2 \circ i_2 = \mathrm{id}_V$. Посмотрим на отображение $\mathrm{pr}_1\colon U \oplus V \to U$. Его ядро состоит из пар (u, v) таких, что $u = 0$, то есть, из пар $(0, v)$. Это означает, что $\mathrm{Ker}(\mathrm{pr}_1) \cong V$. В то же время, легко видеть что pr_1 сюръективно, то есть, $\mathrm{Im}(\mathrm{pr}_1) = U$. По теореме о гомоморфизме получаем $(U \oplus V)/\mathrm{Ker}(\mathrm{pr}_1) \cong \mathrm{Im}(\mathrm{pr}_1)$, то есть, $(U \oplus V)/V \cong U$. Аналогично, $(U \oplus V)/U \cong V$.

7.5 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 7.5.1. Набор векторов $v_1, \ldots, v_n \in V$ называется **линейно независимым над** U, если из $v_1\lambda_1 + \ldots v_n\lambda_n \in U$ следует, что $\lambda_1 = \cdots = \lambda_n = 0$. Набор векторов $v_1, \ldots, v_n \in V$ называется **порождающей системой над** U (или **системой образующих** V **над** U), если любой вектор V можно

представить в виде $v_1\lambda_1 + \cdots + v_n\lambda_n + u$ для некоторых $\lambda_1, \ldots, \lambda_n \in k$ 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 $v_1\lambda_1 + \cdots + v_n\lambda_n + u$ для некоторого $u \in U$ *единственным образом.*

Теорема 7.5.2. *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 $(v_1+U)\lambda_1+\cdots+(v_n+U)\lambda_n = 0 \in V/U$, то $(v_1\lambda_1+\cdots+v_n\lambda_n)+U = 0 \in V/U$. Это означает, что $v_1\lambda_1+\cdots+v_n\lambda_n \in U$, откуда по определению линейной независимости над U следует $\lambda_1 = \cdots = \lambda_n = 0$. Кроме того, любой вектор $v \in V$ можно представить в виде $v = v_1\lambda_1 + \cdots + v_n\lambda_n + u$ для некоторых $\lambda_1, \ldots, \lambda_n \in k$ и $u \in U$. Тогда $\overline{v} = \overline{v_1}\lambda_1 + \cdots + \overline{v_n}\lambda_n$, поскольку $\overline{u} = 0$. Значит, $\overline{v_1}, \ldots, \overline{v_n}$ — система образующих V/U.

$1 \Rightarrow 2$ 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 $v+U = (v_1+U)\lambda_1+\cdots+(v_n+U)\lambda_n = (v_1\lambda_1+\cdots+v_n\lambda_n)+U$. Поэтому $v \sim_U v_1\lambda_1+\cdots+v_n\lambda_n$ и $v-(v_1\lambda_1+\cdots+v_n\lambda_n) = u \in U$. Разложим вектор u по базису $u_1, \ldots, u_k$: $u = u_1\mu_1 + \cdots + u_k\mu_k$. Получаем, что $v = v_1\lambda_1 + \cdots+v_n\lambda_n+u_1\mu_1+\cdots+u_k\mu_k$. Это доказывает, что $v_1, \ldots, v_n, u_1, \ldots, u_k$ — базис V. Наконец, если $v_1\lambda_1 + \cdots + v_n\lambda_n + u_1\mu_1 + \cdots + u_k\mu_k = 0$, то $v_1\lambda_1 + \cdots + v_n\lambda_n = -u_1\mu_1 - \cdots - u_k\mu_k \in U$, поэтому $\overline{v_1\lambda_1 + \cdots + v_n\lambda_n} = \overline{0}$, и в силу линейной независимости $\overline{v_1}, \ldots, \overline{v_n}$ в V/U из этого следует, что $\lambda_1 = \cdots = \lambda_n = 0$.

$3 \Rightarrow 4$ 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 $v = v_1\lambda_1+\cdots+v_n\lambda_n \in U$, то $v \in U \cap U'$, откуда $v = 0$, и в силу линейной независимости v_i, получаем $\lambda_1 = \cdots = \lambda_n = 0$. Наконец, любой вектор $v \in V$ можно представить в виде $v = u+u'$ для некоторых $u \in U$, $u' \in U'$. Запишем $u' = v_1\lambda_1 + \cdots + v_n\lambda_n$; получаем, что $v = v_1\lambda_1 + \cdots + v_n\lambda_n + u$.

□

Следствие 7.5.3. *Пусть $U \leqslant V$ — векторные пространства. Тогда* $\dim(V) = \dim(U) + \dim(V/U)$.

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 7.5.2 набор $u_1, \ldots, u_k, v_1, \ldots, v_n$ является базисом в V, состоящим из $k+n$ элементов. □

Следствие 7.5.4. *Пусть* $\varphi\colon U \to V$ *— гомоморфизм векторных пространств. Тогда*

$$\dim(\operatorname{Ker}(\varphi)) + \dim(\operatorname{Im}(\varphi)) = \dim(U).$$

Доказательство. По теореме о гомоморфизме (7.3.3) имеем $U/\operatorname{Ker}(\varphi) \cong \operatorname{Im}(\varphi)$. Выбирая базисы в $\operatorname{Ker}(\varphi)$ и $U/\operatorname{Ker}(\varphi)$ и пользуясь предыдущим следствием, получаем нужное равенство. □

Следствие 7.5.5. *Пусть* $W = U \oplus V$. *Тогда* $\dim(W) = \dim(U) + \dim(V)$.

Доказательство. Обсуждение в конце раздела 7.4 показывает, что $U \cong W/V$; осталось применить предыдущее следствие. □

Следствие 7.5.6. *Пусть* U, V *— подпространства векторного пространства* W. *Тогда* $\dim(U+V) = \dim(U) + \dim(V) - \dim(U \cap V)$.

Доказательство. Рассмотрим отображение $\psi\colon U \oplus V \to U+V$, заданное формулой $(u, v) \mapsto u + v$. Очевидно, что оно линейно и сюръективно (по определению суммы $U+V$). Кроме того, если $(u, v) \in \operatorname{Ker}(\psi)$, то $u + v = 0$, поэтому $u = -v \in U \cap V$. Поэтому $\operatorname{Ker}(\psi)$ можно отождествить с подпространством $U \cap V \leqslant V$, сопоставив паре $(u, v) \in \operatorname{Ker}(\psi)$ вектор u. По следствию 7.5.4 имеем $\dim(U \cap V) + \dim(U+V) = \dim(U \oplus V)$, а по следствию 7.5.5 правая часть равна $\dim(U) + \dim(V)$. □

7.6 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 $f, g\colon U \to V$ — два линейных отображения. Определим их **сумму** $f+g$ поточечным образом: положим $(f+g)(u) = f(u)+g(u)$ для любого $u \in U$. Получим некоторое отображение $f+g\colon U \to V$. Покажем, что оно является линейным. Действительно, $(f+g)(u_1+u_2) = f(u_1+u_2)+g(u_1+u_2) = f(u_1)+f(u_2)+g(u_1)+g(u_2) = (f+g)(u_1)+(f+g)(u_2)$ и $(f+g)(u\lambda) = f(u\lambda)+g(u\lambda) = f(u)\lambda + g(u)\lambda = (f(u)+g(u))\lambda = ((f+g)(u))\lambda$.

Мы научились складывать линейные отображения. Обозначим множество всех линейных отображений из U в V через $\operatorname{Hom}(U, V)$ (иногда мы будем писать $\operatorname{Hom}_k(U, V)$, если нужно подчеркнуть, над каким полем рассматриваются пространства U и V). Мы ввели на $\operatorname{Hom}(U, V)$ бинарную операцию сложения. Очевидно, что эта операция ассоциативна, коммутативна, обладает нейтральным элементом (это нулевое отображение) и у каждого $f \in \operatorname{Hom}(U, V)$ есть противоположный, определенный равенством $(-f)(u) = -f(u)$ для всех $u \in U$.

Научимся теперь умножать линейные отображения на константы. Пусть $f \in \operatorname{Hom}(U, V)$, $\lambda \in k$. Определим $f\lambda \in \operatorname{Hom}(U, V)$ равенством $(f\lambda)(u) = f(u)\lambda$. Нетрудно видеть (упражнение!), что сложение и умножение на скаляры задают на $\operatorname{Hom}(U, V)$ структуру векторного пространства над k.

Пусть теперь $f\colon U \to V$ и $g\colon V \to W$ — два линейных отображения. Тогда определена их композиция $g \circ f$. Покажем, что она также является линейным отображением. Действительно, $(g \circ f)(u_1 + u_2) = g(f(u_1 + u_2)) = g(f(u_1) + f(u_2)) = g(f(u_1)) + g(f(u_2)) = (g \circ f)(u_1) + (g \circ f)(u_2)$ и $(g \circ f)(u\lambda) = g(f(u\lambda)) = g(f(u)\lambda) = g(f(u))\lambda = ((g \circ f)(u))\lambda$.

Кроме того, если $f\colon U \to V$ и $g_1, g_2\colon V \to W$ — линейные отображения, то $(g_1 + g_2) \circ f = g_1 \circ f + g_2 \circ f$. Действительно, для проверки совпадения этих двух линейных отображений достаточно проверить, что $((g_1+g_2)\circ f)(u) = (g_1 \circ f + g_2 \circ f)(u)$ для всех $u \in U$. Но левая часть равна $(g_1+g_2)(f(u)) = g_1(f(u))+g_2(f(u))$, а правая часть равна $(g_1 \circ f)(u)+(g_2 \circ f)(u) = g_1(f(u))+ g_2(f(u))$. Это означает, что композиция линейных отображений дистрибутивна относительно сложения; аналогично, если $f_1, f_2\colon U \to V$ и $g\colon V \to W$ — линейные отображения, то $g \circ (f_1 + f_2) = g \circ f_1 + g \circ f_2$, поскольку $(g \circ (f_1 + f_2))(u) = g((f_1 + f_2)(u)) = g(f_1(u) + f_2(u)) = g(f_1(u)) + g(f_2(u)) = (g \circ f_1)(u) + (g \circ f_2)(u) = (g \circ f_1 + g \circ f_2)(u)$.

Рассмотрим теперь множество $\mathrm{Hom}(U, U)$ для векторного пространства U. Напомним, что линейное отображение из U в U мы называем также *эндоморфизмом* пространства U; поэтому иногда мы будем обозначать $\mathrm{Hom}(U, U)$ через $\mathrm{End}(U)$. Мы знаем, что $\mathrm{End}(U)$ само является векторным пространством. Кроме того, на нем определена операция композиции. Оказывается, $\mathrm{End}(U)$ является ассоциативным кольцом с единицей относительно сложения линейных отображений и их композиции. Действительно, необходимые свойства сложения мы уже проверили, когда доказывали, что $\mathrm{End}(U)$ является векторным пространством, дистрибутивность была проверена только что, id_U играет роль единичного элемента относительно композиции и в теоретико-множественном случае, и ассоциативность композиции мы доказали еще в теореме 1.3.1. Кольцо $\mathrm{End}(U)$ называется **кольцом эндоморфизмов** векторного пространства U. Отметим, что это кольцо не обязано быть коммутативным: чуть позже мы докажем, что оно изоморфно кольцу квадратных матриц размера $\dim(U)$ над полем k, некоммутативность которого при $\dim(U) \geqslant 2$ нам уже известна (замечание 5.3.5).

7.7 Универсальное свойство базиса

Литература: [КМ], ч. 1, § 3, п. 3.

Следующая теорема говорит, что для задания линейного отображения из пространства U достаточно задать его на каком-нибудь базисе U.

Теорема 7.7.1. *Пусть U, V — векторные пространства над k, $u_1, \ldots, u_n$ — базис U, и $x_1, \ldots, x_n \in V$ — некоторые элементы V. Тогда существует единственное линейное отображение $\varphi\colon U \to V$ такое, что $\varphi(u_i) = x_i$ для всех $i = 1, \ldots, n$.*

Доказательство. Пусть $u \in U$. Разложим u по базису $u_1, \ldots, u_n$: $u = u_1\lambda_1 + \cdots + u_n\lambda_n$ и положим $\varphi(u) = x_1\lambda_1 + \cdots + x_n\lambda_n$. В силу однозначности разложения по базису мы получили корректно определенное отображение $\varphi\colon U \to V$. При этом, очевидно, $\varphi(u_i) = x_i$. Проверим линейность φ.

Если, кроме того, $u' \in U$ и $u' = u_1\mu_1 + \cdots + u_n\mu_n$, то $\varphi(u) = x_1\lambda_1 + \cdots + x_n\lambda_n$, $\varphi(u') = x_1\mu_1 + \cdots + x_n\mu_n$, откуда $\varphi(u) + \varphi(u') = x_1(\lambda_1 + \mu_1) + \cdots + x_n(\lambda_n + \mu_n)$. Но $u + u' = u_1(\lambda_1 +$

$\mu_1) + \cdots + u_n(\lambda_n + \mu_n)$ — разложение по нашему базису вектора $u + u'$, поэтому $\varphi(u + u') = x_1(\lambda_1 + \mu_1) + \cdots + x_n(\lambda_n + \mu_n) = \varphi(u) + \varphi(u')$.

Наконец, если $\lambda \in k$, то $\varphi(u)\lambda = (x_1\lambda_1 + \cdots + x_n\lambda_n)\lambda = x_1\lambda_1\lambda + \cdots + x_n\lambda_n\lambda$. С другой стороны, $u\lambda = u_1\lambda_1\lambda + \cdots + u_n\lambda_n\lambda$, поэтому $\varphi(u\lambda) = x_1\lambda_1\lambda + \cdots + x_n\lambda_n\lambda = \varphi(u)\lambda)$.

Осталось доказать, что такое φ единственно. Пусть φ' — другое отображение с тем же свойством; тогда $\psi = \varphi - \varphi'$ — линейное отображение, и $\psi(u_i) = \varphi(u_i) - \varphi'(u_i) = x_i - x_i = 0$ для любого i. Но тогда для любого $u = u_1\lambda_1 + \cdots + u_n\lambda_n$ получаем $\varphi(u) - \varphi'(u) = \psi(u) = \psi(u_1)\lambda_1 + \cdots + \psi(u_n)\lambda_n = 0$; поэтому $\varphi(u) = \varphi'(u)$ для всех $u \in U$. □

Следствие 7.7.2. *Два векторных пространства над полем k изоморфны тогда и только тогда, когда их размерности равны.*

Доказательство. Очевидно, что размерности изоморфных пространств равны. Обратно, если U, V — векторные пространства над k такие, что $\dim(U) = \dim(V) = n$, пусть $u_1, \ldots, u_n$ — произвольный базис U, а $v_1, \ldots, v_n$ — произвольный базис V. Тогда по теореме 7.7.1 найдется линейное отображение $\varphi\colon U \to V$ такое, что $\varphi(u_i) = v_i$ для всех $i = 1, \ldots, n$. Покажем, что φ — изоморфизм. Если $u \in \operatorname{Ker}(\varphi)$, запишем $u = u_1\lambda_1 + \cdots + u_n\lambda_n$. Тогда $0 = \varphi(u) = v_1\lambda_1 + \cdots + v_n\lambda_n$, и в силу линейной независимости $v_1, \ldots, v_n$ из этого следует, что $\lambda_1 = \cdots = \lambda_n = 0$ и $u = 0$. Это показывает, что φ инъективно. Пусть теперь $v \in V$; запишем $v = v_1\lambda_1 + \cdots + v_n\lambda_n$ и рассмотрим вектор $u = u_1\lambda_1 + \cdots + \lambda_n$. Тогда $\varphi(u) = \varphi(u_1)\lambda_1 + \cdots + \varphi(u_n)\lambda_n = v_1\lambda_1 + \cdots + v_n\lambda_n = v$, и φ сюръективно. □

7.8 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 U, V — векторные пространства над k, $\varphi\colon U \to V$ — линейное отображение между ними. Если выбрать базисы в U и V, каждое из этих пространств отождествится с некоторым пространством столбцов над k. Сейчас мы увидим, что при этом линейное отображение φ можно записать как матрицу таким образом, что действие φ будет состоять в умножении слева на столбцы.

А именно, пусть $\mathcal{B} = \{e_1, \ldots, e_n\}$ — базис в U, а $\mathcal{B}' = \{f_1, \ldots, f_m\}$ — базис в V. Применим φ к каждому вектору базиса $\mathcal{B}$ и разложим результат по базису $\mathcal{B}'$. Мы получим n координатных столбцов $[e_1]_{\mathcal{B}'}, \ldots, [e_n]_{\mathcal{B}'}$. Матрица, составленная из этих столбцов, называется **матрицей линейного отображения φ в базисах $\mathcal{B}$, $\mathcal{B}'$**. Мы будем обозначать ее через $[\varphi]_{\mathcal{B},\mathcal{B}'}$:

$$[\varphi]_{\mathcal{B},\mathcal{B}'} = \begin{pmatrix} [e_1]_{\mathcal{B}'} & \ldots & [e_n]_{\mathcal{B}'} \end{pmatrix} \in M(m, n, k).$$

Следующая теорема утверждает, что если мы будем записывать элементов U как столбцы координат высоты n в базисе $\mathcal{B}$, а элементы V как столбцы координат высоты m в базисе $\mathcal{B}'$, то действие φ заключается в умножении на матрицу отображения φ в этих базисах.

Теорема 7.8.1. *Пусть U, V — векторные пространства над K, $\mathcal{B}$ — базис U, $\mathcal{B}'$ — базис V. Тогда*

$$[\varphi(u)]_{\mathcal{B}'} = [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}}$$

для всех $u \in U$.

Доказательство. Запишем $[u]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$. Это означает $u = e_1\lambda_1 + \cdots + e_n\lambda_n$. Тогда $\varphi(u) = \varphi(e_1)\lambda_1 + \cdots + \varphi(e_n)\lambda_n$. Значит,

$$[\varphi(u)]_{\mathcal{B}'} = [\varphi(e_1)]_{\mathcal{B}'}\lambda_1 + \cdots + [\varphi(e_n)]_{\mathcal{B}'}\lambda_n = \begin{pmatrix} [\varphi(e_1)]_{\mathcal{B}'} & \ldots & [\varphi(e_n)]_{\mathcal{B}'} \end{pmatrix} \cdot \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix} = [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}}.$$

□

Следствие 7.8.2. *Линейное отображение* $\varphi\colon U \to V$ *полностью определяется своей матрицей* $[\varphi]_{\mathcal{B},\mathcal{B}'}$. *Кроме того, если для некоторой матрицы* $A \in M(m, n, k)$ *выполнено равенство* $[\varphi(u)]_{\mathcal{B}'} = A \cdot [u]_{\mathcal{B}}$ *для всех* $u \in U$, *то* $A = [\varphi]_{\mathcal{B},\mathcal{B}'}$.

Доказательство. Теорема 7.8.1 утверждает, что (после выбора базисов $\mathcal{B}$ и $\mathcal{B}'$) если мы знаем u и матрицу $[\varphi]_{\mathcal{B},\mathcal{B}'}$, то мы знаем координаты $\varphi(u)$ в базисе $\mathcal{B}'$, а следовательно, и сам вектор $\varphi(u)$ для каждого $u \in U$.

Если $[\varphi(u)]_{\mathcal{B}'} = A \cdot [u]_{\mathcal{B}}$ и $[\varphi(u)]_{\mathcal{B}'} = [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}}$ для всех $u \in U$, то, в частности, эти равенства выполнено для векторов $u = e_1, \ldots, e_n$ из базиса $\mathcal{B}$. Но при подстановке вектора $u = e_i$ в эти два равенства получаем равенство i-го столбца матрицы A и i-го столбца матрицы $[\varphi]_{\mathcal{B},\mathcal{B}'}$. Варьируя i, получаем равенство указанных матриц. □

Таким образом, для задания линейного отображения достаточно задать его матрицу (в какой-нибудь паре базисов). Посмотрим, что происходит с матрицами линейных отображений при известных нам операциях. Ничего удивительного: при сложении линейных отображений их матрицы складываются, при умножении на скаляр — умножаются на скаляр. Кроме того, композиции линейных отображений соответствует произведение матриц.

Предложение 7.8.3. *Пусть* $\varphi, \psi\colon U \to V$ — *линейные отображения,* $\lambda \in k$, $\mathcal{B}$ — *базис в* U, $\mathcal{B}'$ — *базис в* V. *Тогда* $[\varphi + \psi]_{\mathcal{B},\mathcal{B}'} = [\varphi]_{\mathcal{B}} + [\psi]_{\mathcal{B}'}$ *и* $[\varphi\lambda]_{\mathcal{B},\mathcal{B}'} = [\varphi]_{\mathcal{B},\mathcal{B}'}\lambda$.

Доказательство. По определению суммы имеем $(\varphi + \psi)(u) = \varphi(u) + \psi(u)$ для всех $u \in U$. Теорема 7.8.1 говорит нам, что $[\varphi(u)]_{\mathcal{B}'} = [\varphi_{\mathcal{B},\mathcal{B}'}] \cdot [u]_{\mathcal{B}}$ и $[\psi(u)]_{\mathcal{B}'} = [\psi_{\mathcal{B},\mathcal{B}'}] \cdot [u]_{\mathcal{B}}$. Поэтому

$$\begin{aligned} [(\varphi + \psi)(u)]_{\mathcal{B}'} &= [\varphi(u) + \psi(u)]_{\mathcal{B}'} \\ &= [\varphi(u)]_{\mathcal{B}'} + [\psi(u)]_{\mathcal{B}'} \\ &= [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}} + \\ [\psi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}} &= ([\varphi]_{\mathcal{B},\mathcal{B}'} + [\psi]_{\mathcal{B},\mathcal{B}'}) \cdot [u]_{\mathcal{B}}. \end{aligned}$$

Полученное равенство выполняется при всех $u \in U$. По следствию 7.8.2 это означает, что $[\varphi + \psi]_{\mathcal{B},\mathcal{B}'} = [\varphi]_{\mathcal{B},\mathcal{B}'} + [\psi]_{\mathcal{B},\mathcal{B}'}$.

Аналогично,

$$[(\varphi\lambda)(u)]_{\mathcal{B}'} = [\varphi(u)\lambda]_{\mathcal{B}'} = [\varphi(u)]_{\mathcal{B}'}\lambda = [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}} \cdot \lambda = ([\varphi]_{\mathcal{B},\mathcal{B}'}\lambda) \cdot [u]_{\mathcal{B}}.$$

По следствию 7.8.2 это означает, что $[\varphi\lambda]_{\mathcal{B},\mathcal{B}'} = [\varphi]_{\mathcal{B},\mathcal{B}'}\lambda$. □

Предложение 7.8.3, таким образом, утверждает, что пространство $\mathrm{Hom}(U, V)$ изоморфно пространству матриц $M(n, m, k)$. Этот изоморфизм, однако, зависит от выбора базисов в U и V, то есть, от выбора изоморфизмов $U \cong k^n$ и $V \cong k^m$.

Предложение 7.8.4. *Пусть* $\varphi\colon U \to V$, $\psi\colon V \to W$ — *линейные отображения,* $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ — *базисы в* U, V, W *соответственно. Тогда*

$$[\psi \circ \varphi]_{\mathcal{B},\mathcal{B}''} = [\psi]_{\mathcal{B}',\mathcal{B}''} \cdot [\varphi]_{\mathcal{B},\mathcal{B}'}.$$

Доказательство. Пусть $u \in U$; тогда

$$[(\psi \circ \varphi)(u)]_{\mathcal{B}''} = [\psi(\varphi(u))]_{\mathcal{B}''} = [\psi]_{\mathcal{B}',\mathcal{B}''} \cdot [\varphi(u)]_{\mathcal{B}'} = [\psi]_{\mathcal{B}',\mathcal{B}''} \cdot [\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}}.$$

По следствию 7.8.2 получаем нужное равенство. □

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 7.8.5. *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

$$[\varphi(u)_{\mathcal{C}'} = (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi(u)]_{\mathcal{B}'} = (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi]_{\mathcal{B},\mathcal{B}'} \cdot [u]_{\mathcal{B}} = (\mathcal{C}' \rightsquigarrow \mathcal{B}')[\varphi]_{\mathcal{B},\mathcal{B}'} \cdot (\mathcal{B} \rightsquigarrow \mathcal{C})[u]_{\mathcal{C}}.$$

По следствию 7.8.2 (с учетом того, что $(\mathcal{B}' \rightsquigarrow \mathcal{C}')^{-1} = (\mathcal{C}' \rightsquigarrow \mathcal{B}')$) получаем нужное равенство. □

Итак, при замене базисов в пространствах U и V матрица отображения $\varphi\colon U \to V$ домножается слева на матрицу замены базиса в U и справа — 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 7.8.6 (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 (7.3.3) 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 7.5.2 мы получим базис в U; пусть это $e_1, \ldots, e_n$, причем $e_1, \ldots, e_r$ — относительный базис U над $\operatorname{Ker}(\varphi)$, а $e_{r+1}, \ldots, e_n$ — базис $\operatorname{Ker}(\varphi)$. Базису $\overline{e_1}, \ldots, \overline{e_r}$ в $U/\operatorname{Ker}(\varphi)$ в силу изоморфизма $\widetilde{\varphi}$ соответствует базис $f_1, \ldots, f_r$ в $\operatorname{Im}(\varphi)$; при этом $\varphi(e_i) = f_i$ для $i = 1, \ldots, r$, и видно, что $r = \dim(\operatorname{Im}(\varphi))$. Наконец, поскольку $\operatorname{Im}(\varphi) \leqslant V$, можно дополнить систему $f_1, \ldots, f_r$ до базиса V векторами $f_{r+1}, \ldots, f_m$. Поскольку $\varphi(e_i) = f_i$ для $i = 1, \ldots, r$ и $\varphi(e_i) = 0$ для $i \geqslant r+1$, матрица φ в базисах $(e_1, \ldots, e_n)$, $(f_1, \ldots, f_m)$ имеет нужный вид. □

Фактически мы получили еще одно доказательство следствия 5.4.2.

Замечание 7.8.7. Размерность образа отображения φ называется **рангом** φ; по теореме 7.8.6 ранг линейного отображения равен рангу его матрицы (в любой паре базисов, поскольку при домножении на обратимые матрицы ранг не меняется).

7.9 Приложения к системам линейных уравнений

Литература: [F], гл. IV, § 4, пп. 1–4; [K1], гл. 2, § 2, пп. 1–3; [vdW], гл. IV, § 22.

Проиллюстрируем полученные результаты на примере систем линейных уравнений. Пусть $AX = B$ — система линейных уравнений, где $A \in M(m, n, k)$, $B \in k^m$. Рассмотрим линейное отображение $\varphi\colon k^n \to k^m$, $X \mapsto AX$. Его матрица в стандартных базисах пространств столбцов равна A. Ядро этого отображения — множество решений однородной системы $AX = 0$; образ этого отображения — множество столбцов B, для которых система $AX = B$ имеет решение.

Теорема 7.9.1. *Пусть* $A \in M(m, n, k)$. *Размерность пространства решений однородной линейной системы уравнений* $AX = 0$ *равна* $n - \operatorname{rk}(A)$.

Доказательство. По замечанию 7.8.7 ранг $\operatorname{rk}(A)$ матрицы A равен рангу отображения φ, то есть, $\dim(\operatorname{Im}(\varphi))$. С другой стороны, размерность пространства решений однородной системы — это $\dim(\operatorname{Ker}(\varphi))$. По следствию 7.5.4 их сумма равна n. □

Следствие 7.9.2. *Пусть* $A \in M(m, n, k)$. *Однородная линейная система уравнений* $AX = 0$ *имеет нетривиальное (то есть, ненулевое) решение тогда и только тогда, когда* $\operatorname{rk}(A) < n$. *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 6.5.8 равносильно необратимости A. □

Теорема 7.9.3 (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 $\varphi\colon k^n \to k^m$, $x \mapsto Ax$. Образ φ — это подпространство, порожденное векторами $\varphi(e_1), \ldots, \varphi(e_n)$, то есть, пространство столбцов матрицы A. Значит, B лежит в $\mathrm{Im}(\varphi)$ тогда и только тогда, когда столбец φ является линейной комбинацией столбцов матрицы A. По предложению 5.8.6 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 $\mathrm{Ker}(\varphi) = 0$, то есть, когда $\mathrm{rk}(A) = \dim(\mathrm{Im}(\varphi)) = n$. □

8 Жорданова нормальная форма

В прошлой главе мы выяснили, что для линейного отображения $\varphi\colon U \to V$ можно выбрать базисы в U и в V так, что матрица φ в этих базисах будет окаймленной единичной. Пусть теперь $a\colon V \to V$ — 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** $a\colon V \to V$ в базисе $\mathcal{B}$ называется матрица отображения a в базисах $\mathcal{B}$, $\mathcal{B}$. Мы будем обозначать ее через $[a]_{\mathcal{B}}$ вместо $[a]_{\mathcal{B},\mathcal{B}}$. Цель настоящей главы — выяснить, к какому наиболее простому виду можно привести матрицу оператора a с помощью выбора базиса в V. По теореме 7.8.5 при замене базиса $\mathcal{B}$ на $\mathcal{B}'$ матрица оператора a домножается справа на матрицу замены базиса и слева на обратную к ней. Таким образом, если $A = [a]_{\mathcal{B}}$, $A' = [a]_{\mathcal{B}'}$, T — матрица перехода от $\mathcal{B}$ к $\mathcal{B}'$, то $A' = T^{-1}AT$. Эта процедура называется **сопряжением**: говорят, что $T^{-1}AT$ — матрица, **сопряженная** к матрице A при помощи T.

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?

8.1 Собственные значения и собственные векторы

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.

Определение 8.1.1. Пусть a — оператор на V. Элемент $\lambda \in k$ называется **собственным значением** (или **собственным числом**), если для некоторого ненулевого вектора $v \in V$ выполнено $a(v) = \lambda v$. Такой вектор v при этом называется **собственным вектором**, соответствующим собственному числу λ.

Определение 8.1.2. Пусть a — оператор на V, $\mathcal{B}$ — некоторый базис в V, $\dim(V) = n$. Пусть A — матрица оператора a в базисе $\mathcal{B}$: $A = [a]_{\mathcal{B}}$. Рассмотрим матрицу $A - tE_n \in M(n, k[t])$. Определитель этой матрицы $p_A(t)$ лежит в $k[t]$ и называется **характеристическим многочленом** матрицы A. Также он называется **характеристическим многочленом** оператора a и обозначается через $p_a(t)$.

Замечание 8.1.3. Заметим, что если $\mathcal{B}'$ — другой базис V, и $C = (\mathcal{B} \rightsquigarrow \mathcal{B}')$ — матрица перехода, то $[a]_{\mathcal{B}'} = C^{-1}[a]_{\mathcal{B}}C$, поэтому

$$\begin{aligned}\det([a]_{\mathcal{B}'} - tE_n) &= \det(C^{-1}[a]_{\mathcal{B}}C - tC^{-1}C) \\ &= \det(C^{-1}([a]_{\mathcal{B}} - tE_n)C) \\ &= \det(C^{-1})\det([a]_{\mathcal{B}} - tE_n)\det(C) \\ &= \det([a]_{\mathcal{B}} - tE_n).\end{aligned}$$

Это означает, что характеристический многочлен оператора корректно определен: он не зависит от выбора базиса в V. Нетрудно видеть, что степень характеристического многочлена равна $n = \dim(V)$, а его старший коэффициент равен $(-1)^n$.

Теорема 8.1.4. *Пусть a — оператор на V. Элемент $\lambda \in k$ является собственным числом оператора a тогда и только тогда, когда λ — корень характеристического многочлена оператора a.*

Доказательство. Пусть $\mathcal{B}$ — некоторый базис в V. Равенство $a(v) = \lambda v$ равносильно равенству $[a]_{\mathcal{B}}[v]_{\mathcal{B}} = \lambda[v]_{\mathcal{B}}$, что равносильно равенству $([a]_{\mathcal{B}} - \lambda E_n)[v]_{\mathcal{B}} = 0$. Это однородная система линейных уравнений с матрицей $[a]_{\mathcal{B}} - \lambda E_n$; по теореме Кронекера–Капелли (7.9.3) она имеет нетривиальное решение тогда и только тогда, когда ее ранг меньше n, что по следствию 6.5.8 и следствию 5.8.3 равносильно тому, что ее определитель $p_a(\lambda)$ равен нулю. □

Предложение 8.1.5. *Пусть λ — собственное число оператора $a\colon V \to V$. Множество всех собственных векторов a, соответствующих λ, вместе с $0 \in V$ образует подпространство в V.*

Доказательство. По определению множество собственных векторов — это множество ненулевых решений уравнения $a(v) = \lambda v$, а множество собственных векторов вместе с 0 — это множество всех решений этого уравнения. Перепишем его следующим равносильным образом: $(a - \lambda \mathrm{id}_V)(v) = 0$. Множество решений этого уравнения — это просто ядро оператора $a - \lambda \mathrm{id}$. По предложению 7.3.1 оно является подпространством в V. □

Определение 8.1.6. Подпространство из предложения 8.1.5 обозначается через V_λ и называется **собственным подпространством**, соответствующим собственному числу λ.

8.2 Диагонализуемые операторы

Литература: [К2], гл. 2, § 3, п. 4; [КМ], ч. 1, § 8.

Определение 8.2.1. Оператор $a\colon V \to V$ называется **диагонализуемым**, если существует базис $\mathcal{B}$ пространства V, в котором матрица оператора a диагональна.

Предложение 8.2.2. *Оператор $a\colon V \to V$ диагонализуем тогда и только тогда, когда у V имеется базис, состоящий из собственных векторов оператора a.*

Доказательство. Если для некоторого базиса $\mathcal{B} = (v_1, \ldots, v_n)$ пространства V выполнено

$$[a]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix},$$

то (по определению матрицы линейного оператора) $a(v_1) = \lambda_1 v_1, \ldots, a(v_n) = \lambda_n v_n$, то есть, каждый вектор базиса $\mathcal{B}$ — собственный. Обратно, если $\mathcal{B} = (v_1, \ldots, v_n)$ — базис V и каждый вектор v_i собственный, например, $a(v_i) = \lambda_i v_i$, то в базисе $\mathcal{B}$ матрица оператора a имеет диагональный вид, в котором на диагонали стоят $\lambda_1, \ldots, \lambda_n$. □

Определение 8.2.3. Пусть $\lambda \in k$ — собственное число оператора $a\colon V \to V$. **Алгебраической кратностью** λ называется его кратность как корня характеристического многочлена $p_a(t)$. **Геометрической кратностью** λ называется размерность собственного подпространства V_λ.

Теорема 8.2.4. *Геометрическая кратность собственного числа λ оператора a не превосходит его алгебраической кратности.*

Доказательство. Пусть k — геометрическая кратность λ, то есть, $\dim(V_\lambda) = k$. Выберем некоторый базис $e_1, \ldots, e_k$ в V_λ и дополним его до базиса $\mathcal{B} = e_1, \ldots, e_n$ всего пространства V. При этом $a(e_i) = \lambda e_i$ для $i = 1, \ldots, k$, поэтому матрица оператора a в базисе $\mathcal{B}$ имеет вид

$$[a]_{\mathcal{B}} = \begin{pmatrix} \lambda & 0 & \ldots & 0 & * & \ldots & * \\ 0 & \lambda & \ldots & 0 & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \lambda & * & \ldots & * \\ 0 & 0 & \ldots & 0 & * & \ldots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 0 & * & \ldots & * \end{pmatrix}.$$

После вычитания tE_n получаем

$$[a]_{\mathcal{B}} - tE_n = \begin{pmatrix} \lambda - t & 0 & \dots & 0 & * & \dots & * \\ 0 & \lambda - t & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda - t & * & \dots & * \\ 0 & 0 & \dots & 0 & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & * & \dots & * \end{pmatrix}.$$

Это блочная верхнетреугольная матрица, поэтому ее определитель равен произведению определителя диагональной матрицы $k \times k$ с числами $\lambda - t$ на диагонали и некоторой матрицы T размера $(n-k) \times (n-k)$. Таким образом, $p_a(t) = (\lambda - t)^k \det(T)$. Кратность λ как корня $p_a(t)$ равна сумме k и кратности λ как корня многочлена $\det(T)$, поэтому она не меньше k. □

Пример 8.2.5. Рассмотрим оператор $a\colon k^2 \to k^2$, задаваемый матрицей $A = \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix}$. Его характеристический многочлен равен $(t-\lambda)^2$, поэтому λ — собственное число a с алгебраической кратностью 2. С другой стороны, решения однородной системы линейных уравнений $(A - \lambda E_n)X = 0$ имеют вид $\begin{pmatrix} x \\ 0 \end{pmatrix}$, $x \in k$, так что геометрическая кратность λ равна 1. По предложению 8.2.2 оператор a не является диагонализуемым: все его собственные векторы лежат в одномерном собственном пространстве V_λ и не могут образовывать базис V. Чуть позже мы увидим (теорема 8.2.9), что у диагонализуемого оператора алгебраическая кратность любого собственного числа совпадает с его геометрической кратностью.

Нам понадобится распространить определение прямой суммы на случай нескольких подпространств.

Определение 8.2.6. Пусть V — векторное пространство над полем k, $V_1, \dots, V_m \leqslant V$. Будем говорить, что V является **прямой суммой**, если любой вектор $v \in V$ единственным образом представляется в виде $v = v_1 + \dots + v_m$ для $v_i \in V_i$. Обозначение: $V = V_1 \oplus \dots \oplus V_m$.

Предложение 8.2.7. *Пусть V — векторное пространство над k, $V_1, \dots, V_m$. Пространство V является прямой суммой $V_1, \dots, V_m$ тогда и только тогда, когда выполняются следующие два условия:*

1. *V является суммой $V_1, \dots, V_m$: $V = \sum_{i=1}^m V_i$;*
2. *для любого $i = 1, \dots, n$ пересечение подпространства V_i с суммой остальных тривиально, то есть, $V_i \cap \left(\sum_{j \neq i} V_j\right) = \{0\}$.*

Доказательство. Если $V = V_1 \oplus \dots \oplus V_m$, то любой вектор $v \in V$ представляется в виде суммы векторов из V_i, поэтому $V = \sum_{i=1}^m V_i$. Если же $v \in V_i \cap \left(\sum_{j \neq i} V_j\right)$, то, с одной стороны,

$v \in V_i$, с другой стороны, $v = \sum_{j \neq i} v_j$, где $v_j \in V_j$. Таким образом, $v = 0 + \cdots + v + \cdots + 0 = v_1 + \cdots + 0 + \cdots + v_m$ — два представления v, которые должны совпадать по определению прямой суммы, откуда $v = 0$.

Обратно, из выполнения условия (1) следует, что любой вектор v имеет представление в виде $v = v_1 + \cdots + v_m$ для $v_i \in V_i$. Покажем, что такое представление единственно. Пусть $v = v_1' + \cdots + v_m'$ — другое такое представление, и $v_i \neq v_i'$ для некоторого i. Вычитая эти выражения, получаем $0 = (v_1 - v_1') + \cdots + (v_m - v_m')$. Перенесем $v_i - v_i'$ в левую часть: $v_i' - v_i = (v_1 - v_1') + \cdots + (v_m - v_m')$. С одной стороны, $v_i' - v_i \in V_i$; с другой стороны, правая часть является суммой векторов вида $v_j - v_j'$ по $j \neq i$, то есть, лежит в $\sum_{j \neq i} V_j$. По предположению это возможно только если $v_i' - v_i = 0$, то есть, $v_i' = v_i$ — противоречие. □

Предложение 8.2.8. *Собственные векторы, соответствующие различным собственным числам оператора* $a\colon V \to V$*, линейно независимы. Сумма* $\sum_\lambda V_\lambda$ *(где* λ *пробегает все собственные числа оператора* a*) является прямой.*

Доказательство. Пусть $\lambda_1, \ldots, \lambda_m$ — набор попарно различных собственных чисел. Предположим, что $v_1, \ldots, v_m \in V$ — ненулевые векторы такие, что $a(v_i) = v_i\lambda_i$ для всех i. Предположим, что между ними имеется нетривиальная линейная комбинация $v_1c_1 + v_2c_2 + \cdots + v_mc_m = 0$, и выберем такую комбинацию, в которой количество ненулевых коэффициентов минимально. Можно считать, что $c_1 \neq 0$. При этом среди $c_1, \ldots, c_m$ есть хотя бы два ненулевых коэффициента: иначе $v_1c_1 = 0$, откуда $v_1 = 0$. Применим оператор a: $0 = a(v_1c_1 + v_2c_2 + \cdots + v_mc_m) = a(v_1)c_1 + a(v_2)c_2 + \cdots + a(v_m)c_m = v_1\lambda_1c_1 + v_2\lambda_2c_2 + \cdots + v_m\lambda_mc_m$. Умножим исходное равенство на λ_1:

$$v_1c_1\lambda_1 + v_2c_2\lambda_1 + \cdots + v_mc_m\lambda_1 = 0.$$

Вычитая два полученных выражения, получаем

$$v_2(\lambda_2 - \lambda_1)c_2 + \cdots + v_m(\lambda_m - \lambda_1)c_m = 0.$$

Эта линейная комбинация нетривиальна, поскольку среди $c_2, \ldots, c_m$ есть хотя бы один ненулевой, и все множители вида $(\lambda_i - \lambda_1)$ отличны от 0 в силу попарной различности $\lambda_1, \ldots, \lambda_m$. С другой стороны, количество ненулевых коэффициентов в этой линейной комбинации меньше, чем в исходной, поскольку пропал коэффициент c_1. Получаем противоречие. □

Теорема 8.2.9. *Линейный оператор* $a\colon V \to V$ *диагонализуем тогда и только тогда, когда все корни характеристического многочлена лежат в* k *и геометрическая кратность каждого собственного значения* a *совпадает с алгебраической кратностью.*

Доказательство. Пусть оператор a диагонализуем. Перестановкой векторов базиса можно добиться, чтобы одинаковые значения на диагонали стояли рядом. Поэтому можно считать, что матрица оператора a диагональна, и на диагонали стоят элементы

$$\lambda_1, \ldots, \lambda_1, \lambda_2, \ldots, \lambda_2, \ldots, \lambda_m, \ldots, \lambda_m,$$

где все λ_i попарно различны и λ_i встречается n_i раз. При этом, очевидно, $n_1 + n_2 + \cdots + n_m = n$. Характеристический многочлен a, очевидно, равен $\prod_{i=1}^{m} (\lambda_i - t)^{n_i}$, поэтому все его

корни лежат в k. Алгебраическая кратность собственного значения λ_i равна n_i. Наконец, векторы базиса, соответствующие столбцам, в которых стоит λ_i, являются собственными векторами, соответствующими собственному значению λ_i, поэтому геометрическая кратность λ_i не меньше n_i. По теореме 8.2.4 из этого следует, что она равна n_i.

Обратно, пусть $\lambda_1, \ldots, \lambda_m \in k$ — все корни характеристического многочлена a; пусть λ_i имеет алгебраическую (и геометрическую) кратность n_i. Пусть $V_i = V_{\lambda_i}$ — собственное подпространство, соответствующее собственному числу λ_i. Тогда $\dim(V_i) = n_i$. По предложению 8.2.8 сумма подпространств V_i является прямой: $V_1 \oplus \cdots \oplus V_m \leqslant V$. С другой стороны, ее размерность равна сумме размерностей V_i, то есть, $n_1 + \cdots + n_m = n$. Поэтому, на самом деле, $V_1 \oplus \cdots \oplus V_m = V$. Выберем в каждом из V_i по базису. Объединение этих базисов тогда будет базисом V, состоящим из собственных векторов. По предложению 8.2.2 из этого следует, что a диагонализуем. □

8.3 Инвариантные подпространства

Литература: [F], гл. XII, § 5, п. 3; [K2], гл. 2, § 3, п. 2; [KM], ч. 1, § 8; [vdW], гл. XII, § 87.

Определение 8.3.1. Пусть $a\colon V \to V$ — линейный оператор на V. Заметим, что он действует не только на векторах V, но и на подпространствах V. Действительно, для $U \leqslant V$ положим $a(U) = \{a(u) \mid u \in U\}$. Подпространство $U \leqslant V$ называется **инвариантным** относительно оператора a, если $a(U) \subseteq U$.

Замечание 8.3.2. Несложно видеть, что одномерное инвариантное подпространство — это в точности линейная оболочка одного собственного вектора.

Определение 8.3.3. Пусть $a\colon V \to V$ — линейный оператор, и подпространство $U \leqslant V$ инвариантно относительно a. Оператор $a|_U\colon U \to U$, $u \mapsto a(u)$ будем называть **ограничением** оператора a на инвариантное подпространство U.

Наличие инвариантного подпространства позволяет выбрать базис, в котором матрица оператора a выглядит чуть проще.

Предложение 8.3.4. *Пусть $a\colon V \to V$ — линейный оператор. В пространстве V имеется инвариантное подпространство тогда и только тогда, когда матрица оператора a в некотором базисе принимает клеточно-верхнетреугольный вид*

$$\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}.$$

Пространство V является прямой суммой двух инвариантных подпространств U и U' тогда и только тогда, когда матрица оператора a в некотором базисе принимает клеточно-диагональный вид

$$\begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}.$$

При этом диагональные блоки — это матрицы ограничения оператора a на U и U'.

Доказательство. Пусть $[a]_{\mathcal{B}} = \begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$, где $\mathcal{B} = (e_1, \ldots, e_n)$ —- некоторый базис, и диагональные блоки указанной матрицы имеют размер $m \times m$ и $(n-m) \times (n-m)$. Тогда подпространство $\langle e_1, \ldots, e_m \rangle$ является инвариантным. Если же $[a]_{\mathcal{B}} = \begin{pmatrix} * & 0 \\ 0 & * \end{pmatrix}$, то и подпространство $\langle e_{m+1}, \ldots, e_n \rangle$ является инвариантным.

Обратно, если $U \leqslant V$ — инвариантное подпространство, выберем базис $e_1, \ldots, e_m$ в U и дополним его до базиса $(e_1, \ldots, e_n)$ всего пространства V; нетрудно видеть, что тогда матрица a в этом базисе имеет вид $\begin{pmatrix} * & * \\ 0 & * \end{pmatrix}$. Если же $V = U \oplus U'$ для инвариантных подпространств U, U', выберем базисы $\mathcal{B} = (e_1, \ldots, e_m)$ в U и $\mathcal{B}' = (e_{m+1}, \ldots, e_n)$ в U'. Тогда нетрудно видеть, что $e_1, \ldots, e_m, e_{m+1}, \ldots, e_n$ — базис V, в котором матрица a имеет клеточно-диагональный вид $\begin{pmatrix} [a]_{\mathcal{B}} & 0 \\ 0 & [a]_{\mathcal{B}'} \end{pmatrix}$. □

8.4 Многочлены от операторов и теорема Кэли–Гамильтона

Литература: [F], гл. XII, § 5, пп. 2–6; гл. IV, § 7, п. 2; [K2], гл. 2, § 4, п. 1; [KM], ч. 1, § 8.

Пусть $f(t) = c_0 + c_1 t + c_2 t^2 + \cdots + c_d x^d \in k[t]$ — многочлен от одной переменной, $A \in M(n, k)$. Определим матрицу $f(A)$, называемую результатом подстановки A в многочлен f, равенством $f(A) = c_0 E_n + c_1 A + c_2 A^2 + \cdots + c_d A^d \in M(n, k)$.

Пусть теперь $a\colon V \to V$ — линейный оператор. Определим оператор $f(a)$ равенством $f(a) = \mathrm{id}_V c_0 + a c_1 + a^2 c_2 + \cdots + a^d c_d$, где id_V — тождественный оператор на V, а $a^j = a \circ \cdots \circ a$ — j-кратная композиция оператора a.

Следующая лемма очевидным образом следует из определения операций над многочленами.

Лемма 8.4.1. *Пусть* $f, g \in k[t]$, $c \in k$, $A \in M(n, k)$. *Тогда* $f(A) + g(A) = (f+g)(A)$, $f(A)g(A) = (fg)(A)$, $c \cdot f(A) = (cf)(A)$.

Следствие 8.4.2. *Пусть* $f, g \in k[t]$, $A \in M(n, k)$. *Тогда матрицы* $f(A)$ *и* $g(A)$ *коммутируют:* $f(A)g(A) = g(A)f(A)$.

Доказательство. $f(A)g(A) = (fg)(A) = (gf)(A) = g(A)f(A)$. □

Определения подстановки матрицы и оператора в многочлен согласованы:

Лемма 8.4.3. *Пусть* $a\colon V \to V$ — *линейный оператор,* $f \in k[t]$, $\mathcal{B}$ — *некоторый базис пространства* V. *Тогда* $[f(a)]_{\mathcal{B}} = f([a]_{\mathcal{B}})$.

Доказательство. Следует из того, что операции над операторами соответствуют тем же операциям над их матрицами (см. предложения 7.8.3 и 7.8.4). □

Теорема 8.4.4 (Кэли–Гамильтона). *Если* $a\colon V \to V$ *— линейный оператор,* $p_a \in k[t]$ *— его характеристический многочлен, то* $p_a(a) = 0$*. Эквивалентно, для матрицы* $A \in M(n,k)$ *выполнено* $p_A(A) = 0$*.*

Доказательство. Утверждение для операторов следует из утверждения для матриц в силу леммы 8.4.3. Рассмотрим матрицу $A - tE_n$. По следствию 5.8.2 в кольце $M(n,k[t])$ выполнено равенство $\operatorname{adj}(A - tE_n) \cdot (A - tE_n) = \det(A - tE_n) \cdot E_n$. Заметим, что $M(n,k[t]) = (M(n,k))[t]$. Иными словами, матрицу с многочленами t в качестве коэффициентов можно расписать как многочлен от t с матрицами в качестве коэффициентов. Так, правая часть нашего равенства, $\det(A - tE_n) \cdot E_n$ — это диагональная матрица, в которой все диагональные элементы равны $p_A(t) = \det(A - tE_n)$. Запишем $p_A(t) = c_0 + c_1 t + c_2 t^2 + \cdots + (-1)^n t^n$. Тогда

$$\det(A - tE_n) \cdot E_n = (c_0 E_n) + (c_1 E_n)t + (c_2 E_n)t^2 + \cdots + ((-1)^n E_n)t^n.$$

Вообще, диагональная матрица, в которой на диагонали стоят одинаковые элементы, называется скалярной. Таким образом, все коэффициенты при t в правой части оказались скалярными матрицами из $M(n,k)$.

Теперь распишем левую часть равенства: пусть $\operatorname{adj}(A - tE_n) = d_0 + d_1 t + d_2 t^2 + \cdots + d_{n-1} t^{n-1}$ (нетрудно видеть, что матрица $\operatorname{adj}(A - tE_n)$ составлена из миноров размера $(n-1)$ матрицы $(A - tE_n)$, поэтому степень каждого многочлена, стоящего в ней, не превосходит $n-1$) для некоторых матриц $d_1, \ldots, d_n \in M(n,k)$. Тогда левая часть равна

$$(d_0 + d_1 t + d_2 t^2 + \cdots + d_{n-1} t^{n-1})(A - tE_n).$$

Мы хотим подставить в полученное равенство многочленов над $M(n,k)$ матрицу A вместо переменной t. Здесь нужно действовать аккуратно, поскольку коэффициенты наших многочленов лежат в некоммутативном кольце, а операция перемножения многочленов над некоммутативным кольцом не обладает многими хорошими свойствами. Связано это с тем, что при перемножении двух мономов вида $(\alpha_i t^i)(\beta_j t^j)$ мы хотим получить $(\alpha_i \beta_j) t^{i+j}$; однако, при подстановке элемента A вместо t может оказаться, что $(\alpha_i A^i)(\beta_j A^j) \neq \alpha_i \beta_j A^{i+j}$. К тому же, не совсем понятно, должны ли мы получить $(\alpha_i \beta_j) t^{i+j}$ или $(\beta_j \alpha_i) t^{i+j}$. Для того, чтобы проблем не возникло, хочется, чтобы все коэффициенты многочленов в нашем равенстве коммутировали друг с другом и с матрицей A, которую мы в него подставляем.

В правой части нашего равенства коэффициенты являются скалярными матрицами, то есть, матрицами вида cE_n для $c \in k$; такие матрицы коммутируют с любыми матрицами из $M(n,k)$: $(cE_n)B = cB = B(cE_n)$. В левой части имеется множитель $(A - tE_n)$, коэффициенты которого — единичная матрица и матрица A. Очевидно, они коммутируют с матрицей A. Наконец, покажем, что коэффициенты d_j в оставшемся множителе являются многочленами от A, и поэтому коммутируют с A и друг с другом (следствие 8.4.2). Действительно, посмотрим на наше равенство

$$(d_0 + d_1 t + d_2 t^2 + \cdots + d_{n-1} t^{n-1})(A - tE_n) = (c_0 E_n) + (c_1 E_n)t + (c_2 E_n)t^2 + \cdots + ((-1)^n E_n)t^n.$$

Приравняем в нем коэффициенты при всех степенях t, начиная со старшей. Мы получим цепочку равенств

$$\begin{aligned} -d_{n-1} &= (-1)^n; \\ -d_{n-2} + d_{n-1}A &= c_{n-1}E_n; \\ &\vdots \\ -d_{i-1} + d_iA &= c_iE_n; \\ &\vdots \\ d_0A &= c_0E_n. \end{aligned}$$

Из первого равенства получаем $d_{n-1} = (-1)^{n-1}$. Подставляя это во второе равенство, получаем $d_{n-2} = (-1)^{n-1}A - c_{n-1}E_n$, что является многочленом от A. Докажем по индукции, что все d_i являются многочленом от A. Действительно, если для $d_{n-1}, d_{n-2}, \ldots, d_i$ мы уже это проверили, заметим, что $d_{i-1} = c_iE_n - d_iA$. По предположению индукции d_i является многочленом от A; при умножении его на A и добавлении свободного члена c_iE_n получаем снова многочлен от A.

Итак, в наше равенство можно подставить A вместо переменной t. Но тогда в левой части возникнет множитель $(A - AE_n)$, равный нулю, а в правой части стоит $p_A(A)E_n = p_A(A)$. Теорема доказана. □

Определение 8.4.5. Пусть $a\colon V \to V$ — линейный оператор. Многочлен $f \in k[t]$ называется **аннулирующим многочленом вектора** $v \in V$, если $f(a)(v) = 0$, то есть, v лежит в ядре оператора $f(a)$. Многочлен $f \in k[t]$ называется **аннулирующим многочленом оператора** a, если он является аннулирующим многочленом каждого вектора из V: $f(a)(v) = 0$ для всех $v \in V$. Эквивалентно, это означает, что оператор $f(a)$ нулевой.

Замечание 8.4.6. По теореме Кэли-Гамильтона 8.4.4 характеристический многочлен p_a является нетривиальным аннулирующим многочленом оператора a (а значит, и каждого вектора $v \in V$). Это оправдывает следующее определение.

Определение 8.4.7. Многочлен $m_a \in k[t]$ называется **минимальным многочленом оператора** a, если он является аннулирующим многочленом оператора a минимальной степени, и имеет старший коэффициент 1. Аналогично, многочлен $m_{a,v} \in k[t]$ называется **минимальным многочленом вектора** $v \in V$ (относительно оператора a), если он является аннулирующим многочленом вектора v минимальной степени, и имеет старший коэффициент 1.

Замечание 8.4.8. Заметим, что минимальные многочлены существуют: в силу предыдущего замечания найдется (ненулевой) аннулирующий многочлен оператора a минимальной степени; после деления на старший коэффициент можно считать, что старший коэффициент его равен 1. Аналогично, для любого $v \in V$ найдется аннулирующий многочлен вектора v минимальной степени.

Теорема 8.4.9. *Любой аннулирующий многочлен оператора* a *делится на* m_a. *Любой аннулирующий многочлен вектора* v *делится на* $m_{a,v}$.

Доказательство. Пусть $f \in k[t]$ таков, что $f(a) = 0$. Поделим его с остатком на m_a: $f = m_a q + r$. Если остаток $r \neq 0$, то это нетривиальный многочлен степени строго меньше, чем $\deg m_a$. Кроме того, $f(a) = m_a(a)q(a) + r(a)$. Но $f(a) = 0$ и $m_a(a) = 0$, поэтому и $r(a) = 0$. Получаем, что r — аннулирующий многочлен оператора a, что противоречит минимальности m_a. Поэтому $r = 0$, то есть, f делится на m_a.

Аналогично, пусть $f \in k[t]$ таков, что $f(a)(v) = 0$. Поделим с остатком f на $m_{a,v}$: $f = m_{a,v}q + r$. Если $r \neq 0$, то это нетривиальный многочлен степени строго меньше, чем $\deg m_{a,v}$. Кроме того, $f(a)(v) = (m_a(a)q(a))(v) + r(a)(v)$. Но $f(a)(v) = 0$ и $m_a(a)(v) = 0$, откуда $r(a)(v) = 0$. Значит, r аннулирует v, что противоречит минимальности $m_{a,v}$. Поэтому $r = 0$, то есть, f делится на $m_{a,v}$. □

Пример 8.4.10. Пусть $a = \mathrm{id}_V \colon V \to V$ — тождественный оператор. Нетрудно видеть, что его характеристический многочлен равен $p_a(t) = (1-t)^n$, а минимальный многочлен равен $m_a(t) = 1 - t$.

8.5 Корневое разложение

Литература: [F], гл. XII, § 6, п. 2; [K2], гл. 2, § 4, п. 3; [KM], ч. 1, § 9.

Мы видели, что у диагонализуемого оператора найдется базис, состоящий из собственных векторов. В случае произвольного оператора собственных векторов оказывается недостаточно. Нам понадобится небольшое обобщение понятия собственного вектора.

Определение 8.5.1. Пусть $a \colon V \to V$ — линейный оператор, $\lambda \in k$ — его собственное число. Вектор $v \in V$ называется **корневым вектором**, соответствующим собственному числу λ, если $(a - \lambda \mathrm{id})^m(v) = 0$ для некоторого натурального m. Наименьшее такое m называется **высотой** корневого вектора v.

Замечание 8.5.2. Собственные векторы — это в точности корневые векторы высоты 1.

Предложение 8.5.3. *Пусть* $a \colon V \to V$ — *линейный оператор,* $\lambda \in k$ — *его собственное число. Все корневые векторы, соответствующие* λ, *образуют подпространство в* V.

Доказательство. По определению корневые векторы, соответствующие λ — это те, которые лежат в $\mathrm{Ker}((a - \lambda \mathrm{id})^m)$ для некоторого m. Рассмотрим подпространства $\mathrm{Ker}(a - \lambda \mathrm{id})$, $\mathrm{Ker}((a - \lambda \mathrm{id})^2)$, $\mathrm{Ker}((a - \lambda \mathrm{id})^3)$, ... Нетрудно видеть, что это возрастающая цепочка вложенных подпространств, поскольку из $(a - \lambda \mathrm{id})^i(v) = 0$ следует $(a - \lambda \mathrm{id})^{i+1}(v) = 0$:

$$\mathrm{Ker}(a - \lambda \mathrm{id}) \subseteq \mathrm{Ker}((a - \lambda \mathrm{id})^2) \subseteq \mathrm{Ker}((a - \lambda \mathrm{id})^3) \subseteq \cdots \leqslant V.$$

Посмотрим на их размерности: это неубывающая последовательность натуральных чисел, и все они ограничены размерностью V. Такая последовательность не может возрастать бесконечно, поэтому она должна стабилизироваться. Из этого следует, что, начиная с некоторого

m, выполнено

$$\operatorname{Ker}((a-\lambda\operatorname{id})^m)=\operatorname{Ker}((a-\lambda\operatorname{id})^{m+1})=\operatorname{Ker}((a-\lambda\operatorname{id})^{m+2})=\dots.$$

Но тогда множество корневых векторов, соответствующих λ, это в точности $\operatorname{Ker}((a-\lambda\operatorname{id})^m)$, а это подпространство в V. □

Определение 8.5.4. Подпространство корневых векторов, соответствующих собственному числу $\lambda\in k$, называется **корневым подпространством**, соответствующим λ, и обозначается через $V(\lambda)$.

Лемма 8.5.5. *Пусть* $f,g\in k[t]$, $a\colon V\to V$ — *линейный оператор. Тогда* $\operatorname{Ker}(f(a))$ *является инвариантным подпространством для* $g(a)$.

Доказательство. Пусть $v\in\operatorname{Ker}(f(a))$. Мы хотим показать, что $g(a)(v)\in\operatorname{Ker}(f(a))$. Действительно, $f(a)(g(a)(v))=(f(a)g(a))(v)=(g(a)f(a))(v)=g(a)(f(a)(v))=g(a)(0)=0$. □

Лемма 8.5.6. *Пусть* $f,g\in k[t]$, $a\colon V\to V$ — *линейный оператор. Тогда если* f *делится на* g, *то* $\operatorname{Ker}(g(a))\leqslant\operatorname{Ker}(f(a))$.

Доказательство. Пусть $f=gh$. Если $v\in\operatorname{Ker}(g(a))$, то

$$f(a)(v)=(g(a)h(a))(v)=h(a)(g(a)(v))=h(a)(0)=0,$$

поэтому $v\in\operatorname{Ker}(f(a))$. □

Теорема 8.5.7. *Пусть* $a\colon V\to V$ — *линейный оператор, а многочлен* $f=gh$, *причем* $g,h\in k[t]$ — *взаимно простые многочлены:* $\gcd(g,h)=1$. *Тогда* $\operatorname{Ker}(f(a))=\operatorname{Ker}(g(a))\oplus\operatorname{Ker}(h(a))$.

Доказательство. По лемме 8.5.6 $\operatorname{Ker}(g(a))$ и $\operatorname{Ker}(h(a))$ — подпространства в $\operatorname{Ker}(f(a))$. В силу взаимной простоты g и h найдутся такие многочлены $r,s\in k[t]$, что $rg+sh=1\in k[t]$. Подставляя оператор a, получаем $r(a)g(a)+s(a)h(a)=\mathrm{id}_V$.

Покажем, что $\operatorname{Ker}(g(a))\cap\operatorname{Ker}(h(a))=0$. Действительно, применяя полученное равенство к вектору $v\in\operatorname{Ker}(g(a))\cap\operatorname{Ker}(h(a))$, получаем $(r(a)g(a))(v)+(s(a)h(a))(v)=v$. Но из равенств $g(a)(v)=0$ и $h(a)(v)=0$ следует, что левая часть равна нулю.

Осталось показать, что $\operatorname{Ker}(f(a))=\operatorname{Ker}(g(a))+\operatorname{Ker}(h(a))$. Снова запишем

$$v=(r(a)g(a))(v)+(s(a)h(a))(v)$$

для произвольного $v\in\operatorname{Ker}(f(a))$. Проверим, что первое слагаемое лежит в $\operatorname{Ker}(h(a))$, а второе — в $\operatorname{Ker}(g(a))$. Действительно, $h(a)(r(a)g(a)(v))=(r(a)f(a))(v)=0$, и $g(a)(s(a)h(a)(v))=(s(a)f(a))(v)=0$. □

Сейчас мы сведем изучение произвольного оператора к оператору, характеристический многочлен которого является степенью неприводимого. Пусть $a\colon V\to V$ — линейный оператор. Рассмотрим каноническое разложение его характеристического многочлена $p_a\in k[t]$ на неприводимые множители: $p_a=(-1)^n p_1^{l_1}\dots p_m^{l_m}$, где p_i — попарно различные неприводимые многочлены со старшим коэффициентом 1.

Теорема 8.5.8. *При этих условиях пространство* V *раскладывается в прямую сумму* $V = \operatorname{Ker}(p_1^{l_1}(a)) \oplus \cdots \oplus \operatorname{Ker}(p_m^{l_m}(a))$. *Каждое из подпространств в этой сумме инвариантно для оператора* a.

Доказательство. Проведем индукцию по m. Для $m = 1$ получаем $V = \operatorname{Ker}(p_a(a))$, а это в точности теорема Кэли–Гамильтона. Пусть теперь $m > 1$ и $g = (-1)^n p_1^{l_1} \cdot \dots \cdot p_{m-1}^{l_{m-1}}$, так что $p_a = g p_m^{l_m}$. Так как $\gcd(p_i, p_m) = 1$ для всех $i = 1, \dots, m-1$, то $\gcd(g, p_m^{l_m}) = 1$. Поэтому мы можем применить теорему 8.5.7 и получить $V = \operatorname{Ker}(p_a(a)) = \operatorname{Ker}(g(a)) \oplus \operatorname{Ker}(p_m^{l_m}(a))$. По лемме 8.5.5 пространства $\operatorname{Ker}(g(a))$ и $\operatorname{Ker}(p_m^{l_m}(a))$ являются инвариантными для оператора a. При этом характеристический многочлен оператора $a|_{\operatorname{Ker}(g(a))}$ имеет меньше различных неприводимых множителей, чем характеристический многочлен a, поэтому к нему можно применить предположение индукции. □

Посмотрим, что произойдет, если характеристический многочлен оператора a полностью раскладывается на линейные множители (например, это так, если основное поле k алгебраически замкнуто). Пусть $p_a = (-1)^n (t-\lambda_1)^{l_1} \dots (t-\lambda_m)^{l_m}$, где $\lambda_1, \dots, \lambda_m \in k$ попарно различны.

Лемма 8.5.9. *В этом случае* $\operatorname{Ker}((t-\lambda_i)^{l_i}(a)) = V(\lambda_i)$ *— корневое подпространство, соответствующее собственному числу* λ_i.

Доказательство. Если $v \in \operatorname{Ker}((t-\lambda_i)^{l_i}(a)) = \operatorname{Ker}((a-\lambda_i)^{l_i})$, то $(a-\lambda_i)^{l_i}(v) = 0$, и по определению вектор v является корневым, то есть, лежит в $V(\lambda_i)$. Поэтому $\operatorname{Ker}((t-\lambda_i)^{l_i}(a)) \leqslant V(\lambda_i)$. Обратно, если $v \in V(\lambda_i)$, то $(a-\lambda_i)^d(v) = 0$ для некоторого d; рассмотрим минимальное такое d и покажем, что $d \leqslant l_i$, и, таким образом, $v \in \operatorname{Ker}((a-\lambda_i)^{l_i})$. Минимальный многочлен элемента v обязан быть делителем $(t-\lambda_i)^d$, поэтому в силу минимальности d он равен $(t-\lambda_i)^d$. С другой стороны, $p_a(a)(v) = 0$, поэтому (по теореме 8.4.9) многочлен p_a делится на минимальный многочлен $(t-\lambda_i)^d$ элемента v. Из разложения p_a на неприводимые следует, что $d \leqslant l_i$. □

Теорема 8.5.10 (корневое разложение). *Если характеристический многочлен* p_a *оператора* $a\colon V \to V$ *полностью раскладывается на линейные множители, то* $V = V(\lambda_1) \oplus \dots V(\lambda_m)$.

Доказательство. Вытекает из теоремы 8.5.8 и леммы 8.5.9. □

8.6 Жордановы клетки

Литература: [F], гл. XII, § 6, п. 4; [K2], гл. 2, § 4, пп. 1, 2; [KM], ч. 1, § 9; [vdW], гл. XII, § 87.

Определение 8.6.1. Жордановой клеткой $J_r(\lambda)$ размера r с собственным значением λ называется матрица вида

$$J_r(\lambda) = \begin{pmatrix} \lambda & 1 & 0 & \dots & 0 \\ 0 & \lambda & 1 & \dots & 0 \\ 0 & 0 & \lambda & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & \lambda \end{pmatrix}.$$

Жордановой матрицей называется блочная матрица, диагональные блоки которой являются жордановыми клетками (возможно, различного размера и с различными собственными значениями), а остальные блоки нулевые. **Жордановым базисом** пространства V относительно оператора $a\colon V \to V$ называется такой базис пространства V, в котором матрица a является жордановой.

Пусть матрица оператора a в некотором базисе является жордановой клеткой $J_r(\lambda)$. Тогда характеристический многочлен оператора a равен $(\lambda - t)^n$. Поэтому минимальный многочлен a должен иметь вид $(t-\lambda)^j$ для некоторого $j \leqslant n$. Но прямое вычисление показывает, что матрица оператора $(a-\lambda)^j$ в том же базисе имеет вид

$$(J_r(\lambda) - \lambda E_r)^j = \begin{pmatrix} 0 & 0 & \dots & 1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & 0 & \dots & 0 \end{pmatrix},$$

где единицы стоят на j-ой диагонали выше главной. Поэтому $j = n$. Таким образом, минимальный многочлен такого оператора a совпадает (с точностью до знака) с характеристическим.

Наша цель — показать, что у любого оператора над алгебраически замкнутым полем имеется жорданов базис. А именно, мы докажем следующую теорему.

Теорема 8.6.2. *Если характеристический многочлен оператора $a\colon V \to V$ полностью раскладывается над полем k на линейные множители, то в пространстве V существует жорданов базис относительно a. Получившаяся жорданова форма оператора a единственна с точностью до перестановки жордановых клеток.*

Следствие 8.6.3. *Над алгебраически замкнутым полем k каждая матрица приводится к жордановой форме.*

Фактически, мы уже свели доказательство этой теоремы к случаю одного собственного числа. Действительно, корневое разложение (теорема 8.5.10) утверждает, что $V = V(\lambda_1) \oplus \dots \oplus V(\lambda_m)$, где $\lambda_1, \dots, \lambda_m$ — попарно различные собственные числа оператора a. Поэтому матрица оператора a имеет блочно-диагональный вид, блоки которой — матрица ограничения оператора a на $V(\lambda_i)$. Характеристический многочлен этого ограничения равен $(\lambda_i - t)^{l_i}$, поэтому каждое такое ограничение имеет ровно одно собственное число. Поэтому можно считать, что оператор a имеет единственное собственное число λ, так что $V = V(\lambda)$.

Рассмотрим теперь оператор $a - \lambda \mathrm{id}_V$. Нетрудно видеть, что этот оператор имеет единственное собственное число 0, и для любого базиса $\mathcal{B}$ матрица $[a - \lambda \mathrm{id}_V]_{\mathcal{B}}$ равна $[a]_{\mathcal{B}} - \lambda E$. Поэтому достаточно доказать существование жорданова базиса только для случая $\lambda = 0$. В этом случае $p_a = (-t)^n$, $V = V(0) = \mathrm{Ker}(a^n)$, поэтому $a^n = 0$. Такой оператор называется **нильпотентным**.

Значит, для доказательства теоремы 8.6.2 осталось доказать существование жорданова базиса и единственность набора размеров жордановых клеток для нильпотентного оператора.

8.7 Жорданов базис 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.

Теорема 8.7.1. *Пусть $a\colon V \to V$ — нильпотентный оператор. Тогда у пространства V существует жорданов базис, причем размеры жордановых клеток определены однозначно.*

Оставшаяся часть раздела посвящена доказательству этой теоремы.

Пусть m — наименьшее число, для которого $a^m = 0$. Рассмотрим ядра последовательных степеней оператора a:

$$0 \leqslant \operatorname{Ker}(a) \leqslant \operatorname{Ker}(a^2) \leqslant \cdots \leqslant \operatorname{Ker}(a^{m-1}) \leqslant \operatorname{Ker}(a^m) = V.$$

На первом шаге построим относительный базис $u_1, \ldots, u_{n_m}$ пространства V относительно $\operatorname{Ker}(a^{m-1})$. Применим к каждому из векторов $u_1, \ldots, u_{n_m}$ оператор a до тех пор, пока не получится 0. Мы получим векторы $u_i, a(u_i), a^2(u_i), \ldots, a^{m-1}(u_i)$ для $i = 1, \ldots, n_m$. Для каждого фиксированного i эти векторы порождают пространство размерности n_m, и в базисе из этих векторов матрица ограничения оператора a является жордановой клеткой $J_{n_m}(0)$ размера m (мы пока не знаем, что эти векторы линейно независимы; это будет доказано позже). Мы получим, таким образом, n_m жордановых клеток порядка m.

Ключевое соображение для продолжения этой процедуры сформулировано в следующей лемме.

Лемма 8.7.2. *Предположим, что векторы $v_1, \ldots, v_s \in \operatorname{Ker}(a^r)$ линейно независимы относительно $\operatorname{Ker}(a^{r-1})$. Тогда их образы $a(v_1), \ldots, a(v_s) \in \operatorname{Ker}(a^{r-1})$ линейно независимы относительно $\operatorname{Ker}(a^{r-2})$.*

Доказательство. Предположим, что эти векторы линейно зависимы над $\operatorname{Ker}(a^{r-2})$. Это означает, что

$$a(v_1)c_1 + \cdots + a(v_s)c_s \in \operatorname{Ker}(a^{r-2})$$

для некоторых $c_1, \ldots, c_s \in k$, не равных 0 одновременно. Это равенство означает, что $a(v^1c_1 + \cdots + v_sc_s) \in \operatorname{Ker}(a^{r-2})$, то есть, что

$$a^{r-2}(a(v_1c_1 + \cdots + v_sc_s)) = a^{r-1}(v_1c_1 + \cdots + v_sc_s) = 0.$$

Но это означает, что

$$v_1c_1 + \cdots + v_sc_s \in \operatorname{Ker}(a^{r-1}),$$

то есть, что векторы $v_1, \ldots, v_s$ линейно зависимы относительно $\operatorname{Ker}(a^{r-1})$, что противоречит условию. □

Эта лемма позволяет нам сделать следующий шаг в конструкции жорданова базиса. А именно, поскольку $u_1, \ldots, u_{n_m}$ образуют относительный базис $\operatorname{Ker}(a^m)$ над $\operatorname{Ker}(a^{m-1})$, то векторы $a(u_1), \ldots, a(u_{n_m}) \in \operatorname{Ker}(a^{m-1})$ линейно независимы над $\operatorname{Ker}(a^{m-2})$. Этот набор векторов

можно дополнить до относительного базиса $\operatorname{Ker}(a^{m-1})$ над $\operatorname{Ker}(a^{m-2})$ векторами $v_1, \dots, v_{n_{m-1}}$. Как и ранее, применим к каждому из только что добавленных векторов оператор a до тех пор, пока не получится 0. Мы получим векторы $v_i, a(v_i), a^2(v_i), \dots, a^{m-2}(v_i)$ для всех $i = 1, \dots, n_{m-1}$. Каждый такой набор для фиксированного i порождает инвариантное подпространства, и матрица ограничения оператора a на него является жордановой клеткой $J_{m-1}(0)$ порядка m. Мы получим, таким образом, n_{m-1} жордановых клеток порядка $m-1$.

Теперь уже ясно, что мы можем продолжать действовать таким образом. Опишем еще один шаг этой процедуры. Согласно лемме 8.7.2 векторы

$$a^2(u_1), a^2(u_2), \dots, a^2(u_{n_m}), a(v_1), \dots, a(v_{n_{m-1}}) \in \operatorname{Ker}(a^{m-2})$$

линейно независимы над $\operatorname{Ker}(a^{m-3})$. Поэтому можно дополнить их векторами $w_1, \dots, w_{n_{m-2}}$ до базиса $\operatorname{Ker}(a^{m-2})$. Применяем оператор a к каждому из добавленных векторов, получаем цепочки $w_i, a(w_i), a^2(w_i), \dots, a^{m-3}(w_i)$ для $i = 1, \dots, n_{m-2}$. Каждая такая цепочка дает нам инвариантное подпространство, матрица ограничения оператора a на которое является жордановой клеткой порядка $m-2$. Мы получили n_{m-2} жордановых клеток порядка $m-2$.

Мы будем продолжать так, пока не дойдем до собственных векторов: на последнем шаге мы выбираем дополнение $x_1, \dots, x_{n_1}$ векторов

$$a^{m-1}(u_1), \dots, a^{m-1}(u_{n_m}), a^{m-2}(v_1), \dots, a^{m-2}(v_{n_{m-1}}), a^{m-3}(w_1), \dots, a^{m-3}(w_{n_{m-2}}), \dots$$

до базиса $\operatorname{Ker}(a)$. Этим векторам отвечают n_1 жордановых клеток оператора a порядка 1.

Мы утверждаем, что все построенные векторы образуют жорданов базис пространства V. Докажем сначала, что они линейно независимы. Рассмотрим произвольную линейную зависимость между ними. Она имеет вид $u_1\alpha_1 + \dots + u_{n_m}\alpha_{n_m} + (\dots) = 0$. При этом «хвост» $(\dots)$ лежит в $\operatorname{Ker}(a^{m-1})$. Мы выбирали $u_1, \dots, u_{n_m}$ так, чтобы они образовывали относительный базис $V = \operatorname{Ker}(a^m)$ над $\operatorname{Ker}(a^{m-1})$, поэтому $\alpha_1 = \dots = \alpha_{n_m} = 0$.

Значит, оставшаяся часть линейной зависимости равна 0. Она имеет вид $a(u_1)\beta_1 + \dots + a(u_{n_m})\beta_{n_m} + v_1\gamma_1 + \dots + v_{n_{m-1}}\gamma_{n_{m-1}} + (\dots) = 0$. При этом «хвост» $(\dots)$ лежит в $\operatorname{Ker}(a^{m-2})$. Из построения $v_1, \dots, v_{n_{m-1}}$ (они вместе с $a(u_1), \dots, a(u_{n_m})$ образуют относительный базис $\operatorname{Ker}(a^{m-1})$ над $\operatorname{Ker}(a^{m-2})$) следует, что $\beta_1 = \dots = \beta_{n_m} = \gamma_1 = \dots = \gamma_{n_{m-1}} = 0$.

Продолжая действовать таким образом, мы видим, что все коэффициенты линейной зависимости между указанными векторами равны 0. Поэтому они линейно независимы.

С другой стороны, эти векторы порождают V: по построению векторы $x_1, \dots, x_{n_1}$ вместе с векторами из $\operatorname{Ker}(a)$, полученными из предыдущих шагов, образуют базис $\operatorname{Ker}(a)$. Перед этим мы выбрали несколько векторов так, чтобы они образовывали относительный базис $\operatorname{Ker}(a^2)$ над $\operatorname{Ker}(a)$, поэтому, если добавить их к базису $\operatorname{Ker}(a)$, получится базис $\operatorname{Ker}(a^2)$. И так далее: на последнем шаге мы убеждаемся, что построенные векторы образуют базис $\operatorname{Ker}(a^m) = V$, что и требовалось.

Матрица оператора a в этом базисе имеет блочно-диагональный вид, и на диагонали стоит n_m матриц вида $J_m(0)$, n_{m-1} матриц вида $J_{m-1}(0)$, и вообще количество жордановых клеток размера k равно n_k.

Осталось доказать единственность набора размеров жордановых клеток. Заметим, что $\dim(\mathrm{Ker}(a)) = n_1 + \cdots + n_m$, $\dim(\mathrm{Ker}(a^2)) - \dim(\mathrm{Ker}(a)) = n_2 + \cdots + n_m$, и так далее, до $\dim(\mathrm{Ker}(a^{m-1}) - \dim(\mathrm{Ker}(a^{m-2}) = n_{m-1} + n_m$ и $n - \dim(\mathrm{Ker}(a^{m-1})) = n_m$. Таким образом, количество жордановых клеток размера i равно $n_i = 2\dim(\mathrm{Ker}(a^i)) - \dim(\mathrm{Ker}(a^{i+1})) - \dim(\mathrm{Ker}(a^{i-1}))$ и, таким образом, однозначно определяется размерностями ядер операторов a^i. Поэтому набор $(n_1, \ldots, n_m)$ является инвариантом самого оператора a и не зависит от выбора жорданова базиса.

9 Теория групп

9.1 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 5.5.1):

Определение 9.1.1. 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 $a \circ a^{-1} = a^{-1} \circ a = e$.

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 5.5.2):

Определение 9.1.2. 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 5.5 мы видели, что группа S_n не является абелевой при $n \geqslant 3$.

На самом деле мы встречали и другие группы.

Примеры 9.1.3. 1. Пусть R — кольцо (см.определение 2.8.4). 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 $k^* = k \setminus \{0\}$ 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.

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 5.7.5). 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. 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.

9.2 Подгруппы

Литература: [F], гл. X, § 1, пп. 3–4, § 3, п. 6; [vdW], гл. 2, § 7; [Bog], гл. 1, § 1.

Ситуация, описанная в примере 9.1.3 (6), встречается достаточно часто:

Определение 9.2.1. 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 9.2.2.
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 5.5.7), причем произведение четных перестановок четно (теорема 5.5.12), и обратная к четной перестановке четна (следствие 5.5.13). 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 9.2.3. *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 $G \setminus \{0\}$. 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 9.2.4. *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 9.2.5. 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 9.2.6. 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 9.2.5 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 9.2.7. *Пусть* G *— группа,* $X \subseteq G$. *Пересечение всех подгрупп в* G, *содержащих* X, *является подгруппой в* G, *порожденной множеством* X.

Доказательство. По лемме 9.2.4 пересечение всех подгрупп в G, содержащих X, является подгруппой в G. Обозначим ее через $\langle X\rangle$ и проверим, что она удовлетворяет определению 9.2.5. 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 9.2.8. Обратите внимание на сходство предложения 9.2.7 и определения линейной оболочки 6.2.7. Понятие подгруппы, порожденной множеством элементов G, является точным аналогом понятия линейной оболочки множества элементов векторного пространства. Следующая лемма совершенно аналогична лемме 6.2.8.

Лемма 9.2.9. *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 9.2.7.

Для доказательства обратного включения заметим, что множество Y само является подгруппой в G, содержащей множество X. В силу определения 9.2.5 из этого следует, что $\langle X \rangle \leqslant Y$. □

Следующее понятие продолжает эту мысль, вводя аналог понятия *системы образующих* векторного пространства (см. определение 6.2.9).

Определение 9.2.10. 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 9.2.11.
1. Предложение 5.5.4 в точности показывает, что группа S_n порождается множеством всех транспозиций, а вместе с предложением 5.5.5 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: $-n = \underbrace{(-1) + (-1) + \dots + (-1)}$.

9.3 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 9.3.1. 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 9.3.2. *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 9.3.2 чрезвычайно похоже на теорему 1.5.4 о разбиении на классы эквивалентности. Это не случайно: за смежными классами стоят достаточно естественные отношения эквивалентности.

Определение 9.3.3. 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 9.3.4. *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 9.3.5. 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 9.3.6. Отношения $\sim_H$ и ${}_H\sim$ являются прямыми аналогами сравнения по модулю подпространства (см. определение 7.2.1); однако, отсутствие коммутативности приводит к тому, что необходимо рассматривать два варианта обобщения: условие $v_1 - v_2 \in U$ из определения 7.2.1 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 7.2.2). 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 9.3.7. Пусть G — группа. Подгруппа $H \leqslant G$ называется **нормальной** (обозначение: $H \trianglelefteq G$), если для любого элемента $g \in G$ его левый и правый смежный классы совпадают: $Hg = gH$.

Полезны следующие переформулировки нормальности.

Лемма 9.3.8. *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 9.3.9. 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 9.3.10. Из замечания 9.3.6 следует, что все подгруппы абелевой группы нормальны.

Примеры 9.3.11. 1. $SL(n,k) \trianglelefteq GL(n,k)$. Действительно, если $h \in SL(n,k)$ и $g \in GL(n,k)$, то $\det(ghg^{-1}) = \det(g) \cdot \det(h) \cdot \det(g^{-1}) = \det(h) = 1$, поэтому ${}^g h \in SL(n,k)$.

2. $A_n \trianglelefteq S_n$. Это доказывается совершенно аналогично предыдущему примеру, с заменой определителя на знак перестановки. Нормальность в обоих этих примерах также следует из леммы 9.4.5.

3. Любая подгруппа индекса 2 нормальна. Мы докажем это чуть позже.

9.4 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 9.4.1. 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 9.4.2. *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 9.4.3. 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 5.5.12).

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 5.7.5).

Определение 9.4.4. 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 9.4.5. *Образ гомоморфизма $\varphi\colon G \to H$ является подгруппой в H, а его ядро — нормальной подгруппой в G: $\mathrm{Im}(\varphi) \leqslant H$, $\mathrm{Ker}(\varphi) \trianglelefteq G$.*

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 9.3.8 этого достаточно для доказательства нормальности $\mathrm{Ker}(\varphi) \trianglelefteq G$. □

Замечание 9.4.6. Сравните с предложением 7.3.1. 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 9.4.7. 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 9.4.8. *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 9.4.9. 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 9.4.10. *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. □

9.5 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 $H \trianglelefteq G$ — 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 9.5.1. *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 9.5.2.

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.

Теперь мы можем доказать аналог теоремы о гомоморфизме 7.3.3.

Теорема 9.5.3 (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 9.4.10 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 9.4.8 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 $\widetilde{\varphi}(g\operatorname{Ker}(\varphi)) = \varphi(g) = h$. □

9.6 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 $g^n = \underbrace{g \cdot \ldots \cdot g}_{n}$ и $g^{-n} = \underbrace{g^{-1} \cdot \ldots \cdot g^{-1}}_{n}$. 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 9.4.5 является подгруппой в G.

Лемма 9.6.1. *Образ отображения* pow_g *совпадает с* $\langle g\rangle$ *(подгруппой, порожденная* g*).*

Доказательство. Прежде всего, $\operatorname{Im}(\operatorname{pow}_g)$ содержит g, поэтому и $\langle g\rangle \subseteq \operatorname{Im}(\operatorname{pow}_g)$. С другой стороны, любой элемент $\operatorname{Im}(\operatorname{pow}_g)$ имеет вид g^n для некоторого n, и содержится в $\langle g\rangle$, поскольку $\langle g\rangle$ — подгруппа в G. □

Определение 9.6.2. 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 9.6.3 (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 9.6.1 его образ совпадает с $\langle g \rangle = G$. По теореме о гомоморфизме 9.5.3 имеем $\mathbb{Z}/\operatorname{Ker}(\mathrm{pow}_g) \cong G$. По теореме 9.2.3 $\operatorname{Ker}(\mathrm{pow}_g)$, будучи подгруппой в $\mathbb{Z}$, имеет вид $m\mathbb{Z}$ для некоторого натурального m, что и требовалось доказать. □

Следствие 9.6.4. *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 9.6.3. □

Определение 9.6.5. 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 9.6.6. Заметим, что порядок нейтрального элемента равен 1, и это единственный элемент порядка 1 в любой группе.

9.7 Теорема Лагранжа

Литература: [F], гл. X, § 1, пп. 5, 7; [K3], гл. 1, § 2, п. 1; [Bog], гл. 1, § 2.

Определение 9.7.1. 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 9.7.2. *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 9.3.2 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 $g \in Ha_i^{-1}$ равносильно тому, что $g = ha_i^{-1}$ для некоторого H, откуда $g^{-1} = (ha_i^{-1})^{-1} = a_i h^{-1} \in a_i H$. Но это равенство выполнено ровно для одного индекса $i \in I$, поэтому g лежит ровно в одном множестве вида Ha_i^{-1}, что и требовалось доказать. □

Замечание 9.7.3. По определению фактор-множество G/H состоит из правых смежных классов G по H, так что $|G : H| = |G/H|$.

Теорема 9.7.4 (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 9.7.5. *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 9.7.6. *Пусть* G *— конечная группа. Тогда* $g^{|G|} = 1$ *для любого* $g \in G$.

В качестве примера приложения теоремы Лагранжа выведем из нее теорему Эйлера 2.11.2 (и, как следствие, малую теорему Ферма 2.11.3).

Теорема 9.7.7. *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 9.1.3 (4)). Порядок этой группы равен $\varphi(m)$ (предложение 2.10.3). Класс $\overline{a}$ элемента a в $\mathbb{Z}/m\mathbb{Z}$ обратим, поскольку $a \perp m$ (предложение 2.8.10). Применение следствия 9.7.6 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 9.7.8. *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 9.6.6 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 9.6.3) изоморфна $\mathbb{Z}/p\mathbb{Z}$. □

9.8 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 9.8.1. Множество $G \times H$ с такой операцией называется **прямым произведением** групп G и H.

Предложение 9.8.2. *Пусть G, H — группы. Рассмотрим отображения*

$$\begin{gathered} i_1 \colon G \to G \times H,\ \ g \mapsto (g, e), \\ i_2 \colon H \to G \times H,\ \ h \mapsto (e, h), \\ \pi_1 \colon G \times H \to G,\ \ (g, h) \mapsto g, \\ \pi_2 \colon G \times H \to H,\ \ (g, h) \mapsto h. \end{gathered}$$

1. *i_1, i_2 — инъективные, а π_1, π_2 — сюръективные гомоморфизмы групп;*
2. *$\operatorname{Im}(i_1) = \operatorname{Ker}(\pi_2) = G \times \{e\}$, $\operatorname{Im}(i_2) = \operatorname{Ker}(\pi_1) = \{e\} \times H$ — 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. $\operatorname{Im}(i_1)$ состоит в точности из элементов вида (g, e), а $\operatorname{Ker}(\pi_2)$ 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 9.4.5. 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 9.8.3. *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 9.4.8). 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 $\pi_2(\varphi(j_2(h))) = \pi_2(\varphi(h)) = \pi_2((e, h)) = h$. □

9.9 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 9.9.1. Перестановка $\pi \in S_n$ называется **циклом длины** k, если для некоторых различных $i_1, \ldots, i_k \in \{1, \ldots, n\}$ выполнено $\pi(i_1) = i_2$, $\pi(i_2) = i_3$, …, $\pi(i_{k-1}) = i_k$, $\pi(i_k) = i_1$, и для всех $j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$ 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 9.9.2. 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 9.9.3. *Независимые циклы коммутируют: если $\pi, \rho \in S_n$ — независимые циклы, то $\pi\rho = \rho\pi$.*

Доказательство. Непосредственное вычисление. □

Определение 9.9.4. 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 9.9.5. *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 $j \in \{1, \ldots, n\} \setminus \{i_1, \ldots, i_k\}$.

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 $\mathrm{Fix}(\pi) = \mathrm{Fix}(\pi') \setminus \{i_1, \ldots, i_k\}$. □

Определение 9.9.6. Запись элемента $\pi \in S_n$ в виде, указанном в теореме, называется **цикленной записью перестановки** π.

Пример 9.9.7. 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 9.9.8. Как мы видели выше (замечание 9.9.2), запись цикла в виде $(i_1\ \ i_2\ \ \ldots\ \ i_k)$ не вполне однозначна: на первое место можно поставить любой элемент из $i_1, \ldots, i_k$. 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 9.9.9. *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 9.9.10. 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 9.9.11. *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 9.9.9 имеем ${}^{\pi}\rho_i = \tau_i$ для всех $i = 1, \ldots, n$. Поэтому и ${}^{\pi}\rho = \tau$. □

Замечание 9.9.12. Из доказательства теоремы 9.9.11 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 9.9.13 (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: $G = \{g_1, \ldots, g_n\}$. Сопоставим каждому элементу $g \in G$ перестановку $\pi_g \in S_n$ следующим образом: для $i = 1, \ldots, n$ 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 $\{1, \ldots, n\}$ (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 $i = 1, \ldots, n$ 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 9.5.3 изоморфен фактору G по ядру гомоморфизма π, которое тривиально. Поэтому группа $\mathrm{Im}(\pi)$ изоморфна G и является подгруппой в S_n. □

9.10 Диэдральная группа

Литература: [К3], гл. 1, § 4, п. 5.

Рассмотрим на эвклидовой плоскости правильный n-угольник с вершинами $A_1, \ldots, A_n$ и центром в начале координат (точке O). Множество всех поворотов плоскости, переводящих этот n-угольник в себя, образует группу (см. пример 9.1.3 (7)). 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, $A_{n+1} = A_1$, $A_{n+2} = A_2$, $\ldots$). 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): $1, \varepsilon, \varepsilon^2, \ldots, \varepsilon^{n-1}$. 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 $k = 1, \ldots, n-1$:

$$yx^k = \begin{pmatrix} \cos(2\pi k/n) & \sin(2\pi k/n) \\ \sin(2\pi k/n) & -\cos(2\pi k/n) \end{pmatrix}.$$

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 9.10.1. *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 $\langle x\rangle \trianglelefteq D_n$ (впрочем, нормальность следует и из примера 9.3.11 (3): $\langle x\rangle$ имеет индекс 2 в D_n). □

Замечание 9.10.2. 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 9.7.8) и изоморфна G. Однако, D_n не является прямым произведением G и H (при $n > 2$): не хватает условия 3 из теоремы 9.8.3. 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 9.8.2, пункт 2).

10 Эвклидовы и унитарные пространства

10.1 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 10.1.1. 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: $u \perp v$. Если вектор ортогонален сам себе, он называется **изотропным** относительно формы B.

Определение 10.1.2. 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 все векторы пространства V **изотропны** относительно этой формы, то есть, если $B(u, u) = 0$ для всех $u \in V$.

Замечание 10.1.3. 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 10.1.4. 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 10.1.5. 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

$$B(u,v) = B(u,v)^T = (u^Tv)^T = v^Tu = B(v,u).$$

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 10.1.6. 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 10.1.7. 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 10.1.8. 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.

10.2 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 10.2.1. 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 10.2.2. Полуторалинейная форма $B\colon V\times V\to\mathbb{C}$ называется **эрмитовой**, если $B(u,v)=\overline{B(v,u)}$ для всех $u,v\in V$.

Замечание 10.2.3. 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 10.2.4. Пусть $V=\mathbb{C}^n$ — пространство столбцов высоты n над k. Определим форму $B\colon V\times V\to\mathbb{C}$ формулой $B(u,v)=\overline{u_1}v_1+\dots+\overline{u_n}v_n$. Иными словами, $B(u,v)=\overline{u}^Tv$. Нетрудно видеть, что эта форма полуторалинейная

$$\begin{aligned}
&B(u,v_1+v_2)=u^T(v_1+v_2)=\overline{u}^Tv_1+\overline{u}^Tv_2=B(u,v_1)+B(u,v_2)\\
&B(u,v\lambda)=\overline{u}^T(v\lambda)=\lambda(\overline{u}^Tv)=\lambda B(u,v)\\
&B(u_1+u_2,v)=\overline{(u_1+u_2)}^Tv=\overline{u_1}^tv+\overline{u_2}^Tv=B(u_1,v)+B(u_2,v)\\
&B(u\lambda,v)=\overline{(u\lambda)}^Tv=\overline{\lambda}(\overline{u}^Tv)=\overline{\lambda}B(u,v)
\end{aligned}$$

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 10.2.5. 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 10.2.6. 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 10.2.7. Как и в эвклидовом случае (см. замечание 10.1.8), 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: $B(u,v) = \overline{B(v,u)}$.

10.3 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 10.3.1. Пусть (V, B) — эвклидово или унитарное пространство, $v \in V$. Будем называть число $\|v\| = \sqrt{B(v,v)}$ **длиной** v.

Лемма 10.3.2. *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 10.3.3. Пусть (V, B) — эвклидово пространство. Лемма 10.3.2 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 10.3.4. Заметим, что угол $\angle(u, v)$ равен $\pi/2$ тогда и только тогда, когда $B(u, v)$, то есть, когда векторы u и v ортогональны в смысле определения 10.1.1.

10.4 Матрица Грама

Литература: [F], гл. XIII, § 1, п. 4; [КМ], ч. 2, § 2, пп. 2–3; [КМ], 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 $\overline{B(v, u)} = \overline{B(v, u)}^T = (\overline{[v]}^T G[u])^T = \overline{[u]^T G^T [v]} = \overline{[u]}^T \overline{G}^T [v]$. 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 $e_1, \ldots, e_n$, 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 10.4.1. 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 10.1.5 и 10.2.4. 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 $k = \mathbb{C}$:

$$B(u, v_1 + v_2) = \overline{u}^T G(v_1 + v_2) = \overline{u}^T G v_1 + \overline{u}^T G v_2 = B(u, v_1) + B(u, v_2)$$
$$B(u, v\lambda) = \overline{u}^T G(v\lambda) = (\overline{u}^T G v)\lambda = B(u, v)\lambda$$
$$B(u_1 + u_2, v) = \overline{u_1 + u_2}^T G v = \overline{u_1}^T G v + \overline{u_2}^T G v = B(u_1, v) + B(u_2, v)$$
$$B(u\lambda, v) = \overline{u\lambda}^T G v = \overline{\lambda}(\overline{u}^T G v) = \overline{\lambda} B(u, v)$$

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 $\mathcal{E} = \{e_1, \ldots, e_n\}$ и $\mathcal{F} = \{f_1, \ldots, f_n\}$ — 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 6.6.1).

Теорема 10.4.2 (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 6.6.3 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 10.4.3. *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 7.9.3) уравнение $GX = 0$ имеет ненулевое решение: найдется столбец $X_0 \in k^n \setminus \{0\}$, 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. □

10.5 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 10.5.1. 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 10.5.2. *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 6.4.6 следует, что они образуют базис (действительно, размерность их линейной оболочки совпадает с размерностью V, поэтому эта линейная оболочка равна V). □

Замечание 10.5.3. 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 10.1.5 и 10.2.4: стандартные базисы пространства столбцов являются ортонормированными относительно рассматриваемых там форм).

Лемма 10.5.4 (Процесс ортогонализации Грама–Шмидта). *Пусть* (V, B) — *эвклидово или унитарное пространство,* $e_1, \dots, e_{n-1}$ — *семейство попарно ортогональных ненулевых векторов,* $v \notin \langle e_1, \dots, e_{n-1}\rangle$. *Тогда существует вектор* $e_n \in V$ *такой, что* e_n *ортогонален всем векторам* $e_1, \dots, e_{n-1}$ *и, кроме того,* $\langle e_1, \dots, e_{n-1}, v\rangle = \langle e_1, \dots, e_{n-1}, e_n\rangle$.

Доказательство. Будем искать вектор e_n в виде

$$e_n = v - e_1\lambda_1 - e_2\lambda_2 - \dots - e_{n-1}\lambda_{n-1}.$$

Подберем коэффициенты $\lambda_1, \dots, \lambda_{n-1} \in \mathbb{F}$ так, чтобы e_n был ортогонален каждому e_i, $i = 1, \dots, n-1$. Посмотрим на скалярное произведение e_n и e_i. Поскольку e_i ортогонален всем векторам из $e_1, \dots, e_{n-1}$, 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 $i = 1, \dots, n-1$. Кроме того, v выражается через $e_1, \dots, e_n$, поэтому $v \in \langle e_1, \dots, e_n\rangle$, и e_n выражается через $e_1, \dots, e_{n-1}, v$, поэтому $e_n \in \langle e_1, \dots, e_{n-1}, v\rangle$. Это и означает равенство нужных линейных оболочек. □

Следствие 10.5.5. *Пусть* (V, B) — *эвклидово или унитарное пространство, и пусть* $\mathcal{F} = (f_1, \dots, f_n)$ — *базис* V. *Тогда существует ортогональный базис* $\mathcal{E} = (e_1, \dots, e_n)$ *пространства* V *такой, что* $\langle e_1, \dots, e_k\rangle = \langle f_1, \dots, f_k\rangle$ *для всех* $k = 1, \dots, n$.

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 $f_1, \dots, f_{n-1}$: $W = \langle f_1, \dots, f_{n-1}\rangle$. По предположению индукции найдется ортогональный базис $e_1, \dots, e_{n-1}$ пространства W такой, что $\langle e_1, \dots, e_k\rangle = \langle f_1, \dots, f_k\rangle$ для всех $k = 1, \dots, n-1$.

Применим лемму 10.5.4 к набору $e_1, \dots, e_{n-1}$ и вектору f_n. Мы найдем вектор e_n такой, что $e_1, \dots, e_n$ — ортогональная система векторов, и $\langle e_1, \dots, e_n\rangle = \langle f_1, \dots, f_n\rangle = V$, то есть, $e_1, \dots, e_n$ — базис V. Очевидно, что условие $\langle e_1, \dots, e_k\rangle = \langle f_1, \dots, f_k\rangle$ теперь выполняется для всех $k = 1, \dots, n$. □

Следствие 10.5.6. *В любом [конечномерном] эвклидовом или унитарном пространстве существует ортогональный (и даже ортонормированный) базис.*

Доказательство. Применим следствие 10.5.5 к произвольному базису пространства V. Получим ортогональный базис $e_1, \dots, e_n$. Положим $e'_i = e_i/\|e_i\|$; легко видеть, что $\|e'_i\| = 1$ и векторы $e'_1, \dots, e'_n$ все еще попарно ортогональны. Мы получили ортонормированный базис пространства V. □

Следствие 10.5.7. *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 10.5.5, воспользуемся леммой 10.5.4 для индуктивного построения нужного базиса. □

10.6 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 10.6.1. 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 10.6.2. 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 10.6.3. *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 10.1.5).*
4. *строки C образуют ортонормированный базис в эвклидовом пространстве ${}^n\mathbb{R}$ со стандартным эвклидовым скалярным произведением.*

Лемма 10.6.4. *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 10.2.4).*
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 10.5.2 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 10.6.5. *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 10.4.2 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. □

10.7 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 10.7.1. Пусть (V, B) — эвклидово или унитарное пространство, $U \leqslant V$ — подпространство. **Ортогональным дополнением** к подпространству U в V называется $U^\perp = \{v \in V \mid \forall u \in U \;\; B(u, v) = 0\}$.

Теорема 10.7.2. *Пусть (V, B) — конечномерное эвклидово или унитарное пространство, $U \leqslant V$ — подпространство. Тогда*

1. *$U^\perp$ является подпространством в V;*
2. *$0^\perp = V$, $V^\perp = 0$;*
3. *$\dim(U^\perp) + \dim(U) = \dim(V)$.*
4. *$U \cap U^\perp = 0$; $U + U^\perp = V$.*
5. *$U^{\perp\perp} = U$;*
6. *сопоставление $U \rightsquigarrow U^\perp$ обращает включения, то есть, для любых двух подпространств U, W в V из $U \leqslant W$ следует, что $U^\perp \geqslant W^\perp$;*

7. $(U + W)^\perp = U^\perp \cap W^\perp$;

8. $(U \cap W)^\perp = U^\perp + W^\perp$.

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 $e_1, \ldots, e_m$ — некоторый ортонормированный базис подпространства U (такой существует по следствию 10.5.6). Рассмотрим следующее отображение $\varphi\colon V \to V$:

$$w \mapsto w - B(e_1, w)e_1 - B(e_2, w)e_2 - \cdots - B(e_m, w)e_m.$$

Нетрудно видеть, что оно линейно (например, потому, что это сумма тождественного отображения и отображений вида $w \mapsto -B(e_i, w)e_i$, которые линейны в силу линейности B по второму аргументу). Заметим, что $\varphi(w) \perp e_i$ для любого $i = 1, \ldots, m$. Действительно,

$$B(e_i, \varphi(w)) = B(e_i, w - \sum B(e_j, w)e_j) = B(e_i, w) - B(e_i, w)B(e_i, e_i) = 0.$$

Поэтому $\varphi(w)$ ортогонален любой линейной комбинации векторов e_i, и, значит, $\varphi(w) \in U^\perp$. Стало быть, $\mathrm{Im}(\varphi) \subseteq U^\perp$. Верно и обратное включение: если $w \in U^\perp$, то $\varphi(w) = w$, и потому $U^\perp \subseteq \mathrm{Im}(\varphi)$. Мы получили, что $\mathrm{Im}(\varphi) = U^\perp$.

Покажем, что $\mathrm{Ker}(\varphi) = U$. Действительно, $\varphi(e_i) = e_i - B(e_i, e_i)e_i = 0$, поэтому $e_i \in \mathrm{Ker}(\varphi)$, и, значит, $U \subseteq \mathrm{Ker}(\varphi)$. Обратно, если $\varphi(w) = 0$, то $w - \sum_i B(e_i, w)e_i = 0$, и w является линейной комбинацией векторов $e_1, \ldots, e_m$; это означает, что $\mathrm{Ker}(\varphi) \subseteq U$.

Осталось применить теорему о сумме размерностей ядра и образа (теорема 7.5.4) к отображению φ: $\dim(\mathrm{Im}(\varphi)) + \dim(\mathrm{Ker}(\varphi)) = \dim(V)$.

4. Пусть вектор $u \in V$ лежит одновременно в U и в $U^\perp$. Тогда, в частности, $B(u, u) = 0$, откуда $u = 0$. Поэтому $U \cap U^\perp = 0$. Воспользуемся следствием 7.5.6: $\dim(U + U^\perp) = \dim(U) + \dim(U^\perp) - \dim(U \cap U^\perp)$; в предыдущем пункте мы показали, что $\dim(U) + \dim(U^\perp) = \dim(V)$, поэтому $\dim(U + U^\perp) = \dim(V)$, откуда $U + U^\perp = V$.

5. Покажем сначала, что $U \leqslant U^{\perp\perp}$. Возьмем $u \in U$; нам нужно показать, что $B(u', u) = 0$ для всех $u' \in U^\perp$. Но из определения $U^\perp$ следует, что $B(u, u') = 0$; поэтому и $B(u', u) = 0$. Теперь заметим, что $\dim(U^\perp) = \dim(V) - \dim(U)$, поэтому $\dim(U^{\perp\perp}) = \dim(V) - \dim(U^\perp) = \dim(V) - (\dim(V) - \dim(U)) = \dim(U)$, поэтому во включении $U \leqslant U^{\perp\perp}$ имеет место равенство.

6. Предположим, что $U \leqslant W$, и рассмотрим вектор $v \in W^\perp$. По определению $B(w, v) = 0$ для всех $w \in W$; в частности, $B(w, v) = 0$ для всех $w \in U$. Но это и означает, что $v \in U^\perp$.

7. Поскольку $U \leqslant U + W$, и $W \leqslant U + W$, из предыдущего пункта следует, что $(U + W)^\perp \leqslant U^\perp$ и $(U + W)^\perp \leqslant W^\perp$; поэтому $(U + W)^\perp \leqslant U^\perp \cap W^\perp$. Обратно, если $v \in U^\perp \cap W^\perp$, то v ортогонален всем векторам из U и всем векторам из W. Возьмем произвольный $x \in U + W$ и запишем его в виде $x = u + w$; тогда $B(x, v) = B(u, v) + B(w, v) = 0$; поэтому v ортогонален всем векторам из $U + W$.

8. Подставим в предыдущий пункт $U^\perp$, $W^\perp$ вместо U, W: получим, что $(U^\perp + W^\perp)^\perp = U^{\perp\perp} \cap W^{\perp\perp}$, откуда $(U^\perp + W^\perp)^\perp = U \cap W$. Переходя к ортогональным дополнениям, получаем, что $U^\perp + W^\perp = (U \cap W)^\perp$, что и требовалось.

□

Пусть теперь (V, B) и (V', B') — два эвклидовых или унитарных пространства. Определим на их прямой сумме $V \oplus V'$ форму $B \oplus B'$ следующим образом: для векторов $(v_1, v_1'), (v_2, v_2') \in V \oplus V'$ положим $(B \oplus B')((v_1, v_1'), (v_2, v_2')) = B(v_1, v_2) + B'(v_1', v_2')$. Непосредственная проверка показывает, что мы получили билинейную форму, если B и B' были билинейными, и полуторалинейную форму для полуторалинейных B, B'. Аналогично, из симметричности B и B' следует симметричность $B \oplus B'$, а из эрмитовости B и B' следует эрмитовость $B \oplus B'$. Наконец, если B и B' — положительно определенные формы, то и $B \oplus B'$ положительно определена. Действительно, для $(v, v') \in V \oplus V'$ по определению $(B \oplus B')((v, v'), (v, v')) = B(v, v) + B'(v', v')$, и каждое из слагаемых неотрицательно. Если эта сумма равна 0, то каждое слагаемое нулевое, откуда $v = 0$ и $v' = 0$, поэтому и $(v, v') = 0 \in V \oplus V'$.

Таким образом, мы получили новое эвклидово или унитарное пространство $(V \oplus V', B \oplus B')$. Заметим, что в векторное пространство $V \oplus V'$ вложены подпространства V и V' стандартным образом: $v \mapsto (v, 0)$, $v' \mapsto (0, v')$. Ограничение формы $B \oplus B'$ на эти пространства совпадают с B и B' соответственно. Действительно, $(B \oplus B')((v_1, 0), (v_2, 0)) = B(v_1, v_2) + B'(0, 0) = B(v_1, v_2)$ и $(B \oplus B')((0, v_1'), (0, v_2')) = B(0, 0) + B'(v_1', v_2') = B'(v_1', v_2')$. При этом все векторы из V оказались ортогональны всем векторам из V': $(B \oplus B')((v, 0), (0, v')) = B(v, 0) + B'(0, v') = 0$. Поэтому мы будем называть пространство $V \oplus V'$ с формой $B \oplus B'$ **ортогональной прямой суммой** эвклидовых (или унитарных) пространств (V, B) и (V', B'). Обозначение: $V \boxplus V' = (V \oplus V', B \oplus B')$. Это аналог понятия *внешней прямой суммы* пространств (см. раздел 7.4). Несложно определить и аналог *внутренней прямой суммы*: говорят, что эвклидово или унитарное пространство (V, B) является [внутренней] ортогональной прямой суммой своих подпространств $U, W \leqslant V$, если, во-первых, V является прямой суммой U и W, и, во-вторых, любой вектор из U ортогонален любому вектору из W, то есть, $B(u, w) = 0$ для всех $u \in U, w \in W$. Например, свойство (4) теоремы 10.7.2 означает, что V является ортогональной прямой суммой подпространств U и $U^\perp$.

Лемма 10.7.3. *Пусть (V, B) — эвклидово или унитарное пространство, $U \leqslant V$ — подпространство, $\mathcal{E} = (e_1, \ldots, e_m)$ — ортогональный базис U. Тогда любое дополнение $\mathcal{E}$ до ортогонального базиса всего пространства V является ортогональным базисом*

пространства $U^\perp$. *Иными словами, если* $(e_1, \dots, e_m, f_1, \dots, f_k)$ — *ортогональный базис пространства* V, *то* $f_1, \dots, f_k$ — *ортогональный базис пространства* $U^\perp$.

Доказательство. Для всех $i = 1, \dots, k$ вектор f_i ортогонален векторам $e_1, \dots, e_m$, поэтому f_i ортогонален и всем их линейным комбинациям. Значит, $f_i \in U^\perp$. По теореме 10.7.2 размерность пространства $U^\perp$ равна $\dim(V) - \dim(U) = (k+m) - m = k$. Таким образом, векторы $f_1, \dots, f_k \in U^\perp$ линейно независимы и их число равно размерности $U^\perp$; поэтому они образуют базис $U^\perp$. □

10.8 Сопряженные отображения

Литература: [F], гл. XIII, § 4, п. 2; [K2], гл. 3, § 3, п. 1; [KM], ч. 2, § 8, пп. 1–3.

Определение 10.8.1. Пусть (V, B) и (V', B') — эвклидовы или унитарные пространства, $\varphi\colon V \to V'$ — линейное отображение. Линейное отображение $\psi\colon V' \to V$ называется **сопряженным** к отображению φ, если $B'(\varphi(v), v') = B(v, \psi(v'))$ для всех векторов $v \in V$ и $v' \in V'$.

Покажем, что у каждого линейного отображения между эвклидовыми или унитарными пространствами имеется единственное сопряженное.

Для этого нам понадобятся две леммы.

Лемма 10.8.2. *Пусть* (V, B) — *эвклидово или унитарное пространство,* $v_1, v_2 \in V$. *Предположим, что* $B(v, v_1) = B(v, v_2)$ *для всех* $v \in V$. *Тогда* $v_1 = v_2$.

Доказательство. Если $B(v, v_1) = B(v, v_2)$ для всех $v \in V$, то $B(v, v_1 - v_2) = 0$ для всех $v \in V$. Это означает, что $v_1 - v_2 \in V^\perp$. Но по теореме 10.7.2 $V^\perp = 0$, поэтому $v_1 = v_2$. □

Лемма 10.8.3 (Теорема Риса). *Пусть* (V, B) — *конечномерное эвклидово или унитарное пространство. Любое линейное отображение* $f\colon V \to k$ *имеет вид* $f(v) = B(v_f, v)$ *для некоторого однозначно определенного вектора* $v_f \in V$. *В унитарном случае, кроме того, любое полулинейное отображение* $g\colon V \to k$ *имеет вид* $g(v) = B(v, v_g)$ *для некоторого вектора* $v_g \in V$.

Доказательство. Пусть $e_1, \dots, e_n$ — произвольный ортонормированный базис пространства V. Положим $v_f = e_1\overline{f(e_1)} + e_2\overline{f(e_2)} + \dots + e_n\overline{f(e_n)}$. Отображение $v \mapsto B(v_f, v)$ является линейным; проверим, что оно совпадает с f. Для этого достаточно проверить, что оно совпадает с f на базисных векторах $e_1, \dots, e_n$. Действительно, $B(v_f, e_i) = B(e_1\overline{f(e_1)} + e_2\overline{f(e_2)} + \dots + e_n\overline{f(e_n)}, e_i) = f(e_1)B(e_1, e_i) + f(e_2)B(e_2, e_i) + \dots + f(e_n)B(e_n, e_i) = f(e_i)B(e_i, e_i) = f(e_i)$. Покажем, что такой вектор v_f единственный: если v_f' — другой такой вектор, то $B(v_f, v) = f(v) = B(v_f', v)$ для всех $v \in V$, и по лемме 10.8.2 имеем $v_f = v_f'$.

Если теперь g — полулинейное отображение, то отображение $f\colon v \mapsto \overline{g(v)}$ является линейным; по уже доказанному, найдется вектор $v_f \in V$ такой, что $f(v) = B(v, f_v)$ для всех $v \in V$. Но тогда $B(f_v, v) = \overline{B(v, f_v)} = \overline{f(v)} = g(v)$ для всех $v \in V$. □

Предложение 10.8.4. *Пусть* (V, B) *и* (V', B') *— эвклидовы или унитарные пространства,* $\varphi\colon V \to V'$ *— линейное отображение. Существует линейное отображение* $\psi\colon V' \to V$ *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 (лемма 10.8.3) найдется вектор $v_f \in V$ такой, что $B(v, v_f) = f(v) = B'(\varphi(v), v')$ для всех $v \in V$. Положим $\psi(v') = v_f$.

Таким образом, для каждого $v' \in V'$ мы нашли вектор $\psi(v') \in V$ такой, что $B(v, \psi(v')) = B'(\varphi(v), v')$ для всех $v \in V$. Проверим, что полученное отображение $\psi\colon V' \to V$ является линейным. Действительно.

$$\begin{aligned} B(v, \psi(v'_1)\lambda_1 + \psi(v'_2)\lambda_2) &= B(v, \psi(v'_1))\lambda_1 + B(v, \psi(v'_2))\lambda_2 \\ &= B'(\varphi(v), v'_1)\lambda_1 + B'(\varphi(v), v'_2))\lambda_2 \\ &= B'(\varphi(v), v'_1\lambda_1 + v'_2\lambda_2). \end{aligned}$$

С другой стороны, по определению ψ выполнено $B(v, \psi(v'_1\lambda_1 + v'_2\lambda_2)) = B'(\varphi(v), v'_1\lambda_1 + v'_2\lambda_2)$. Поэтому $B(v, \psi(v'_1\lambda_1 + v'_2\lambda_2)) = B(v, \psi(v'_1)\lambda_1 - \psi(v'_2)\lambda_2)$ для всех $v \in V$, откуда по лемме 10.8.2 следует, что $\psi(v'_1\lambda_1 + v'_2\lambda_2) = \psi(v'_1)\lambda_1 - \psi(v'_2)\lambda_2$.

Осталось показать единственность отображения ψ с указанным свойством. Но если $\widetilde{\psi}$ — другое такое отображение, то $B(v, \psi(v')) = B'(\varphi(v), v') = B(v, \widetilde{\psi}(v'))$ для всех $v \in V$, $v' \in V'$. По лемме 10.8.2 из этого следует, что $\psi(v') = \widetilde{\psi}(v')$ для каждого v'. □

Сопряженное к отображению φ мы будем обозначать через φ^*. Таким образом, если $\varphi\colon V \to V'$, то $\varphi^*\colon V' \to V$ и $B'(\varphi(v), v') = B(v, \varphi^*(v'))$ для всех $v \in V$, $v \in V'$.

Выясним, как выглядит матрица сопряженного отображения в ортонормированных базисах.

Предложение 10.8.5. *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* $[\varphi^*]_{\mathcal{E}',\mathcal{E}} = \overline{[\varphi]_{\mathcal{E},\mathcal{E}'}}^{\mathsf{T}}$.

Доказательство. Обозначим $A = [\varphi]_{\mathcal{E},\mathcal{E}'}$, $A^* = [\varphi^*]_{\mathcal{E}',\mathcal{E}}$. По основному свойству матрицы линейного отображения (теорема 7.8.1) 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

$$\overline{[\varphi(v)]_{\mathcal{E}'}}^{\mathsf{T}} \cdot [v']_{\mathcal{E}'} = B'(\varphi(v), v') = B(v, \varphi^*(v')) = \overline{[v]_{\mathcal{E}}}^{\mathsf{T}} \cdot [\varphi^*(v')]_{\mathcal{E}}.$$

Подставляя сюда выражения для столбцов координат $\varphi(v)$ и $\varphi^*(v')$, получаем

$$\overline{A \cdot [v]_{\mathcal{E}}}^{\mathsf{T}} \cdot [v']_{\mathcal{E}'} = \overline{[v]_{\mathcal{E}}}^{\mathsf{T}} \cdot A^* \cdot [v']_{\mathcal{E}'},$$

откуда

$$\overline{[v]_{\mathcal{E}}}^{\mathsf{T}} \cdot \overline{A}^{\mathsf{T}} \cdot [v']_{\mathcal{E}'} = \overline{[v]_{\mathcal{E}}}^{\mathsf{T}} \cdot A^* \cdot [v']_{\mathcal{E}'}.$$

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 $A^* = \overline{A}^{\mathsf{T}}$. □

Лемма 10.8.6. *Пусть* (V, B), (V', B'), (V'', B'') — *эвклидовы или унитарные пространства,* $f, g\colon V \to V'$, $f'\colon V' \to V''$ — *линейные отображения,* $\lambda \in k$. *Тогда*

1. $(f + g)^* = f^* + g^*$;
2. $(\lambda f)^* = \overline{\lambda} f^*$,
3. $(f' \circ f)^* = f^* \circ f'^*$.

Доказательство. Все эти свойства немедленно следует из леммы 10.8.5, соответствия между линейными отображениями и матрицами (раздел 7.8), и свойств транспонирования (теорема 5.3.2). Их можно доказать и напрямую: к примеру, $B(v, (f+g)^*(v')) = B'((f+g)(v), v') = B'(f(v) + g(v), v') = B'(f(v), v') + B'(g(v), v') = B(v, f^*(v')) + B(v, g^*(v') = B(v, f^*(v') + g^*(v')) = B(v, (f^* + g^*)(v'))$ для всех $v \in V$, $v' \in V'$, откуда по лемме 10.8.2 следует, что $(f + g)^*(v') = (f^* + g^*)(v')$ для всех $v' \in V'$, что доказывает (1). □

10.9 Нормальные операторы

Литература: [F], гл. XIII, § 4, п. 3; [K2], гл. 3, § 3, п. 7; [KM], ч. 2, § 8, п. 11.

Определение 10.9.1. Пусть (V, B) — эвклидово или унитарное пространство. Линейный оператор $a\colon V \to V$ называется **нормальным**, если он коммутирует со своим сопряженным: $a^* \circ a = a \circ a^*$.

Лемма 10.9.2 (Свойства нормальных операторов). *1. Тождественный оператор нормален.*

2. Сопряженный к нормальному оператору нормален.

Доказательство. Очевидно. □

Лемма 10.9.3 (Матрица нормального оператора). *Пусть* (V, B) — *эвклидово или унитарное пространство,* $a\colon V \to V$ — *линейный оператор,* $\mathcal{E}$ — *ортонормированный базис пространства* V, $A = [a]_{\mathcal{E}}$ — *матрица оператора* a *в этом базисе. Оператор* a *нормален тогда и только тогда, когда* $A \cdot \overline{A}^{\mathsf{T}} = \overline{A}^{\mathsf{T}} \cdot A$.

Доказательство. Поскольку базис $\mathcal{E}$ ортонормирован, $[a^*]_{\mathcal{E}} = \overline{A}^{\mathsf{T}}$ (предложение 10.8.5). Поэтому $[a \circ a^*]_{\mathcal{E}} = A\overline{A}^{\mathsf{T}}$ и $[a^* \circ a]_{\mathcal{E}} = \overline{A}^{\mathsf{T}} A$; осталось заметить, что операторы равны тогда и только тогда, когда равны их матрицы. □

Следующая теорема верна только в унитарном пространстве!

Теорема 10.9.4. *Пусть* (V, B) *— конечномерное унитарное пространство. Линейный оператор* $a\colon V \to V$ *является нормальным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *диагональна.*

Доказательство. Предположим, что матрица оператора a в базисе $\mathcal{E}$ диагональна. Обозначим $A = [a]_{\mathcal{E}}$, тогда $\overline{A}^T = [a^*]_{\mathcal{E}}$ по предложению 10.8.5. Пусть

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}, \overline{A}^T = \begin{pmatrix} \overline{\lambda_1} & 0 & \dots & 0 \\ 0 & \overline{\lambda_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \overline{\lambda_n} \end{pmatrix}.$$

Тогда матрицы операторов $a \circ a^*$ и $a^* \circ a$ в том же базисе выглядят так:

$$[a \circ a^*]_{\mathcal{E}} = A \cdot \overline{A}^T = \begin{pmatrix} |\lambda_1|^2 & 0 & \dots & 0 \\ 0 & |\lambda_2|^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & |\lambda_n|^2 \end{pmatrix} = \overline{A}^T \cdot A = [a^* \circ a]_{\mathcal{E}},$$

поэтому и сами операторы совпадают.

Обратно, пусть теперь a — нормальный оператор. Покажем, что в некотором ортонормированном базисе матрица оператора a диагональна. Доказательство будет проведено индукцией по размерности пространства V. Случай $\dim(V) = 1$ очевиден. Пусть теперь $n = \dim(V) \geqslant 2$. Поскольку поле $\mathbb{C}$ алгебраически замкнуто, у оператора a имеются собственные числа, и, таким образом, хотя бы один собственный вектор. Пусть это вектор $v \in V$, соответствующий собственному числу $\lambda \in \mathbb{C}$; иными словами, $a(v) = \lambda v$, $v \neq 0$. После домножения на скаляр можно считать, что длина вектора v равна 1. По следствию 10.5.7 вектор v можно дополнить до ортонормированного базиса. Матрица оператора a в этом базисе тогда имеет клеточно-верхнетреугольный вид

$$\begin{pmatrix} \lambda & U \\ 0 & B \end{pmatrix} \in M(n, \mathbb{C}),$$

где $U = (u_1, \dots, u_{n-1}) \in M(1, n-1, \mathbb{C})$ — некоторая строка, а $B \in M(n-1, \mathbb{C})$ — квадратная матрица некоторого размера (см. предложение 8.3.4). Покажем, что на самом деле $U = 0$. Действительно, матрица оператора a^* в этом же базисе имеет вид

$$\begin{pmatrix} \overline{\lambda} & 0 \\ \overline{U}^T & \overline{B}^T \end{pmatrix},$$

и потому матрицы операторов $a \circ a^*$ и $a^* \circ a$ имеют вид

$$\begin{pmatrix} \lambda\overline{\lambda} + U\overline{U}^T & U\overline{B}^T \\ B\overline{U}^T & B\overline{B}^T \end{pmatrix} \text{ и } \begin{pmatrix} \lambda\overline{\lambda} & \overline{\lambda}B \\ \lambda\overline{U}^T & \overline{U}^T U + \overline{B}^T B \end{pmatrix},$$

соответственно. Но оператор a нормален, поэтому полученные матрицы должны совпадать. В частности, $\lambda\overline{\lambda} + U\overline{U}^T = \lambda\overline{\lambda}$, откуда $U\overline{U}^T = 0$. Поэтому $u_1\overline{u_1} + u_2\overline{u_2} + \cdots + u_{n-1}\overline{u_{n-1}} = 0$ и $u_1 = u_2 = \cdots = u_n$. Итак, $U = 0$ и матрица оператора a имеет клеточно-диагональный вид

$$\begin{pmatrix} \lambda & 0 \\ 0 & B \end{pmatrix}.$$

Более того, перемножая матрицы операторов a и a^* мы видим, что из нормальности оператора a следует, что $B\overline{B}^T = \overline{B}^T B$. Таким образом, V распадается в прямую сумму двух подпространств, инвариантных относительно оператора a (см. предложение 8.3.4): это подпространство $\langle v \rangle$ и подпространство V', натянутое на все остальные векторы выбранного базиса. При этом ограничение оператора a на V' имеет матрицу B, поэтому оно также является нормальным оператором (лемма 10.9.3). К нему теперь можно применить предположение индукции: в подпространстве V' имеется ортонормированный базис $\mathcal{E}'$, в котором матрица ограничения $a|_{V'}$ диагональна. Осталось вспомнить, что подпространство V' было ортогонально вектору v, поэтому и каждый вектор базиса $\mathcal{E}'$ ортогонален вектору v. Значит, при добавлении вектора v к $\mathcal{E}'$ мы получим ортонормированный базис пространства V, и легко видеть, что матрица оператора a в нем диагональна. □

Замечание 10.9.5. Напомним, что если матрица оператора a диагональна, то на диагонали в ней стоят собственные числа оператора a (с учетом кратности), и геометрическая кратность каждого собственного числа равна его алгебраической кратности В этом случае пространство раскладывается в прямую сумму собственных подпространств, соответствующих различным собственным числам (см. предложение 8.2.8 и теорему 8.2.9). Нормальный оператор имеет диагональный вид в ортонормированном базисе, поэтому сумма собственных подпространств для него является не просто прямой, но и ортогональной.

10.10 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 10.10.1. 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 10.10.2. Нетрудно видеть, что самосопряженные, кососимметрические, унитарные, ортогональные операторы являются нормальными.

Лемма 10.10.3. *Пусть* (V, B) *— эвклидово или унитарное пространство,* $\mathcal{E}$ *— ортонормированный базис пространства* V*,* $a\colon V \to V$ *— линейный оператор,* $A = [a]_{\mathcal{E}}$ *— матрица оператора* a *в этом базисе.*

1. *Оператор* a *самосопряжен тогда и только тогда, когда* $A = A^*$*.*
2. *Оператор* a *кососимметричен тогда и только тогда, когда* $A = -A^*$*.*
3. *Оператор* a *на унитарном пространстве унитарен тогда и только тогда, когда его матрица унитарна, то есть,* $A^*A = AA^* = E$*.*
4. *Оператор* a *на эвклидовом пространстве ортогонален тогда и только тогда, когда его матрица ортогональна, то есть,* $A^TA = AA^T = E$*.*

Доказательство. Очевидно (см. также лемму 10.9.3). □

В случае унитарного пространства описание нормальных операторов, полученное в теореме 10.9.4, позволяет характеризовать самосопряженные, кососимметрические и унитарные операторы следующим образом.

Теорема 10.10.4. *Пусть* (V, B) *— конечномерное унитарное пространство,* $a\colon V \to V$ *— линейный оператор.*

1. *Оператор* a *является самосопряженным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *диагональна, и все ее диагональные элементы вещественны.*
2. *Оператор* a *является кососимметрическим тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *диагональна, и все ее диагональные элементы — чисто мнимые комплексные числа.*
3. *Оператор* a *является унитарным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *диагональна, и все ее диагональные элементы — комплексные числа, равные по модулю* 1*.*

Доказательство. Если оператор a самосопряженный/кососимметрический/унитарный, то он нормален и по теореме 10.9.4 существует ортонормированный базис $\mathcal{E}$, в котором его матрица $A = [a]_{\mathcal{E}}$ диагональна:

$$A = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

Если теперь a самосопряженный, то равенство $A = A^*$ влечет $\lambda_i = \overline{\lambda_i}$ для всех i, поэтому $\lambda_i \in \mathbb{R}$. Если a кососимметрический, то равенство $A = -A^*$ влечет $\lambda_i = -\overline{\lambda_i}$ для всех i, поэтому $\lambda_i \in i\mathbb{R}$. Если a унитарный, то равенство $AA^* = E$ влечет $\lambda_i\overline{\lambda_i} = 1$ для всех i, поэтому $|\lambda_i| = 1$.

Обратно, если в ортонормированном базисе $\mathcal{E}$ матрица оператора a диагональна, то по теореме 10.9.4 он нормален, и из соответствующих равенств для всех λ_i следуют равенства для его матрицы, а потому и для самого оператора. □

10.11 Нормальные операторы в эвклидовых пространствах

Литература: [F], гл. XIII, § 5; [K2], гл. 3, § 3, пп. 3, 6; [KM], ч. 2, § 7, пп. 4–5; § 8, пп. 2–6, 8.

Перейдем теперь к эвклидовым пространствам. Покажем, что теоремы 10.9.4 и 10.10.4 не могут быть верны в этом случае. Рассмотрим двумерное пространство $V = \mathbb{R}^2$ со стандартным эвклидовым скалярным произведением и пусть $a\colon V \to V$ — оператор на V, задаваемый (в стандартном ортонормированном базисе) матрицей $A = \begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$ для некоторых $\alpha, \beta \in \mathbb{R}$. Тогда оператора a^* задается в том же базисе матрицей $A^T = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$. Нетрудно видеть, что $A \cdot A^T = A^T \cdot A = \begin{pmatrix} \alpha^2 + \alpha^2 & 0 \\ 0 & \alpha^2 + \beta^2 \end{pmatrix}$, и потому оператор a нормален. Кроме того, если $\alpha^2 + \beta^2 = 1$, то оператор a ортогонален, а если $\alpha = 0$, то он кососимметричен. Однако, при $\beta \neq 0$ не существует базиса, в котором этот оператор был бы диагонален. Действительно, в этом случае его собственные числа были бы вещественными. Но нетрудно видеть, что собственные числа оператора a равны $\alpha \pm \beta i$.

Тем не менее, оказывается, что приведенный пример в каком-то смысле единственный. А именно, выполнена следующая теорема о канонической форме нормального оператора в эвклидовом пространстве (аналог теоремы 10.9.4).

Теорема 10.11.1. *Пусть (V, B) — конечномерное эвклидово пространство. Линейный оператор $a\colon V \to V$ является нормальным тогда и только тогда, когда существует ортонормированный базис пространства V, в котором матрица оператора a имеет блочно-диагональный вид, и каждый блок выглядит как (a) для $a \in \mathbb{R}$ или $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$ для $\alpha, \beta \in \mathbb{R}$, $\beta \neq 0$.*

Для доказательства теоремы нам понадобятся две леммы.

Лемма 10.11.2. *Пусть (V, B) — конечномерное эвклидово или унитарное пространство, $a\colon V \to V$ — нормальный оператор. Если вектор $v \in V$ является собственным вектором оператора a, соответствующим собственному числу $\lambda \in \mathbb{C}$, то он является собственным вектором и для оператора a^*, соответствующим собственному числу $\overline{\lambda}$.*

Доказательство. Заметим сначала, что если $c\colon V \to V$ — нормальный оператор, то длины векторов $c(v)$ и $c^*(v)$ совпадают:

$$\|c(v)\|^2 = B(c(v), c(v)) = B(c, c^*(c(v)) = B(v, c(c^*(v))) = B(c^*(v), c^*(v)) = \|c^*(v)\|^2.$$

Кроме того, нетрудно видеть, что из нормальности оператора a следует нормальность оператора $a - \lambda\,\mathrm{id}$. Действительно, $(a-\lambda\,\mathrm{id})(a-\lambda\,\mathrm{id})^* = (a-\lambda\,\mathrm{id})(a^*-\overline{\lambda}\,\mathrm{id}) = aa^* - \overline{\lambda}a - \lambda a^* + |\lambda|^2$. С другой стороны, $(a-\lambda\,\mathrm{id})^*(a-\lambda\,\mathrm{id}) = (a^*-\overline{\lambda}\,\mathrm{id})(a-\lambda\,\mathrm{id}) = a^*a - \overline{\lambda}a - \lambda a^* + |\lambda|^2$, что то же самое.

Поэтому $a(v) = \lambda v$ равносильно $(a - \lambda\,\mathrm{id})(v) = 0$, что равносильно $\|(a - \lambda\,\mathrm{id})(v)\| = 0$. Аналогично, $a^*(v) = \overline{\lambda}v$ равносильно $\|(a^* - \overline{\lambda}\,\mathrm{id})(v)\| = 0$. Но $\|(a - \lambda\,\mathrm{id})(v)\| = 0$ равносильно тому, что $\|(a^* - \overline{\lambda}\,\mathrm{id})(v)\| = 0$ в силу нашего замечания про длины. □

Лемма 10.11.3. *Пусть (V, B) — эвклидово или унитарное пространство, $a\colon V \to V$ — нормальный оператор. Если $v_1, v_2 \in V$ — собственные векторы оператора a, соответствующие различным собственным числам $\lambda_1 \neq \lambda_2$ соответственно, то v_1 ортогонален v_2.*

Доказательство. По условию $a(v_1) = \lambda_1 v_1$, $a(v_2) = \lambda_2 v_2$, и поэтому $\lambda_2 B(v_1, v_2) = B(v_1, \lambda_2 v_2) = B(v_1, a(v_2)) = B(a^*(v_1), v_2) = B(\overline{\lambda_1} v_1, v_2) = \lambda_1 B(v_1, v_2)$. Из этого следует, что $(\lambda_1 - \lambda_2)B(v_1, v_2)$; по предположению $\lambda_1 \neq \lambda_2$, поэтому $B(v_1, v_2) = 0$, что и требовалось. □

Доказательство теоремы 10.11.1. Мы будем проводить доказательство совершенно параллельно доказательству теоремы 10.9.4. Пусть a — нормальный оператор. Если у него имеется собственный вектор, соответствующий некоторому (вещественному!) собственному числу λ, то можно повторить рассуждение из доказательства теоремы 10.9.4: выбрать соответствующий этому числу собственный вектор длины 1 в качестве первого элемента базиса, дополнить его до ортонормированного базиса, тогда матрица оператора a примет блочный вид $\begin{pmatrix} \lambda & u \\ 0 & B \end{pmatrix}$; из условия нормальности следует, что $u = 0$ и B — матрица нормального оператора $a|_W$, где W — подпространство на единицу меньшей размерности, и можно применить индукцию.

Предположим теперь, что у оператора a нет собственных векторов. Пусть A — матрица нашего оператора в *каком-нибудь* ортонормированном базисе. Рассмотрим ее как матрицу с комплексными коэффициентами: $A \in M(n, \mathbb{C})$. У полученной матрицы есть (комплексное!) собственное число $\lambda \in \mathbb{C}$ и соответствующий ему собственный вектор $z \in \mathbb{C}^n$. После домножения на скаляр можно считать, что $\|z\| = 1$. Таким образом, $Az = \lambda z$. Применим комплексное сопряжение к этому равенству (и учтем, что A — вещественная матрица, потому $\overline{A} = A$): $A\overline{z} = \overline{\lambda}\overline{z}$. Это означает, что $\overline{z}$ — собственный вектор матрицы A, соответствующий собственному числу $\overline{\lambda}$. По нашему предположению $\lambda \notin \mathbb{R}$, поэтому $\lambda \neq \overline{\lambda}$. По лемме 10.11.3 из этого следует, что $z \perp \overline{z}$.

Разложим число λ и вектор z на вещественную и мнимую части: запишем $\lambda = \alpha + i\beta$ для $\alpha, \beta \in \mathbb{R}$, и $z = u + vi$ для $u, v \in \mathbb{R}^n$; тогда $\overline{\lambda} = \alpha - i\beta$ и $\overline{z} = u - vi$. Заметим, что

$\|u+vi\|^2 = B(u+vi,u+vi) = B(u,u)+B(u,vi)+B(vi,u)+B(vi,vi) = B(u,u)+B(v,v)+i(B(u,v)-\overline{B(u,v)})$. В то же время, это вещественное число, поэтому $B(u,v)-\overline{B(u,v)}=0$. С другой стороны, $\|u-vi\|^2 = B(u-vi,u-vi) = B(u,u)+B(v,v)-i(B(u,v)-\overline{B(u,v)})$. Поэтому $\|u+vi\| = \|u-vi\|$. В силу нашего выбора вектора z получаем, что $\|z\| = \|\bar{z}\| = 1$. Наконец, $0 = B(z,\bar{z}) = B(u+vi,u-vi) = B(u,u)-B(v,v)-i(B(u,v)+\overline{B(u,v)})$, откуда (заметим, что $B(u,u)$, $B(v,v)$, $B(u,v)$ — вещественные числа!) $B(u,u) = B(v,v)$ и $B(u,v) = 0$. Из этого следует, что $\|z\|^2 = 2\|u\|^2 = 2\|v\|^2$; мы выбрали вектор z так, что $\|z\| = 1$; поэтому векторы u, v после домножения на $\sqrt{2}$ станут ортогональными векторами длины 1.

Пусть V' — двумерное подпространство в V, натянутое на векторы u, v (оно двумерно, поскольку если u, v линейно зависимы, то и $u+vi, u-vi$ линейно зависимы, чего не может быть, поскольку это собственные векторы, соответствующие *различным* собственным числам). Заметим, что подпространство V' инвариантно относительно оператора a: действительно,

$$a(u) = a((z+\bar{z})/2) = (a(z)+a(\bar{z}))/2 = (\lambda z+\overline{\lambda z})/2 = \alpha u - \beta v,$$
$$a(v) = a((z-\bar{z})/(2i)) = (a(z)-a(\bar{z}))/(2i) = (\lambda z-\overline{\lambda z})/(2i) = \alpha v + \beta u.$$

Дополним ортонормированный базис $(u\sqrt{2}, v\sqrt{2})$ до ортонормированного базиса $\mathcal{B}$ всего пространства V. Из нашего вычисления следует, что матрица оператора a в этом базисе равна

$$M = [a]_{\mathcal{B}} = \begin{pmatrix} \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} & U' \\ 0 & B' \end{pmatrix},$$

где $U' \in M(2, n-2, \mathbb{R})$. Покажем, что на самом деле $U' = 0$. Представим матрицу U' в виде $U' = \begin{pmatrix} U'_1 \\ U'_2 \end{pmatrix}$, где U'_1, U'_2 — строчки ширины $n-2$. Матрица M является матрицей нормального оператора a в ортонормированном базисе $\mathcal{B}$, поэтому $MM^T = M^TM$. При этом

$$M = \begin{pmatrix} \alpha & \beta & U'_1 \\ -\beta & \alpha & U'_2 \\ 0 & 0 & B \end{pmatrix}, \quad M^T = \begin{pmatrix} \alpha & -\beta & 0 \\ \beta & \alpha & 0 \\ U'_1 & U'_2 & B^T \end{pmatrix}.$$

Перемножая матрицы и сравнивая коэффициенты, получаем, что $U'_1(U'_1)^T = 0$ и $U'_2(U'_2)^T$, откуда, как и в доказательстве теоремы 10.9.4, следует, что $U'_1 = U'_2 = 0$

Это означает, что матрица M оператора a в базисе $\mathcal{B}$ имеет блочно-диагональный вид с блоками $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ и B; из равенства $MM^T = M^TM$ теперь следует, что $BB^T = B^TB$, поэтому ограничение оператора a на подпространство размерности $n-2$, натянутое на последние $n-2$ вектора базиса B, также является нормальным оператором, и к нему можно применить предположение индукции. □

Следующая теорема — аналог теоремы 10.10.4 для эвклидовых пространств.

Теорема 10.11.4. *Пусть* (V, B) — *конечномерное эвклидово пространство,* $a\colon V \to V$ — *линейный оператор.*

1. *Оператор* a *является самосопряженным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *диагональна.*

2. *Оператор* a *является кососимметрическим тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *имеет блочно-диагональный вид, и каждый блок выглядит как* (0) *или* $\begin{pmatrix} 0 & -\beta \\ \beta & 0 \end{pmatrix}$ *для* $\beta \in \mathbb{R}$*,* $\beta \neq 0$*.*

3. *Оператор* a *является ортогональным тогда и только тогда, когда существует ортонормированный базис пространства* V*, в котором матрица оператора* a *имеет блочно-диагональный вид, и каждый блок выглядит как* (1)*,* (-1) *или* $\begin{pmatrix} \alpha & -\beta \\ \beta & \alpha \end{pmatrix}$ *для* $\alpha, \beta \in \mathbb{R}$*,* $\beta \neq 0$*,* $\alpha^2 + \beta^2 = 1$*.*

Доказательство. Если a самосопряжен, то он нормален, и по теореме 10.11.1 существует ортонормированный базис, в котором его матрица A имеет блочно-диагональный вид с блоками размера 1×1 и 2×2. Однако, диагональные блоки имеют при этом вид $M = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ с $\beta \neq 0$. Из равенства $A = A^T$ следует, что и для каждого такого блока M выполняется равенство $M = M^T$; поэтому $\beta = 0$ — противоречие. Обратно, диагональная матрица A, очевидно, удовлетворяет равенству $A = A^T$, поэтому является матрицей самосопряженного оператора.

Аналогично, если a кососимметричен, то $M^T = -M$ для каждого блока M в матрице A, что для блока (λ) размера 1×1 означает, что $\lambda = 0$, а для блока вида $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ означает, что $\alpha = 0$. Обратно, если матрица A составлена из таких блоков, то $A^T = -A$

Наконец, в случае ортогонального оператора каждый блок M его матрицы удовлетворяет соотношению $M \cdot M^T = E$. Для блока (λ) это означает, что $\lambda^2 = 1$, откуда $\lambda = \pm 1$; а для блока $\begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}$ это означает, что $\alpha^2 + \beta^2 = 1$. Обратно, если матрица A составлена из таких блоков, то $A^T A = E$. □

Определение 10.11.5. 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 10.11.6. *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 $u, v \in V$. По лемме 10.8.2 отсюда следует, что $v = a^*(a(v))$ для всех $v \in V$; значит, $a^* \circ a = \mathrm{id}$.

$2 \Rightarrow 3$ 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

$$2\,\mathrm{Re}(B(a(u), a(v)) = 2\,\mathrm{Re}(B(u, v)), \quad 2\,\mathrm{Im}(B(a(u), a(v)) = 2\,\mathrm{Im}(B(u, v)).$$

Отсюда следует, что $B(a(u), a(v)) = B(u, v)$, что и требовалось.

□

Следствие 10.11.7 (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 10.11.6 оператор a ортогонален. По теореме 10.11.4 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 10.11.8 (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 лемме 10.10.3 оператор q самосопряжен. По теореме 10.11.4 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 7.8.5). Кроме того, поскольку C — матрица перехода между ортонормированными базисами, то C ортогональна (лемма 10.6.3): $C^T = C^{-1}$. Но тогда $D = C^T QC$, и по теореме 10.4.2 это означает, что D — матрица Грама квадратичной формы q в ортонормированном базисе $\mathcal{C}$. □

Замечание 10.11.9. Переформулируем утверждение первого пункта теоремы 10.11.4 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 8.2).

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 $\lambda_1, \dots, \lambda_m$, 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 $\lambda_1, \dots, \lambda_m$.

10.12 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 10.12.1. 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 10.12.2. *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 10.10.4 и 10.11.4 существует ортонормированный базис $\mathcal{B} = (e_1, \dots, e_n)$, в котором a имеет диагональную матрицу

$$[a]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}.$$

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 $\mathcal{B} = \{e_1, \dots, e_n\}$ имеет диагональную матрицу с неотрицательными числами $\lambda_1, \dots, \lambda_n$ на диагонали. По теоремам 10.10.4 и 10.11.4 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 $v = 0$. $\square$

Замечание 10.12.3. 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 10.12.4 (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 10.12.2 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 10.12.2) 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 10.11.9 для оператора $\widetilde{b}$. А именно, пусть $\mu_1, \ldots, \mu_n$ — собственные числа оператора $\widetilde{b}$ с учетом кратности. Тогда $\widetilde{b}$ приводится в некотором базисе к диагональному виду, и на диагонали стоят положительные числа $\mu_1, \ldots, \mu_n$. Но тогда $a = \widetilde{b}^2$ в этом же базисе имеет диагональный вид, и на диагонали стоят числа $\mu_1^2, \ldots, \mu_n^2$. Значит, собственные числа оператора a (с учетом кратности) равны $\mu_1^2, \ldots, \mu_n^2$. С другой стороны, мы знаем, что они равны $\lambda_1, \ldots, \lambda_n$. Мы знаем, что $\mu_i > 0$ для всех i, поэтому набор $\mu_1, \ldots, \mu_n$ совпадает (с точностью до перестановки) с набором $\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n}$.

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 10.11.9.

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 $z \in \mathbb{R} \setminus \{0\}$, то $z = |z| \cdot (\pm 1)$; ортогональный оператор на одномерном пространстве может быть равен лишь 1 или -1.

Теорема 10.12.5 (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 10.12.3). По теореме 10.12.4 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 10.12.4), получаем, что $p = p'$, и, стало быть, $u = u'$. □

Замечание 10.12.6. Даже доказательство теоремы 10.12.5 напоминает доказательство факта про тригонометрическую форму записи комплексного числа: напомним, что модуль ком-

плексного числа z определялся как $\sqrt{z \cdot \overline{z}}$ (см. определение 3.2.3); извлечение корня возможно в силу неотрицательности $z \cdot \overline{z}$.

11 Полилинейная алгебра

11.1 Полилинейные отображения

Литература: [KM], 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 10.1.1).
- 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 5.6).

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.

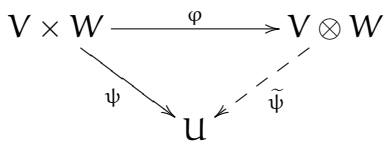
11.2 Тензорное произведение двух пространств

Литература: [F], гл. XIV, § 4, пп. 1, 2; [K2], гл. 6, § 1, п. 5; [KM], ч. 4, § 1, пп. 2–5.

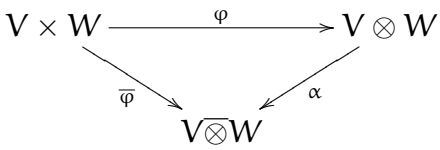
Определение 11.2.1. Пусть V, W — векторные пространства над полем k. **Тензорным произведением** пространств V и W называется векторное пространство $V \otimes W$ вместе с билинейным отображением $\varphi\colon V \times W \to V \otimes W$, удовлетворяющие следующему *универсальному свойству*:

- для любого векторного пространства U и любого билинейного отображения $\psi\colon V \times W \to U$ существует единственное линейное отображение $\widetilde{\psi}\colon V \otimes W \to U$ такое, что $\psi = \widetilde{\psi} \circ \varphi$.

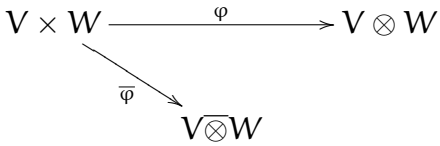
Универсальное свойство можно изобразить следующей диаграммой:



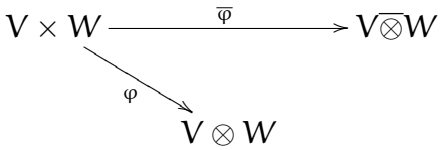
Теорема 11.2.2. *Тензорное произведение любых векторных пространств V, W над полем k существует и единственно с точностью до канонического изоморфизма. Последнее означает, что если $\overline{\varphi}\colon V \times W \to V\overline{\otimes}W$ — еще одно тензорное произведение в смысле определения 11.2.1, то существует единственный изоморфизм векторных пространств $\alpha\colon V \otimes W \to V\overline{\otimes}W$ такой, что $\overline{\varphi} = \alpha \circ \varphi$:*



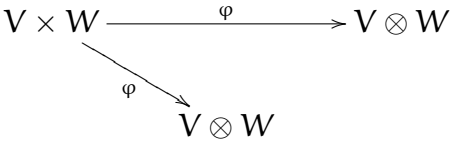
Доказательство. Сначала докажем единственность. Итак, пусть $\varphi\colon V\times W \to V\otimes W$ и $\overline{\varphi}\colon V\times W \to V\overline{\otimes}W$ — два тензорных произведения пространств V и W. Рассмотрим следующую диаграмму:



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



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



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 7.7.1) 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 7.7.1). 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. □

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 $\dim(V \otimes W) = \dim(V) \cdot \dim(W)$.*

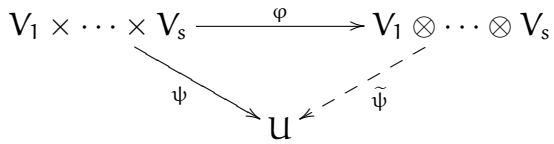
Замечание 11.2.6. Сравните формулу для размерности тензорного произведения с формулой для прямой суммы: $\dim(V \oplus W) = \dim(V) + \dim(W)$. Это свидетельство того, что тензорное произведение и прямая сумма — аналоги умножения и сложения для векторных пространств.

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 $\widetilde{\psi}\colon V_1 \otimes \dots \otimes V_s \to U$ такое, что $\psi = \widetilde{\psi} \circ \varphi$:



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 $\varphi\colon V_1 \times \dots \times V_s \to L \to L/R$ полилинейно. Проверим, что оно универсально: пусть $\psi\colon V_1 \times \dots \times V_s \to U$ — некоторое полилинейное отображение. Сопоставление «$v_1 \otimes \dots \otimes v_s$» $\mapsto \psi(v_1, \dots, v_s)$ 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 $V_1 \otimes \dots \otimes V_s = L/R$. 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 $(v_1, \dots, v_s) \in V_1 \times \dots \times V_s$ в пространстве $V_1 \otimes \dots \otimes V_s$ обозначается через $v_1 \otimes \dots \otimes v_s$ и называется **разложимым тензором**; для задания линейного отображения из $V_1 \otimes \dots \otimes V_s$ 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 $V_1, \dots, V_s$ — векторные пространства над полем k размерностей $n_1, \dots, n_s$; $\mathcal{B}_j = \{e_1^j, \dots, e_{n_j}^j\}$ — базис V_i для каждого $j = 1, \dots, s$. Тогда элементы вида $e_{i_1}^1 \otimes \dots \otimes e_{i_s}^s$, где $1 \leqslant i_k \leqslant n_k$ для всех $k = 1, \dots, s$, образуют базис пространства $V_1 \otimes \dots \otimes V_s$.*

Доказательство. Мы можем повторить доказательство предложения 11.2.3. А именно, рассмотрим векторное пространство W над k, базисом которого являются формальные символы вида $e_{i_1}^1 \otimes \dots \otimes e_{i_s}^s$. Определим полилинейное отображение $\varphi\colon V_1 \times \dots \times V_s \to W$ следующим образом: набор базисных векторов $(e_{i_1}^1, \dots, e_{i_s}^s) \in V_1 \times \dots \times V_s$ отправим в базисный элемент $e_{i_1}^1 \otimes \dots \otimes e_{i_s}^s$, 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 $v_j = \sum_{i_j=1}^{n_j} e^j_{i_j} a_{i_j,j}$. Положим $\varphi(v_1, \ldots, v_s) = \varphi(\sum_{i_1=1}^{n_1} e^1_{i_1} a_{i_1,1}, \ldots, \sum_{i_s=1}^{n_s} e^s_{i_s} a_{i_s,s}) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_s=1}^{n_s} a_{i_1,1} \ldots a_{i_s,s} \varphi(e^1_{i_1}, \ldots, e^s_{i_s}) = \sum_{i_1=1}^{n_1} \cdots \sum_{i_s=1}^{n_s} a_{i_1,1} \ldots a_{i_s,s} e^1_{i_1} \otimes \cdots \otimes e^s_{i_s}$. 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 $\psi\colon V_1 \times \cdots \times V_s \to U$ — 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 $e^1_{i_1} \otimes \cdots \otimes e^s_{i_s}$. Это можно сделать единственным образом: положим $\widetilde{\psi}(e^1_{i_1} \otimes \cdots \otimes e^s_{i_s}) = \psi(e^1_{i_1}, \ldots, e^s_{i_s})$. Композиция $\widetilde{\psi} \circ \varphi$, разумеется, является полилинейным отображением и совпадает с ψ на наборах вида $(e^1_{i_1}, \ldots, e^s_{i_s})$, и цепочка равенств выше показывает, что значение полилинейного отображения на произвольном наборе $(v_1, \ldots, v_s)$ 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 7.6). 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 $\mathcal{B} = (e_1, \ldots, e_n)$ — базис V. По универсальному свойству базиса (теорема 7.7.1) для задания элемента $\varphi \in V^* = \mathrm{Hom}(V, k)$ достаточно задать (произвольным образом) элементы $\varphi(e_1), \ldots, \varphi(e_n) \in k$.

Предложение 11.4.1. *Пусть V — векторное пространство над k с базисом $\mathcal{B} = (e_1, \ldots, e_n)$. 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 $(e_1^*, \ldots, e_n^*)$ — базис пространства V^*.*

Доказательство. Пусть $\varphi\colon V \to k$ — произвольный элемент пространства V^*. Мы знаем (теорема 7.7.1), что задать φ — это то же самое, что задать значения $\varphi(e_1), \ldots, \varphi(e_n) \in k$. Рассмотрим функцию $\varphi(e_1)e_1^* + \cdots + \varphi(e_n)e_n^*$. Покажем, что она совпадает с φ. Действительно, для базисного вектора e_i получаем $(\varphi(e_1)e_1^* + \cdots + \varphi(e_n)e_n^*)(e_i) = \varphi(e_1)e_1^*(e_i) + \cdots + \varphi(e_1)e_n^*(e_i) = \varphi(e_i)e_i^*(e_i) = \varphi(e_i)$. 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 (следствие 7.7.2). Например, имеется изоморфизм $V \to V^*$, отправляющий e_i в φ_i при $i = 1, \dots, n$, если $e_1, \dots, e_n$ — 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 $(v, e_1, e_2, \dots)$ 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 следствию теоремы о гомоморфизме (следствие 7.5.4) 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 $(f_1, \dots, f_m)$ в U и базис $(e_1, \dots, e_n)$ 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 $(\varphi_1, \dots, \varphi_m)$, двойственный базису $(f_1, \dots, f_m)$. Как мы знаем из раздела 7.8, 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 $(f_1, \dots, f_m)$ в пространстве U и базис $(e_1, \dots, e_n)$ 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*

$$\mathrm{Hom}(U \otimes W, V \otimes Z) \cong \mathrm{Hom}(U, V) \otimes \mathrm{Hom}(W, Z).$$

Доказательство. Мы построили отображение $\mathrm{Hom}(U,V)\times\mathrm{Hom}(W,Z)\to\mathrm{Hom}(U\otimes W,V\otimes Z)$, $(\varphi,\psi)\mapsto\varphi\otimes\psi$. По теореме 11.6.1 это сопоставление билинейно, поэтому определяет линейное отображение $\mathrm{Hom}(U,V)\otimes\mathrm{Hom}(W,Z)\to\mathrm{Hom}(U\otimes W,V\otimes Z)$, и обычные рассуждения (например, выбор базисов во всех указанных пространствах) убеждают нас, что получился изоморфизм. Еще один способ доказательства — воспользоваться уже доказанными изоморфизмами:

$$\mathrm{Hom}(U\otimes W,V\otimes Z)\cong(U\otimes W)^*\otimes(V\otimes Z)\cong(U^*\otimes V)\otimes(W^*\otimes Z)\cong\mathrm{Hom}(U,V)\otimes\mathrm{Hom}(W,Z).$$

□

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.
- $T_0^1(V) = V$ — векторы;
- $T_1^0(V) = V^*$ — ковекторы;
- $T_0^2(V) = V \otimes V = (V^* \otimes V^*)^* = \mathrm{Hom}(V^* \otimes V^*, k)$. 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.
- $T_2^0(V) = V^* \otimes V^* = (V \otimes V)^* = \mathrm{Hom}(V \otimes V, k)$. 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.
- $T_2^1(V) = V \otimes V^* \otimes V^* = (V \otimes V)^* \otimes V = \mathrm{Hom}(V \otimes V, V)$; 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:

$$\varphi = \sum \varphi_i e^i = \begin{pmatrix} \varphi_1 & \ldots & \varphi_n \end{pmatrix} \begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix}.$$

При этом в тензорном пространстве T^p_q (для произвольных p, q) возникает тензорный базис, состоящий из векторов вида $e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes \cdots \otimes ej_q$, где $1 \leqslant i_1, \ldots, i_p, j_1, \ldots, j_q \leqslant n$. Таким образом, каждый тензор $x \in T^p_q(V)$ можно единственным образом записать в виде

$$x = \sum_{\substack{i_1,\ldots,i_p \\ j_1,\ldots,j_q}} x^{i_1\ldots i_p}_{j_1\ldots j_q} e_{i_1} \otimes \cdots \otimes e_{i_p} \otimes e^{j_1} \otimes e^{j_q},$$

где $x^{i_1\ldots i_p}_{j_1\ldots j_q} \in k$ — 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 $\mathcal{F} = (f_1, \ldots, f_n)$, который будет называться *новым* (в отличие от *старого* базиса $\mathcal{E} = (e_1, \ldots, e_n)$). Напомним, что мы изучали, как связаны координаты векторов в этих базисах, с помощью [обратимой] матрицы перехода $C = (\mathcal{E} \rightsquigarrow \mathcal{F})$ (см. определение 6.6.1):

$$\begin{pmatrix} f_1 & \ldots & f_n \end{pmatrix} = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} \cdot C.$$

Вспомним, как преобразуются координаты вектора $v = \sum_i e_i v^i$ при замене базиса:

$$v = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} \cdot C \cdot C^{-1} \cdot \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix} = \begin{pmatrix} f_1 & \ldots & f_n \end{pmatrix} \cdot C^{-1} \begin{pmatrix} v^1 \\ \vdots \\ v^n \end{pmatrix}.$$

Таким образом, при переходе в новый базис столбец координат вектора умножается на C^{-1}. Это означает (см. замечание 6.6.4), 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

$$\begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} \cdot \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} = \begin{pmatrix} 1 & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & 1 \end{pmatrix} = E.$$

и аналогично для базиса $\mathcal{F}$. Домножим последнее равенство на C^{-1} слева и на C справа:

$$C^{-1} \begin{pmatrix} e^1 \\ \vdots \\ e^n \end{pmatrix} \cdot \begin{pmatrix} e_1 & \ldots & e_n \end{pmatrix} C = C^{-1} E C = E.$$

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