Конспект лекций по алгебре
Факультет математики и компьютерных наук
СПбГУ
потоки “Математика” и “Науки о данных” – 2022

А.В.Степанов

Оглавление

Глава 1. Введение 7
1. Определение основных алгебраических структур 7
2. Примеры 9

Глава 2. Векторные пространства 10
1. Основные определения 10
2. Матрицы 11
3. Другие определения базиса и его существование 14
4. Размерность пространства 15
5. Изоморфизм и классификация векторных пространств 16
6. Прямая сумма и прямое произведение 17
7. Замена базиса 18
8. Матрица линейного отображения 19
9. Размерность ядра и образа, прямая сумма, формула Грассмана 20
10. Факторпространство 21
11. Ранг, PDQ-разложение 22
12. Разложения Брюа и Гаусса 23

Глава 3. Начала теории групп 26
1. Простейшие конструкции 26
2. Гомоморфизмы, ядро и образ 27
3. Порождение, циклические группы, порядок элемента 27
4. Смежные классы и теорема Лагранжа 28
5. Факторгруппа и теоремы о гомоморфизме 29
6. Сопряженные элементы, коммутаторы и коммутант 30
7. Группа унитреугольных матриц 32
8. Приведенное разложение Брюа и второе доказательство разложения Гаусса 33
9. Симметрическая группа 33
10. Экспонента группы 34

Глава 4. Кольца и алгебры 36
1. Модули, алгебры, гомоморфизмы 36
2. Факторкольцо и теорема о гомоморфизме 37
3. Примеры идеалов и гомоморфизмов колец 38
4. Комплексные числа 38
5. Порождение 41
6. Прямая сумма алгебр 41
7. Евклидовы кольца 42
8. Китайская теорема об остатках 42
9. Простые и максимальные идеалы 43
10. Простые и неприводимые элементы 43
11. Нетеровы кольца и разложение на неприводимые 44
12. Факториальность колец главных идеалов 45
13. Наибольший общий делитель 46

14. Локализация 47
15. Поле частных и разложение на простейшие дроби 49
16. Многочлены от одной переменной 50
17. Формальная производная и кратность корня 51
18. Основная теорема алгебры 53
19. Вещественные многочлены 54
20. Лемма Гаусса 54
21. Экспонента мультипликативной группы кольца вычетов 56
22. О простых числах 58
23. Описание тестов на простоту 59

Глава 5. Определители 62
1. Полилинейные и антисимметричные формы. 62
2. Определение определителя 64
3. Свойства определителя 65
4. Формула для элементов обратной матрицы, формулы Крамера и минорный ранг 67

Глава 6. Собственные числа и жорданова форма 69
1. Собственные числа и вектора 69
2. Жорданова форма и теорема Гамильтона–Кэли 71
3. Другое доказательство жордановой формы 74
4. Разложение Жордана 75
5. Функции от матриц 77
6. Дифференциальные и рекуррентные уравнения 78
7. Свободные модули 79
8. Подмодули свободного модуля 79
9. Конечнопорожденные модули над кольцами главных идеалов 81
10. Фробениусова и жорданова форма 82
11. Единственность разложения 83

Глава 7. Билинейные и квадратичные формы 85
1. Формы и их матрицы 85
2. Диагонализация эрмитовой формы 86
3. Вещественные квадратичные формы 88
4. Пространства со скалярным произведением 90
5. Нормальные операторы 93
6. Матричные разложения 97
7. Гильбертово пространство 99
8. Кватернионы и движения трехмерного пространства 101
9. Теоремы Витта 104
10. Симплектические формы 107

Глава 8. Теория групп 108
1. Свободные группы, задание группы образующими и соотношениями 108
2. Подгруппы свободной группы 111
3. Действие группы на множестве и лемма Бернсайда 114
4. Классификация G-множеств 116
5. Несколько приложений действия группы на множестве 117
6. Теоремы Силова 118
7. Полупрямое произведение 120
8. Лемма о бабочке 122
9. Субнормальные ряды 123

10. Примеры простых групп 124
11. Разрешимые группы 125
12. Нильпотентные группы 126

Глава 9. Начала теории категорий 130
1. Категория, универсальные объекты, типы морфизмов 130
2. Функторы 133
3. Естественные преобразования 134
4. Эквалайзеры, произведения и универсальные квадраты 136
5. Сопряженные функторы 142
6. Универсальные алгебраические конструкции 147
7. Пределы 149
8. Представимые функторы и лемма Йонеды 154
9. Абелевы категории 155

Глава 10. Полилинейная алгебра 157
1. Простейшие свойства тензорного произведение 157
2. Тензоры 159
3. Тензорная алгебра модуля 161
4. Градуированные алгебры 161
5. Симметрическая и внешняя алгебры 163
6. Теорема Бине–Коши 165
7. Разложение определителя по группе столбцов 166
8. Миноры обратной матрицы 167
9. Грассманиан и соотношения Плюккера 168

Глава 11. Многочлены 170
1. Определения кольца многочленов 170
2. Нетеровость кольца многочленов 171
3. Порядок на множестве мономов 172
4. Деление с остатком и базисы Гребнера 173
5. Алгоритм Бухбергера 175
6. Минимальные и редуцированные базисы Гребнера 177
7. Симметрические многочлены 178
8. Степенные суммы 179
9. Результант 180
10. Дискриминант 184
11. Конечнопорожденные алгебры над полем 185
12. Теорема Гильберта о нулях 186
13. Начала алгебраической геометрии 188

Глава 12. Теория представлений 189
1. Основные определения 189
2. Строение артиновых колец 191
3. Лемма Шура 193
4. Полная приводимость 194
5. Характеры представления 195
6. Соотношения ортогональности 197
7. Разложение регулярного представления 199
8. Количество неприводимых представлений 199
9. Таблицы характеров 201
10. Таблица характеров как матрица перехода 202

11. Степени неприводимых представлений 203
12. Представления прямого произведения 204
13. Индуцированные представления 205

Глава 13. Теория Галуа 208
1. Расширения полей 208
2. Поле разложения 209
3. Существование и единственность конечных полей 210
4. Алгебраическое замыкание 210
5. Типы расширений 213
6. Группы автоморфизмов 215
7. Соответствие Галуа 217
8. Приложения теории Галуа 218
9. Норма и след 221

Глава 1

Введение

1. Определение основных алгебраических структур

Определение 1.1. Операцией называется функция $X_1 \times \cdots \times X_n \to X$. Чаще всего рассматривается ситуация, когда $X_1 = \cdots = X_n = X$. В этом случае операция называется n-арной операцией на множестве X. Декартово произведение пустого набора множеств по определению равно одноточечному множеству. Поэтому 0-арная операция на X – это выбор фиксированной точки множества X. 1-арная операция называется унарной, а 2-арная – бинарной. Бинарные операции обычно обозначаются не буквами, а значками, например $\star$, и вместо $\star(x, y)$ пишут $x \star y$.

Пусть X – множество, а $\star$ – бинарная операция на X. Рассмотрим следующие свойства.

(1) $\forall x, y, z \in X:\ (x \star y) \star z = x \star (y \star z)$ (ассоциативность).
(2) $\exists e \in X\ \forall x \in X:\ e \star x = x \star e = x$ (e называется нейтральным элементом).
(3) $\forall x \in X\ \exists x' \in X:\ x \star x' = x' \star x = e$ (x' называется элементом обратным к x). Если выполнено только одно из равенств $x' \star x = e$ или $x \star x' = e$, то x' называют левым или, соответственно, правым обратным к x.
(4) $\forall x, y \in X:\ x \star y = y \star x$ (коммутативность).

Определение 1.2. Множество X с операцией $\star$ называется

– полугруппой, если операция ассоциативна;
– моноидом, если операция ассоциативна и существует нейтральный элемент;
– группой, если выполнены свойства (1)–(3).

Полугруппа, моноид или группа называется коммутативной, если выполнено свойство (4). Коммутативную группу называют абелевой группой.

Элемент моноида называется обратимым, если для него существует (двусторонний) обратный.

Множество функций из X в Y обозначается $\mathrm{Map}(X, Y)$ или $\mathrm{Fun}(X, Y)$. Кроме хорошо известных из школы операций сложения и умножения, основным примером бинарной ассоциативной операции является операция композиции функций на множестве $\mathrm{Map}(X, X)$. Для функций $f, g : X \to X$ композиция определяется формулой $f \circ g(x) = f\big(g(x)\big)$. Нейтральным элементом относительно композиции является тождественная функция $\mathrm{id}(x) = x$. Таким образом, $\mathrm{Map}(X, X)$ является моноидом. Функция $f : X \to X$ обратима тогда и только тогда, когда она биективна, т. е.

$$\forall\, y \in X\ \exists!\, x \in X:\ f(x) = y.$$

Действительно, именно в этом случае $f\big(g(x)\big) = g\big(f(x)\big) = \mathrm{id}(x)$, где g определена по правилу $g(y) = x \iff f(x) = y$. Множество всех биективных функций из X в X является группой, которая обозначается S_X и называется симметрической группой множества X. Если $X = \{1, \ldots, n\}$, то вместо S_X пишут S_n и называют ее симметрической группой степени n или симметрической группой на n элементах.

Нейтральный элемент относительно операции умножения обычно обозначается символом 1, а относительно сложения – 0. Если из контекста неясно, нейтральным элементом какого множества является данный элемент, то пишут e_X, 1_X и 0_X для нейтрального элемента множества X относительно различных операций. Обратный к x элемент относительно сложения обозначается

через $-x$, относительно других операций – через x^{-1}. Сумма n экземпляров элемента x обозначается через nx, а произведение или другая операция – через x^n. По определению $0 \cdot x = 0$, x^0 – нейтральный элемент, $(-n)x = -(nx)$ и $x^{-n} = (x^n)^{-1}$.

Лемма 1.3. *Нейтральный элемент единственен (это утверждение не зависит даже от ассоциативности).*

Если элемента моноида имеет левый и правый обратный, то они совпадают. В частности, обратный элемент единственен.

Лемма 1.4. *Если в моноиде элементы x и y обратимы, то $x \star y$ обратим, причем $(x \star y)^{-1} = y^{-1} \star x^{-1}$. Множество обратимых элементов моноида является группой.*

Определение 1.5. Пусть теперь на множестве R заданы операции сложения и умножения, причем R является абелевой группой по сложению и полугруппой по умножению. Предположим, что выполнено следующее свойство:

5. $\forall x, y, z \in R : (x + y)z = xz + yz$ и $z(x + y) = zx + zy$ (правая и левая дистрибутивность).

Тогда R называется (ассоциативным) кольцом.

Если существует нейтральный элемент по умножению, то кольцо называется кольцом с единицей, если умножение коммутативно, то коммутативным кольцом.

Поле – это коммутативное кольцо с единицей, в котором каждый ненулевой элемент обратим.

Лемма 1.6. *Для любого элемента r произвольного кольца R: $0 \cdot r = r \cdot 0 = 0$.*

Если R – кольцо с единицей, то $(-1) \cdot r = -r$.

Как следует из леммы 1.4, множество обратимых (по умножению) элементов кольца R является группой. Эта группа называется мультипликативной подгруппой кольца и обозначается через $R^\times$.

Определение 1.7. Пусть V – абелева группа в аддитивной записи, F – поле, и задана операция (умножение) $V \times F \to V$. Предположим, что для любых $u, v \in V$ и $\alpha, \beta \in F$ выполнены следующие свойства:

(1) $v(\alpha\beta) = (v\alpha)\beta$;
(2) $v(\alpha + \beta) = v\alpha + v\beta$;
(3) $(u + v)\alpha = u\alpha + v\alpha$;
(4) $v \cdot 1 = v$.

Тогда V называется векторным пространством над полем F.

Определение 1.8. Пусть A – векторное пространство над полем F и, одновременно, кольцо с той же операцией сложения. Если выполнено свойство $(ab)\alpha = a(b\alpha) = (a\alpha)b$ для любых $a, b \in A$ и $\alpha \in F$, то A называется (ассоциативной) алгеброй над полем F.

Если отказаться от аксиомы ассоциативности кольцевого умножения, то получится неассоциативная алгебра. Изучение таких объектов в общем виде бесперспективно, даже если требовать конечномерность над полем. Однако если заменить ассоциативность какой-нибудь другой аксиомой, то получаются очень содержательные объекты. В частности, одной из важнейших алгебраической структур являются алгебры Ли, в которых ассоциативность заменена антисимметричностью и тождеством Якоби:

$$a \times a = 0; \qquad a \times (b \times c) + b \times (c \times a) + c \times (a \times b) = 0$$

для всех a, b, c из данной алгебры Ли.

Если в A есть нейтральный элемент относительно умножения (обозначим его символом e), то элементы $\alpha \in F$ отождествляются с элементами $e \cdot \alpha \in A$. Таким образом, если A – алгебра с единицей, то можно считать, что она содержит поле F. Обратно, если есть кольцо A, содержащее поле F, причем $x\alpha = \alpha x$ для любых $x \in A$ и $\alpha \in F$, то оно естественным образом является алгеброй над F (внешняя операция умножения в векторном пространстве $A \times F \to A$ является сужением операции умножения в кольце $A \times A \to A$).

2. Примеры

- $\mathbb{N} = \{1, 2, \dots\}$ – коммутативная полугруппа по сложению, коммутативный моноид по умножению.
- $\mathbb{N}_0 = \{0, 1, 2, \dots\}$ – коммутативный моноид и по сложению, и по умножению.
- $\mathbb{Z}$ – коммутативное кольцо.
- $\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$ – поля.
- $\mathrm{Map}(X, X)$ – (некоммутативный) моноид по отношению к композиции (коммутативный только в случае $|X| \leqslant 1$).
- S_X – (некоммутативная) группа по отношению к композиции (коммутативная только в случае $|X| \leqslant 2$).
- Если R – коммутативное кольцо с 1, то кольцо многочленов $R[t_1, \dots, t_n]$ также является коммутативным кольцом с 1.
- $\mathbb{Z}_n := \{0, 1, \dots, n-1\}$ с операциями сложения и умножения по модулю n, т. е. выполняется операция после чего берется остаток от деления на n, является коммутативным кольцом с 1 (обозначение $\mathbb{Z}_n$ – временное, позже мы обозначим это по-другому, а $\mathbb{Z}_n$ будет обозначать кольцо n-адических чисел). Кольцо $\mathbb{Z}_n$ является полем $\iff$ n – простое число.
- Множество векторов на плоскости или в 3-мерном пространстве, изучаемое в школьной геометрии, является векторным пространством над полем $\mathbb{R}$.
- Множество столбцов высоты n с элементами из поля F является векторным пространством над F, которое мы обозначаем F^n.
- Если V – абелева группа, в которой для некоторого простого числа p выполнено равенство $px = 0$ для любого $x \in V$, то V является векторным пространством над $\mathbb{Z}_p$ относительно естественных операций.
- Множество всех подмножеств данного множества X называется *булеаном* и обозначается через $2^X = \mathcal{P}(X)$. Оно является алгеброй над полем $\mathbb{Z}_2$ с операциями симметрической разности $Y \triangle Z := Y \cup Z \setminus Y \cap Z$ и пересечения.
- Множество $\mathbb{N}_0$ с операцией Ним-суммы является векторным пространством над $\mathbb{Z}_2$.

Глава 2

Векторные пространства

1. Основные определения

Далее в настоящей главе используются следующие обозначения и соглашения.

- F – поле.
- V – векторное пространство над F.
- F^n – множество столбцов высоты n над F.
- ${}^nF$ – множество строк длины n над F.
- Допуская вольность речи, элементы линейного пространства обычно называют векторами, а элементы поля F – числами;
- По умолчанию, греческие буквы обозначают числа, строчные латинские – элементы линейного пространства и столбцы, а прописные латинские – множества и линейные операторы и матрицы;
- словосочетание "почти все" означает "все, кроме конечного числа".

Подмножество $U \subseteq V$ называется подпространством, если оно само является векторным пространством относительно тех же операций, которые заданы в V.

Предложение 1.1 (критерий подпространства)**.** *Подмножество $U \subseteq V$ является подпространством в том и только том случае, если $x + y$, $x\alpha \in U$ для любых $x, y \in U$ и $\alpha \in F$.*

Пусть $u_1, \ldots, u_n \in V$, а $\alpha_1, \ldots, \alpha_n \in F$. Сумма

$$\sum_{k=1}^{n} u_k \alpha_k$$

называется линейной комбинацией векторов $u_1, \ldots, u_n \in V$ с коэффициентами $\alpha_1, \ldots, \alpha_n \in F$. Линейная комбинация называется тривиальной, если все ее коэффициенты равны нулю. Пусть $S \subseteq V$, и задан набор чисел $\alpha_s \in F$, $s \in S$. Если множество S бесконечно, то операция взятия бесконечной суммы $\sum\limits_{s\in S} s\alpha_s$ не определена. Однако, если почти все α_s равны 0, то в сумме остается только конечное число слагаемых. Таким образом, символ $\sum\limits_{s\in S} s\alpha_s$ будет употребляться в дальнейшем и для бесконечных множеств S при условии, что почти все α_s равны 0.

Линейной оболочкой набора S называется подпространство, порожденное S, т. е. наименьшее подпространство, содержащее S. Она обозначается через $\langle S \rangle$.

Предложение 1.2. $\langle S \rangle = \{ \sum\limits_{k=1}^{n} u_k \alpha_k \mid u_1, \ldots, u_n \in S,\ \alpha_1, \ldots, \alpha_n \in F \}$.

Если $\langle S \rangle = V$, то S называется системой образующих пространства V. Другими словами, S является системой образующих, если любой вектор выражается в виде линейной комбинации векторов из S.

Кортеж векторов $(u_1, \ldots, u_n)$ называется линейно независимым, если нетривиальная линейная комбинация этих векторов не равна нулю. Множество $S \subseteq V$ называется линейно независимым, если любой кортеж, составленный из конечного числа различных векторов из S, является линейно независимым. Другими словами, S линейно независимо, если для любого набора чисел $\alpha_s \in F$, почти все из которых равны нулю, из равенства $\sum\limits_{s\in S} s\alpha_s = 0$ следует, что все α_s равны нулю.

Определение 1.3. Базисом называется линейно независимая система образующих.

2. Матрицы

В дальнейшем мы будем широко использовать матричные обозначения, в частности для линейных комбинаций. Каждому элементу векторного пространства будет сопоставлен (возможно бесконечный) столбец (одномерный массив). Аналогично, линейному отображению будет сопоставлена матриц, т. е. двумерный массив. Сейчас мы введем операции на множестве матриц и укажем их простейшие свойства.

Формально, двумерный массив над множеством X – это функция $I \times J \to X$, где I и J – индексные множества. Обычно рассматриваются матрицы, строки и столбцы которых индексированы натуральными числами и являются элементами некоторого поля или кольца. Приведем сначала определение таких матриц и изучим их свойства, а потом рассмотрим более общий случай.

Определение 2.1. Двумерный массив $m \times n$ элементов поля F называется матрицей размера $m \times n$ над F. Множество всех таких матриц обозначается $\mathrm{M}_{m\times n}(F)$. Если $m = n$, то вместо $\mathrm{M}_{n\times n}(F)$ пишут $\mathrm{M}_n(F)$. Элемент матрицы A в позиции (i, j) (т. е. в i-й строке и j-м столбце) обычно обозначается через a_{ij}.

Для двух матриц одинакового размера их сумма определена поэлементно, т. е. $(A + B)_{ij} = a_{ij} + b_{ij}$. Также поэлементно определяется произведение матрицы на число: $(A\alpha)_{ij} = a_{ij}\alpha$.

Произведением матрицы $A \in \mathrm{M}_{m\times n}(F)$ на матрицу $B \in \mathrm{M}_{n\times k}(F)$ называется матрица $C = AB \in \mathrm{M}_{m\times k}(F)$ элементы которой вычисляются по формуле

$$c_{ij} = \sum_{h=1}^{n} a_{ih}b_{hj}.$$

В случае, когда количество столбцов левой матрицы не равно количеству строк правой, произведение матриц не определено.

Строка отождествляется с матрицей $1 \times n$, а столбец – с матрицей $n \times 1$. Таким образом, произведение строки длины n на столбец высоты n – это матрица 1×1, которая отождествляется с числом. Произведение же столбца на строку определено всегда, но является не числом, а матрицей соответствующего размера. Заметим, что произведение матриц некоммутативно, даже если размеры получившихся матриц равны.

Теорема 2.2. *Множество $\mathrm{M}_{m\times n}(F)$ с операциями сложения и умножения на число является векторным пространством над полем F.*

Произведение матриц ассоциативно, дистрибутивно и перестановочно с умножением на число, т. е. для любых матриц A, B, C и числа $\alpha \in F$, как только определены соответствующие операции, так

$$\begin{gathered}(AB)C = A(BC); \quad (AB)\alpha = A(B\alpha) = (A\alpha)B;\\ A(B + C) = AB + AC; \quad (B + C)A = BA + CA.\end{gathered} \tag{1}$$

Множество $\mathrm{M}_n(F)$ с операциями сложения и умножения является алгеброй с единицей над полем F.

Доказательство. Элемент произведения $(AB)C$ в позиции (i, j) равен

$$\sum_k \Big(\sum_h a_{ih}b_{hk}\Big)c_{kj} = \sum_k \sum_h (a_{ih}b_{hk})c_{kj}.$$

Элемент матрицы $A(BC)$ на соответствующем месте равен

$$\sum_h a_{ih}\Big(\sum_k b_{hk}c_{kj}\Big) = \sum_h \sum_k a_{ih}(b_{hk}c_{kj}).$$

Так как умножение в F ассоциативно, а сложение – ассоциативно и коммутативно, то эти элементы равны.

Обозначим через E квадратную матрицу с 1 на главной диагонали (с левого верхнего в правый нижний угол) и остальными нулями. Другими словами $e_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j \end{cases}$. Такая матрица называется единичной, ее размер обычно определяется из контекста, но при необходимости пишут E_n для обозначения единичной матрицы размера $n \times n$. Нетрудно вычислить, что при умножении данной (не обязательно квадратной) матрицы на единичную слева или справа она не меняется. В частности, E_n является мультипликативным нейтральным элементом в $\mathrm{M}_n(F)$. Остальные утверждения теоремы проверяются непосредственным вычислением. □

Заметим, что в алгебре матриц элементы поля F не принято отождествлять со скалярными матрицами $E\alpha$.

На самом деле мы хотим умножать двумерные массивы более общего вида. Предположим, что даны:

- множества X_{ij}, Y_{jh}, $i = 1, \dots, m$, $j = 1, \dots, n$, $h = 1, \dots, k$
- коммутативные моноиды Z_{ih} в аддитивной записи,
- и операции умножения $X_{ij} \times Y_{jh} \to Z_{ih}$.

Обозначим через X, Y и Z наборы множеств X_{ij}, Y_{jh} и Z_{ih}, соответственно, через $\mathrm{M}(X)$ – множество матриц A с элементами $a_{ij} \in X_{ij}$, и аналогично введем обозначения $\mathrm{M}(Y)$ и $\mathrm{M}(Z)$. Тогда можно определить произведение матриц $A \in \mathrm{M}(X)$ и $B \in \mathrm{M}(Y)$ как матрицу $C = AB \in \mathrm{M}(Z)$ с элементами $c_{ih} = \sum_{j=1}^{n} a_{ij} b_{jh}$.

Если все X_{ij} и Y_{jh} будут коммутативными моноидами, а операции умножения дистрибутивными, то умножение матриц также будет дистрибутивным. Если же для обобщенных матриц $A \in \mathrm{M}(X)$, $B \in \mathrm{M}(Y)$ и $C \in \mathrm{M}(W)$ определены произведения $(AB)C$ и $A(BC)$, а операции умножения будут ассоциативны, то $(AB)C = A(BC)$. Точную формулировку этих свойств приведем после того, как определим матрицы, строки и столбцы которых индексированы элементами произвольных множеств.

Одно из приложений описанной выше конструкции – произведение строки векторов на столбец чисел, что является просто другой записью линейной комбинации. В этих терминах линейная независимость кортежа векторов равносильна возможности сокращать на него. Действительно, если кортеж $v = (v_1, \dots, v_n)$ линейно независим, то для $a, b \in F^n$ имеет место: $va = vb \iff v(a - b) = 0 \iff a - b = 0 \iff a = b$. Очевидно, верно и обратное, т. е. из возможности сокращать следует линейная независимость.

Другое приложение – блочное произведение матриц (это очень легко показать на доске, и очень долго писать, поэтому пока не пишу).

Пусть теперь I и J – произвольные множества (возможно бесконечные), элементами которых мы будем индексировать строки и столбцы матриц. Предположим, что для каждого $i \in I$ и $j \in J$ задано множество X_{ij}, и обозначим этот набор множеств через X.[1] Тогда матрицей размера $I \times J$ над X называется функция $A : I \times J \to \cup X_{ij}$, $(i, j) \mapsto a_{ij}$, такая что $a_{ij} \in X_{ij}$. Множество всех матриц размера $I \times J$ над X обозначется через $\mathrm{M}_{I \times J}(X)$. Если $I = \{1\}$, то матрицы размера $I \times J$ будут называться строками длины J, а если $J = \{1\}$, то столбцами высоты I. Множества всех строк (столбцов) данной длины (соотв. высоты) будет обозначаться через ${}^J X$ (соотв. X^J).

[1] Формально надо говорить, что есть некоторый набор множеств, а X – это отображение из $I \times J$ в этот набор. Но нельзя говорить, что X – отображение из $I \times J$ в множество всех множеств, так как последнего не существует из за парадокса Рассела.

В дальнейшем мы будем рассматривать только ситуацию, когда все X_{ij} являются абелевыми группами в аддитивной записи. Тогда для двух матриц из одинакового размера их сумма определена поэлементно, т. е. $(A+B)_{ij} = a_{ij}+b_{ij}$. Если же все X_{ij} являются векторными пространствами над полем F, то также поэлементно определяется произведение матрицы на число: $(A\alpha)_{ij} = a_{ij}\alpha$.

Теперь, если мы хотим определить умножение матрицы из $\mathrm{M}_{I\times J}(X)$ на матрицу из $\mathrm{M}_{J\times H}(Y)$ нам недостаточно иметь операции умножения $X_{ij} \times Y_{jh} \to Z_{ih}$ как в первой части параграфа, потому что мы можем получить бесконечные суммы, которые не определены. Однако мы можем определить такую сумму, если почти все слагаемые будут равны нулю. Для того чтобы гарантировать последнее условие, мы постулируем, что все операции умножения $X_{ij} \times Y_{jh} \to Z_{ih}$ дистрибутивны, и в каждом столбце матрицы из $\mathrm{M}(Y)$ почти все элементы равны 0. Аналогично, мы могли бы потребовать, чтобы в каждой строке матрицы из $\mathrm{M}(X)$ почти все элементы были бы нулевыми, но в нашей системе обозначений именно условие на столбцы будет выполнено. Дистрибутивность умножения нужна для того, чтобы гарантировать равенство $a \cdot 0 = 0$.

Итак, обозначим через $\mathrm{M}^{c.f.}_{J\times H}(Y)$ подмножество в $\mathrm{M}_{J\times H}(Y)$, состоящее из всех матриц B, у которых для любого фиксированного $h \in H$ почти все элементы b_{jh} равны 0.[2]

Определение 2.3. Предположим, что для любых $i \in I$, $j \in J$, $h \in H$ заданы операции умножения $X_{ij} \times Y_{jh} \to Z_{ih}$, причем для любых $x, x' \in X_{ij}$ и любых $y, y' \in Y_{jh}$ имеют место равенства $(x + x')y = xy + x'y$ и $x(y + y') = xy + xy'$. Из этих равенств легко вывести, что $x \cdot 0 = 0 \cdot y = 0$. Тогда определим произведение матриц $A \in \mathrm{M}_{I\times J}(X)$ и $B \in \mathrm{M}^{c.f.}_{J\times H}(Y)$ как матрицу $AB \in \mathrm{M}_{I\times H}(Z)$ с элементами

$$(AB)_{ih} = \sum_{j\in J} a_{ij}b_{jh}.$$

При этом суммы определены, так как почти все слагаемые равны нулю.

Аналогично определяется умножение матриц $A \in \mathrm{M}^{r.f.}_{I\times J}(X)$ и $B \in \mathrm{M}_{J\times H}(Y)$

Лемма 2.4. *Пусть $A \in \mathrm{M}_{I\times J}(X)$, $B \in \mathrm{M}_{J\times H}(Y)$, а $C \in \mathrm{M}_{H\times K}(Z)$. Предположим, что произведения $(AB)C$ и $A(BC)$ определены, все необходимые операции дистрибутивны, и ассоциативны, т. е. $x(yz) = (xy)z$ для любых $i \in I$, $j \in J$, $h \in H$, $k \in K$, $x \in X_{ij}$, $y \in Y_{jh}$ и $z \in Z_{hk}$. Тогда выполнены все обычные свойства умножения матриц* (1).

В случае, когда все X_{ij} – это одно и то же поле F будем писать $\mathrm{M}_{I\times J}(F)$ вместо $\mathrm{M}_{I\times J}(X)$. Если $I = J$, то пишем $\mathrm{M}_I(F)$ вместо $\mathrm{M}_{I\times I}(F)$. В стандартной ситуации, когда $I = \{1, \dots, n\}$ в обозначениях множеств матриц заменяем I на n. Например, $\mathrm{M}_{n\times m}(F)$ обозначает множество прямоугольных матриц с n строками и m столбцами, элементы которых берутся из поля F.

Определение 2.5. Множество обратимых элементов кольца $\mathrm{M}_n(F)$ называется полной линейной группой степени n над F и обозначается через $\mathrm{GL}_n(F)$.

Для множества $\mathrm{M}^{c.f.}_{I\times\{1\}}(F)$ введем специальное обозначение F^I_{fin} и будем называть его множеством финитных столбцов высоты I над F. Другими словами, F^I_{fin} – это множество финитных функций из I в F, т. е. тех функций, у которых почти все значения равны 0. Аналогично, положим ${}^JF_{fin} = \mathrm{M}^{r.f.}_{\{1\}\times J}(F)$.

Пусть V – векторное пространство над полем F, а V^* – множество линейных отображений $V \to F$ (функция $f : V \to F$ линейна, если $f(x+y) = f(x)+f(y)$ и $f(x\alpha) = f(x)\alpha$). Множество V^* является векторным пространством относительно поточечных операций. Определим умножение (чаще говорят "спаривание") $V^* \times V \to F$ по формуле $fv = f(v)$. Тогда определены операции умножения $(V^*)^m \times {}^nV \to \mathrm{M}_{m\times n}(F)$ и ${}^nV \times \mathrm{M}_n(F) \to {}^nV$. Так как все необходимы свойства

[2] c.f. – сокращение от "column finite", аналогично r.f. будет использоваться, как сокращение для "row finite".

операций выполнены, то

$$\left(\begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} (v_1, \dots, v_n)\right) \cdot \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix} = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix} \cdot \left((v_1, \dots, v_n) \begin{pmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \vdots & \vdots \\ a_{n1} & \dots & a_{nn} \end{pmatrix}\right).$$

Есть еще одна полезная унарная операция с матрицами – транспонирование.

Определение 2.6. Пусть $A \in \mathrm{M}_{I\times J}(F)$. Матрица $A^\intercal \in \mathrm{M}_{J\times I}(F)$ с элементами $(A^\intercal)_{ij} = a_{ji}$ называется транспонированной к A.

На начальном этапе освоения материала польза транспонирования состоит в том, что оно меняет порядок сомножителей.

Предложение 2.7. $(AB)^\intercal = B^\intercal A^\intercal$.

Смысл же этой операции будет ясен при изучении сопряженного пространства.

Из чисто полиграфических соображений (для экономии места) для обозначения столбца часто пишется строка со знаком транспонирования, например, $(a_1, \dots, a_n)^\intercal$.

3. Другие определения базиса и его существование

Теорема 3.1 (эквивалентные определения базиса)**.** *Следующие условия на подмножество v векторного пространства V эквивалентны.*

1. *v – линейно независимая система образующих.*
2. *v – максимальная линейно независимая система.*
3. *v – минимальная система образующих.*
4. *Любой элемент $x \in V$ представляется в виде линейной комбинации набора v, причем единственным образом.*

Доказательство. В обозначениях предыдущего параграфа линейная комбинация набора векторов $v \subseteq V$ может быть записана в виде $va = \sum\limits_{y\in v} ya_y$ для некоторого столбца $a \in F^v_{fin}$. Пусть v – базис пространства V, в частности набор v является системой образующих. По определению это означает, что для любого $x \in V$ существует $a \in F^v_{fin}$, такое что $x = va$. Условие же линейной независимости набора v равносильно единственности такого представления. Действительно, мы уже видели, что линейная независимость равносильна возможности сокращать равенство $va = vb$ на v. Это доказывает эквивалентность пунктов 1 и 4. Доказательство того, что остальные пункты эквивалентны первому, оставляется читателю в качестве упражнения. □

Определение 3.2. Пусть v – базис пространства V, $a \in F^v_{fin}$, а $x \in V$. Если $x = va$, то a называется столбцом координат вектора x в базисе v и обозначается через x_v.

Если $V = F^n$, а e – стандартный базис, т. е. базис, состоящий из столбцов единичной матрицы, то получаем $x = ex_e$. Легко видеть, что можно отождествить строку из столбцов с матрицей. Следовательно, $x = Ex_e = x_e$ (столбец координат столбца в стандартном базисе совпадает с ним самим). Если это формалистика пока непонятна, то просто разложите столбец в линейную комбинацию столбцов единичной матрицы и запишите коэффициенты этой линейной комбинации в столбец.

Теорема 3.3 (частный случай леммы Цорна)**.** *Пусть $\mathscr{A}$ – набор подмножеств множества W. Если для любого линейно упорядоченного по включению набора $\mathscr{L} \subseteq \mathscr{A}$ объединение $\bigcup\limits_{A\in\mathscr{L}} A$ принадлежит $\mathscr{A}$, то $\mathscr{A}$ содержит хотя бы один максимальный элемент.*

Теорема 3.4 (о существовании базиса). *Пусть $X \subseteq Y \subseteq V$, причем набор X линейно независим, а Y – система образующих. Тогда существует базис Z, содержащий X и содержащийся в Y.*

Доказательство. Пусть $\mathscr{A}$ – набор всех линейно независимых подмножеств Y, содержащих X. Этот набор не пуст, так как он содержит X. Пусть $\mathscr{L}$ – линейно упорядоченный поднабор в $\mathscr{A}$. Обозначим через S объединение всех множеств из $\mathscr{L}$. Так как любое подмножество $C \in \mathscr{L}$ лежит между X и Y, то этим свойством обладает и S. Возьмем произвольное конечное подмножество $\{v_1, \dots, v_n\} \subseteq S$. По определению объединения множеств для каждого $i = 1, \dots, n$ существует $B_i \in \mathscr{L}$, содержащее вектор v_i. Так как набор $\mathscr{L}$ линейно упорядочен, среди множеств $B_1, \dots, B_n$ найдется наибольшее, скажем, B_k. Тогда $v_1, \dots, v_n \in B_k$, а так как B_k линейно независимо, то и множество $\{v_1, \dots, v_n\}$ линейно независимо. Следовательно, S линейно независимо, откуда $S \in \mathscr{A}$. По лемме Цорна заключаем, что $\mathscr{A}$ содержит максимальный элемент. Обозначим его через Z. Таким образом, Z – максимальное из линейно независимых подмножеств Y, содержащих X.

Пусть $y \in Y \setminus Z$. По максимальности Z множество $Z \cup \{y\}$ линейно зависимо, т. е. существуют $a \in F_{fin}^Z$ и $a_y \in F$ такие, что $y\alpha_y + Za = 0$. Коэффициент α_y не может быть равен нулю, это противоречило бы линейной независимости множества Z. Следовательно, y принадлежит линейной оболочке множества Z. Поэтому все множество Y содержится в $\langle Z \rangle$. С другой стороны, $V = \langle Y \rangle$ – наименьшее подпространство, содержащее Y. Следовательно, $\langle Z \rangle \supseteq V$, т. е. Z – система образующих, а значит и базис. □

4. Размерность пространства

Лемма 4.1 (о замене). *Пусть $u = \{u_1, \dots, u_n\}$ – линейно независимый набор, а v – система образующих пространства V. Тогда существуют различные элементы $v_1, \dots, v_n \in v$ такие, что множество $w = v \setminus \{v_1, \dots, v_n\} \cup u$ также является системой образующих.*

При этом, если набор v линейно независим, то w обладает тем же свойством, причем $(v \setminus \{v_1, \dots, v_n\}) \cap u = \varnothing$.

Доказательство. Проведем доказательство индукцией по n. В качестве базы индукции можно взять случай $n = 0$, который тривиален. Пусть теперь $n > 0$. По индукционному предположению найдутся вектора $v_1, \dots, v_{n-1} \in v$ такие, что $w' = v \setminus \{v_1, \dots, v_{n-1}\} \cup \{u_1, \dots, u_{n-1}\}$ является системой образующих. Причем, если v был линейно независимым, то w' – базис, а объединение дизъюнктное. Вектор u_n выражается через линейную комбинацию набора w', скажем $u_n = \sum_{i=1}^{n-1} u_i\alpha_i + \sum_{j=1}^{m} w_j\beta_j$ для некоторых $\alpha_i, \beta_j \in F$ и $w_j \in v \setminus \{v_1, \dots, v_{n-1}\}$. Заметим, что хотя бы один из коэффициентов β_j не равен нулю, иначе это противоречило бы линейной независимости набора u. Без ограничения общности можно считать, что $\beta_m \neq 0$. Положим $v_n = w_m$. Тогда v_n выражается через линейную комбинацию набора $w = w' \setminus \{v_n\} \cup \{u_n\}$. Таким образом, $w' \subseteq \langle w \rangle$ и, следовательно, w является системой образующих.

Предположим, что набор v, а, следовательно, и w', линейно независим. Положим $w'' = w' \setminus \{v_n\}$ и рассмотрим линейную комбинацию $w''a + u_n\alpha$ набора w, где $a \in F_{fin}^{w''}$. Подставляя выражение для u_n и приравнивая эту комбинацию к нулю получим

$$0 = w''a + u_n\alpha = w''a + \sum_{i=1}^{n-1} u_i\alpha_i\alpha + \sum_{j=1}^{m} w_j\beta_j\alpha = w''b + v_n\beta_m\alpha,$$

где $b \in F_{fin}^{w''}$ (нетрудно понять, как он выражается через a, α_i, β_j и α). Если $\alpha \neq 0$, то $w''b + v_n\beta_m\alpha$ является нетривиальной линейной комбинацией набора $w'' \cup \{v_n\} = w'$. Равенство такой комбинации нулю противоречит линейной независимости w'. Поэтому $\alpha = 0$, откуда $w''a = 0$, и из линейной независимости $w'' \subseteq w'$ следует, что $a = 0$. Таким образом, все коэффициенты исходной линейной комбинации равны 0. Следовательно, набор w линейно независим, а $u_n \notin w''$. □

Теорема 4.2 (количество элементов в базисе)**.** *Любые два базиса пространства V равномощны.*

Доказательство. Приведем доказательство для случая, когда один из базисов конечен. Доказательство для бесконечномерного случая в каком-то смысле проще (если предполагать случай конечного базиса известным). Его можно найти в курсе лекций Николая Верещагина и Александра Шеня "Введение в теорию множеств", лекция 11, теорема 36 (в настоящий момент она доступна по ссылке `http://www.intuit.ru/studies/courses/1034/144/lecture/3994?page=4`).

Пусть v и $u = \{u_1, \dots, u_n\}$ – базисы пространства V. Не умаляя общности можно считать, что мощность множества v не меньше n. По лемме о замене существует подмножество $\{v_1, \dots, v_n\} \subseteq v$ такое, что $w = v \setminus \{v_1, \dots, v_n\} \cup \{u_1, \dots, u_n\}$ – базис, причем объединение дизъюнктное. Легко видеть, что $u \subseteq w$, а $|v| = |w|$ (сколько элементов выкинули, столько не лежащих в $v \setminus \{v_1, \dots, v_n\}$ и добавили). Однако, так как базис – это максимальная линейно независимая система, то один базис не может строго содержаться в другом. Поэтому $w = u$, откуда $|v| = n$. □

Определение 4.3. Размерностью пространства называется мощность (любого) базиса этого пространства.

Пространство называется конечномерным, если в нем существует конечный базис. В этом случае удобнее индексировать базисные вектора и координаты векторов в этом базисе натуральными числами. Поэтому для конечномерных пространств мы будем считать по умолчанию, что базис – это строка из ${}^nV$, а координаты вектора – столбец из F^n. Тогда равенство $x = vx_v$, где v – базис, а x_v – столбец координат, которое является фактически определением координат, по-прежнему будет выполнено. Другими словами, будем считать, что базис конечномерного пространства – это не множество, а кортеж векторов. При необходимости уточнить, какое из определений базиса имеется в виду, мы будем говорить что кортеж векторов – это упорядоченный базис.

Обратите внимание, что линейная независимость кортежа $(v_1, \dots, v_n)$ равносильна тому, что $v_i \neq v_j$ при всех $i \neq j$ и множество $\{v_1, \dots, v_n\}$ линейно независимо.

5. Изоморфизм и классификация векторных пространств

Определение 5.1. Пусть V и U – векторные пространства, а L – функция $V \to U$. Она называется линейным отображением, если для любых $x, y \in V$ и любого $\alpha \in F$

$$L(x + y) = L(x) + L(y) \text{ и } L(x\alpha) = L(x)\alpha.$$

Другими словами, линейное отображение – это гомоморфизм векторных пространств.

Изоморфизм – это биективное линейное отображение (биективный гомоморфизм).

Линейное отображение из пространства в самого себя обычно называют линейным оператором или эндоморфизмом, хотя некоторые авторы используют термин "оператор", как полный синоним термина "отображение". Отображение из пространства в основное поле часто (особенно в функциональном анализе) называют функционалом. В алгебре принято еще слово "форма" для отображений в основное поле. хотя словосочетание "линейная форма" не так распространено, как "квадратичная" или "билинейная форма", которые мы будем изучать в следующем семестре.

Ясно, что линейные отображения характеризуются тем, что сохраняют линейные комбинации векторов. Для строки векторов $v = (v_1, \dots, v_n)$ и отображения $L : V \to U$ положим $L\big(v\big) = \big(L(v_1), \dots, L(v_n)\big) \in {}^nU$. Если набор v бесконечный, то аналогичное обозначение формально выглядит следующим образом: строка $L(v) \in {}^vU$ задана равенством $L(v)_x = L(x)$ для каждого $x \in v$.

В этих обозначениях свойство линейности можно выразить следующей формулой:

$$L(va) = L(v)a, \text{ где } a \in F^v_{fin}.$$

Лемма 5.2. *Пусть V – векторное пространство над полем F, а v – базис V. Отображение $\varphi_v : V \to F^v_{fin}$, заданное равенством $\varphi_v(x) = x_v$, является изоморфизмом векторных пространств (обратное отображение $\psi_v : F^v_{fin} \to V$, $a \mapsto va$).*

Следствие 5.3 (классификация векторных пространств)**.** *Любое векторное пространство изоморфно пространству F^I_{fin} для некоторого множества I (мощность которого равна размерности пространства).*

Два пространства изоморфны между собой тогда и только тогда, когда их размерности равны.

В каждом поле F есть самое маленькое подполе, которое называется простым подполем. Чтобы отличать натуральные числа от элементов поля обозначим через e нейтральный элемент по умножению поля F. Предположим, что $m \cdot e = 0$ для некоторого натурального m. Наменьшее натуральное p, для которого $p \cdot e = 0$ – простое. Тогда подмножество $\{0, e, 2 \cdot e, \ldots, (p-1) \cdot e\} \cong \mathbb{Z}_p$ является полем. Если же такого m не существует, то все элементы $n \cdot e$ различны. Любое подполе должно содержать все элементы $\frac{a \cdot e}{b \cdot e}$. Отображение $\mathbb{Q} \to F$, $\frac{a}{b} \mapsto \frac{a \cdot e}{b \cdot e}$, сохраняет операции и инъективно. Следовательно, его образ является полем, изморфным $\mathbb{Q}$.

Если K – подполе в F (говорят еще, что F – расширение поля K), то F является векторным пространством над K.

Следствие 5.4. *Количество элементов конечного поля F равно степени простого числа.*

Доказательство. Так как поле F конечно, то простое подполе K в F состоит из p элементов, где p простое. F – векторное пространство K, изоморфное (как векторное пространство) K^n. Тогда $|F| = |K^n| = p^n$. □

Теорема 5.5. *Для любого простого p и натуральноо n существует поле из p^n элементов. Любые 2 конечных поля из одинакового количества элементов изоморфны между собой.*

Конечное поле из q элементов обозначается $\mathbb{F}_q$. Как мы видели, q является степенью простого числа. Если q простое, $\mathbb{F}_q \cong \mathbb{Z}_q$. Если же q не простое, то это неверно, так как $\mathbb{Z}_q$ не является полем.

6. Прямая сумма и прямое произведение

Пусть U и V – подпространства векторного пространства W над полем F. Суммой $U+V$ называется совокупность всевозможных векторов вида $x+y$, где $x \in U$, $y \in V$. Сумма подпространств есть подпространство. Пересечение подпространств является подпространством.

Определение 6.1. Пространство W называется внутренней прямой суммой своих подпространств U и V, если каждый элемент $z \in W$ может быть единственным способом представлен в виде суммы $z = x+y$, где $x \in U$, а $y \in V$. Эквивалентная формулировка: $W = U+V$ и $U \cap V = \{0\}$.

Пусть теперь U и V – произвольные векторные пространства. Их (внешней) прямой суммой называется их декартово произведение с покомпонентными операциями.

Обозначение и внешней и внутренней прямой суммы $W = U \oplus V$.

Пространства U и V естественно вкладываются в их внешнюю прямую сумму: $U \ni x \mapsto (x, 0)$, а $V \ni y \mapsto (0, y)$. Если отождествить U и V с их образами, то внешняя прямая сумма превращается в прямую сумму подпространств.

Предложение 6.2. *Пусть $U, V \leqslant W$, а $U \oplus V$ обозначает их внешнюю прямую сумму. Зададим отображение $\varphi : U \oplus V \to W$ формулой $\varphi(x, y) = x + y$. Тогда отображение φ – изоморфизм тогда и только тогда, когда W является внутренней прямой суммой подпространств U и V.*

Предложение 6.3. *Если $W = U \oplus V$, то объединение базисов подпространств U и V есть базис пространства W. Поэтому* $\dim(U \oplus V) = \dim(U) + \dim(V)$.

Предложение 6.4. *Для любого подпространства $U \leqslant W$ существует подпространство $V \leqslant W$ такое, что $W = U \oplus V$.*

Доказательство. Выберем базис u подпространства U и дополним его до базиса $u \cup v$ всего пространства. Если $V = \langle v \rangle$, то легко проверить, что $W = U \oplus V$. □

Прямая сумма конечного количества подпространств определяется по индукции. Так как $(U \oplus V) \oplus W$ естественно изоморфно $U \oplus (V \oplus W)$ (элемент $((x, y), z)$ отождествляется с $(x, (y, z))$) то в расстановка скобок не играет роли (говорят, что прямая сумма ассоциативна, но надо понимать, что речь не идет об операции на множестве элементов векторного пространства, а об операции на *классе* векторных пространств).

Предложение 6.5. *Для подпространств $U_1, \ldots, U_n \leqslant V$ следующие условия эквивалентны.*

1. *Отображение $U_1 \oplus \cdots \oplus U_n \to V$, $(x_1, \ldots, x_n) \mapsto x_1 + \cdots + x_n$, является изоморфизмом.*
2. *Любой элемент $x \in V$ единственным образом раскладывается в сумму элементов $x_1 \in U_1, \ldots, x_n \in U_n$.*
3. *$V = U_1 + \cdots + U_n$ и $U_i \cap (\sum\limits_{j \neq i} U_j) = \{0\}$ для любого $i = 1, \ldots, n$.*
4. *Объединение базисов подпространств $U_1, \ldots, U_n$ является базисом пространства V.*

Если выполнены условия последнего предложения, то говорят, что V является прямой суммой подпространств $U_1, \ldots . U_n$.

Определение 6.6. Пусть I некоторое (возможно бесконечное) множество, и для каждого $i \in I$ задано пространство U_i. Прямой суммой $\bigoplus\limits_{i \in I} U_i$ называется множество финитных функций $f : I \to \bigcup\limits_{i \in I} U_i$ таких, что $f(i) \in U_i$. Неформально это называется множество потенциально бесконечных последовательностей элементов из U_i.

В отличие от прямой суммы, прямым произведением $\prod\limits_{i \in I} U_i$ называется множество *всех* функций $f : I \to \bigcup\limits_{i \in I} U_i$ таких, что $f(i) \in U_i$ (множество реально бесконечных последовательностей).

Названия “прямая сумма” и “прямое произведение” мотивированы универсальными свойствами, которыми обладают эти объекты. Мы изучим их позже в более общем виде в главе о теории категорий.

7. Замена базиса

В дальнейшем мы формулируем все утверждения для конечномерных пространств и для базисов, индексированных натуральными числами. Формулировка и доказательство бесконечномерных версий оставляется читателю в качестве упражнения.

Лемма 7.1. *Пусть $v \in {}^nV$, а $A \in \mathrm{GL}_n(F)$. Если v линейно независим, то и vA линейно независим. Линейные оболочки v и vA равны.*

Доказательство. Пусть $b \in F^n$. Если $(vA)b = 0$, то $v(Ab) = 0$ и, за счет линейной независимости v получаем $Ab = 0$. Домножая слева на A^{-1} получаем $b = 0$, откуда кортеж vA линейно независим.

Все вектора из vA являются линейными комбинациями векторов из v, т. е. лежат в линейной оболочке v. Следовательно, $\langle vA \rangle \subseteq \langle v \rangle$. Обратное включение следует из равенства $v = (vA)A^{-1}$. □

Предложение 7.2. *Пусть v – базис n-мерного пространства V над полем F. Набор $u \in {}^nV$ является базисом тогда и только тогда, когда существует $A \in \mathrm{GL}_n(F)$ такая, что $u =$*

vA (в бесконечномерной версии матрица A должна быть конечностолбцовой, т. е. почти все элементы каждого столбца должны быть равны 0).

Если u и v – базисы, то A называется матрицей перехода от v к u и обозначается через $C_{v\to u}$. При этом:

(1) $(C_{v\to u})_{*k} = (u_k)_v$. *Столбец матрицы $C_{v\to u}$ с номером k равен столбцу координат вектора u_k в базисе v. Одной формулой:*
(2) $C_{v\to u}^{-1} = C_{u\to v}$.
(3) *Если матрица двусторонне обратима, то она квадратная.*

Если $V = F^n$, а e – стандартный базис, то $C_{e\to u}$ матрица, составленная из столбцов базиса u. Проще всего это запомнить, написав определение матрицы перехода в виде $eC_{e\to u} = u$ и отождествив строчки из столбцов с матрицами.

Предложение 7.3 (Преобразование координат при замене базиса)**.** *Пусть v и u – базисы пространства V. Для $x \in V$ имеет место равенство $x_v = C_{v\to u}x_u$.*

Доказательство. По определению столбца координат $x = vx_v = ux_u$. Заменяя u на $vC_{v\to u}$, получим $vx_v = vC_{v\to u}x_u$. Пользуясь ассоциативностью произведения матриц и сокращая на v (это можно сделать, так как v линейно независим), получаем требуемое равенство. □

8. Матрица линейного отображения

Предложение 8.1. *Пусть $L : U \to V$ – линейное отображение, $u = (u_1, \dots, u_n)$ – базис U, а $v = (v_1, \dots, v_m)$ – базис V. Существует единственная матрица $A \in \mathrm{M}_{m\times n}(F)$ такая, что для любого $x \in U$ имеет место равенство $L(x)_v = Ax_u$. Столбцы матрицы A вычисляются по формуле $a_{*k} = L(u_k)_v$.*

Доказательство. По определению столбца координат $x = ux_u$. Применяя к этому равенству отображения L и φ_v из леммы 5.2 и пользуясь их линейностью, получаем

$$L(x)_v = \varphi_v \circ L(x) = \varphi_v \circ L(ux_u) = \varphi_v\big(L(u)x_u\big) = \varphi_v\big(L(u)\big)x_u.$$

Таким образом, положив $A = \varphi_v\big(L(u)\big) = \big(L(u_1)_v, \dots, L(u_n)_v\big)$ получаем существование матрицы A и формулу для ее столбцов. Единственность следует из несложного утверждения про матрицы: Если $Ac = Bc$ для любого столбца c соответствующей высоты, то матрицы A и B равны. □

Матрица A из последнего предложения называется матрицей отображения L в базисах u и v и обозначается через L_v^u. В случае, когда $U = V$, а $u = v$, говорят о матрице оператора L в базисе u и обозначают ее через L_u. Таким образом, для любого $x \in U$ имеют место равенства

$$L(x)_v = L_v^u x_u \text{ или } L(x)_u = L_u x_u \text{ в случае } U = V,\ u = v.$$

Пусть A – матрица размера $n \times m$, а $L : F^m \to F^n$ – умножение на матрицу A, т. е. $L(x) = Ax$ для всех $x \in F^m$. Если e обозначает стандартный базис (как в F^m, так и в F^n), то легко видеть, что $L_e^e = A$. В связи с этим естественным отождествлением матрицы и оператора умножения на нее, последний будет часто обозначаться той же буквой, что и сама матрица.

Следующее утверждение является причиной именно такого определения произведения матриц. С другой стороны, его доказательство непосредственно вытекает из ассоциативности произведения матриц.

Предложение 8.2. *Матрица композиции линейных операторов является произведением матриц этих операторов. Точнее, если U, V и W – конечномерные линейные пространства с базисами u, v и w, соответственно, а $L : U \to V$ и $M : V \to W$ – линейные отображения, то $(M \circ L)_w^u = M_w^v L_v^u$. В частности, при $U = V = W$ и $u = v = w$ получаем $(M \circ L)_u = M_u L_u$.*

Нетрудно проверить, что множество линейных отображений $V \to V$ с операциями поточечного сложения, композиции и умножения на число является алгеброй с единицей. Эта алгебра обычно обозначается $\mathrm{End}(V)$ и называется кольцом эндоморфизмов пространства V.

Следствие 8.3. *Пусть v – базис n-мерного пространства V. Определим отображение $\theta_v : \mathrm{End}(V) \to \mathrm{M}_n(F)$ формулой $\theta_v(L) = L_v$. Тогда θ_v является изоморфизмом алгебр.*

Предложение 8.4. *Пусть u и u' – базисы пространства U, v и v' – базисы пространства V, а $L : U \to V$ – линейное отображение. Тогда*

$$L_{v'}^{u'} = C_{v' \to v} L_v^u C_{u \to u'}.$$

В частности, при $U = V$, $u = v$ и $u' = v'$ получим

$$L_{u'} = C_{u' \to u} L_u C_{u \to u'}.$$

Нетрудно видеть, что матрица перехода $C_{u \to v}$ между двумя различными базисами u и v пространства V совпадает с матрицей тождественного отображения id_V в базисах v и u.

Лемма 8.5. *Пусть $u = (u_1, \ldots, u_n)$ – базис пространства U, а $v = (v_1, \ldots, v_n) \in {}^nV$ – произвольный набор векторов пространства V. Тогда существует единственное линейное отображение $L : U \to V$ такое, что $L(u) = v$. При этом L инъективно тогда и только тогда, когда v линейно независим; L сюръективно тогда и только тогда, когда v – система образующих; L – изоморфизм тогда и только тогда, когда v – базис.*

Доказательство. Для любого $x \in U$ имеем $x = ux_u$. Тогда $L(x) = L(u)x_u$ для любого линейного отображения L. Таким образом, линейное отображение L удовлетворяющее равенству $L(u) = v$, должно быть задано формулой $L(x) = vx_u$. С другой стороны, несложно проверить, что отображение, заданное такой формулой, линейно.

Остальные утверждения леммы вытекают непосредственно из определений. □

Замечание 8.6. Пусть u и v – базисы пространства V. Тогда матрица отображения L из предыдущего предложения в базисе u совпадает с матрицей перехода $C_{u \to v}$.

9. Размерность ядра и образа, прямая сумма, формула Грассмана

Образом функции $f : X \to Y$ называется множество $\mathrm{Im}\, f = \{f(x) \mid x \in X\}$. По-другому образ обозначается через $f(X)$, что совпадает с общим правилом действия функции на множестве: для $A \subseteq X$ положим $f(A) = \{f(x) \mid x \in A\}$.

Полным прообразом подмножества $B \subseteq Y$ называется множество $f^{-1}(B) = \{x \in X \mid f(x) \in B\}$. Если $B = \{y\}$ – одноточечное множество, то говорят о полном прообразе точки y, который еще называется слоем отображения над этой точкой. В случае биективной функции f обозначение $f^{-1}(y)$ двузначно: оно может обозначать значение обратной функции $f^{-1} : Y \to X$ в точке y, а может обозначать полный прообраз точки y, т. е. одноэлементное множество. Несмотря на то, что одноэлементное множество формально нельзя отождествлять с его единственным элементом, это довольно часто делается. При этом отождествлении двузначность обозначения $f^{-1}(y)$ пропадает.

Определение 9.1. Пусть $L : U \to V$ – линейное отображение. Тогда

$$\mathrm{Ker}\, L = L^{-1}(0) := \{x \in U \mid L(x) = 0\} \text{ – ядро отображения } L;$$
$$\mathrm{Im}\, L = \{L(x) \mid x \in U\} \text{ – образ отображения } L.$$

Предложение 9.2. *Ядро и образ линейного отображения $L : U \to V$ являются подпространствами.*

Все слои отображения L являются сдвигами ядра. Точнее, пусть $L(x) = y$, где $x \in U$. Тогда

$$L^{-1}(y) = x + \mathrm{Ker}\, L.$$

В частности, L инъективно тогда и только тогда, когда $\mathrm{Ker}\, L = \{0\}$.

Лемма 9.3. *Пусть $L : U \to V$ – линейное отображение, а u – базис ядра L. Дополним u до базиса $u \cup w$ пространства U. Тогда $L(w)$ является базисом подпространства $\mathrm{Im}\, L$.*

Доказательство. $y \in \operatorname{Im} L \iff y = L(x) = L(ua + wb) = L(u)a + L(w)b = L(w)b$, следовательно, $L(w)$ порождает $\operatorname{Im} L$.

$0 = L(w)b = L(wb) \iff wb \in \operatorname{Ker} L \iff wb = ua$. Так как набор $u \cup w$ линейно независим, то $b = 0$, откуда вытекает линейная независимость $L(w)$. □

Теорема 9.4 (о размерности ядра и образа). *Пусть $L : U \to V$ – линейное отображение. Тогда* $\dim \operatorname{Im} L + \dim \operatorname{Ker} L = \dim U$.

Следствие 9.5. *Если $\dim U = \dim V$, то инъективность L равносильна сюръективности.*

Теорема 9.6 (формула Грассмана). *Если U и V – подпространства линейного пространства W, то*

$$\dim U + \dim V = \dim(U + V) + \dim(U \cap V).$$

Доказательство. Зададим линейное отображение L из внешней прямой суммы $U \oplus V$ в W формулой $L(u, v) = u + v$. Легко проверить, что $\operatorname{Im} L = U + V$, а $\operatorname{Ker} L = \{(u, -u) \mid u \in U \cap V\} \cong U \cap V$. Теперь теорема следует из теоремы о размерности ядра и образа. □

10. Факторпространство

Задача этого параграфа – построить линейное отображение из данного пространства V с данным ядром U. Из предложения 9.2 мы знаем, что все слои линейного отображения – сдвиги ядра, т. е. точки образа находятся во взаимно однозначном соответствии с аффинными подпространствами $x+U$, где x пробегает V. Это наталкивает на мысль, определить область значений нашего отображения, как множество этих аффинных подпространств. Аффинное подпространство $x+U$ называют еще смежным классом V по U. Заметим, что $y \in x+U \iff y-x \in U$. Можно определить отношение эквивалентности $y \sim_U x \iff y - x \in U$ (проверьте, что это эквивалентность). Тогда смежный класс, это в точности класс эквивалентности.

Определение 10.1. Множество смежных классов V по U с операциями:

$$(x + U) + (y + U) = (x + y) + U,$$
$$(x + U)\alpha = x\alpha + U$$

называется факторпространством V по U и обозначается V/U.

Определения операций в V/U зависят от выбора представителя смежного класса, однако легко показать, что результат не зависит от этого выбора (это тот случай, когда необходимо доказывать корректность определения; как и все непосредственные проверки, это доказательство оставлено читателю в качестве упражнения).

Обозначим через $\pi_U : V \to V/U$ естественную проекцию: $\pi_U(x) = x + U$. Нетрудно проверить, что это отображение линейно, сюръективно, а его ядро равно U. По теореме о размерности ядра и образа получим

$$\dim V/U = \dim V - \dim U.$$

Факторпространство обладает следующим универсальным свойством.

Предложение 10.2. *Пусть $U \leqslant V$. Для любого линейного отображения $L : V \to W$, ядро которого содержит U, существует единственное отображение $\widetilde{L} : V/U \to W$ такое, что $L = \widetilde{L} \circ \pi_U$. При этом, сюръективность $\widetilde{L}$ равносильна сюръективности L, а инъективность $\widetilde{L}$ – тому, что* $\operatorname{Ker} L = U$.

Доказательство. Положим $\widetilde{L}(x+U) = L(x)$. Легко проверить, что эта формула не зависит от выбора представителя смежного класса и задает линейное отображение. С другой стороны, эта формула равносильна равенству $L = \widetilde{L} \circ \pi_U$. Следовательно, $\widetilde{L}$ существует и единственно.

Утверждение о сюръективности сразу следует из сюръективности π_U и равенства $L = \widetilde{L} \circ \pi_U$. Отображение $\widetilde{L}$ инъективно $\iff \operatorname{Ker} \widetilde{L} = \{0 + U\}$ (здесь $0 + U$ – нулевой элемент пространства V/U).

$$x + U \in \operatorname{Ker} \widetilde{L} \iff \widetilde{L}(x + U) = 0 \iff L(x) = 0 \iff x \in \operatorname{Ker} L.$$

Таким образом, если $\operatorname{Ker} L = U$, то $\operatorname{Ker} \widetilde{L} = \{0 + U\}$, и наоборот. □

Следствие 10.3 (теорема о гомоморфизме)**.** *Для любого линейного отображения $L : V \to W$ имеем*

$$V/\operatorname{Ker} L \cong \operatorname{Im} L.$$

Пусть теперь $L : V \to V$ линейный оператор, а $U \leqslant V$. Подпространство U называется инвариантным относительно L (короче, L-инвариантным), если $L(x) \in U$ для любого $x \in U$. Пусть, u – базис U, а $v = u \cup w$ – базис V. Если U является L-инвариантным, то матрица оператора L в базисе v имеет вид $\left(\begin{smallmatrix} A & * \\ 0 & B \end{smallmatrix}\right)$, где A имеет размер $\dim U \times \dim U$, а размер B равен $(\dim V - \dim U) \times (\dim V - \dim U)$. Из формулы для столбцов матрицы оператора видно, что A является матрицей сужения оператора L на подпространство U в базсе u.

Так как U является L-инвариантным, можно определить отображение $\overline{L} : V/U \to V/U$ формулой $\overline{L}(x + U) = L(x) + U$ (проверьте, что определение корректно). Обозначим через $\overline{w} = (w_1 + U, \ldots, w_k + U)$ базис факторпространства V/U (проверьте, что это базис). Тогда матрица B – это матрица оператора $\overline{L}$ в базисе $\overline{w}$.

11. Ранг, PDQ-разложение

Определение 11.1. Рангом набора векторов называется размерность линейной оболочки этого набора. Рангом линейного оператора называется размерность образа этого оператора. Столбцовым (строчным) рангом матрицы называется ранг набора ее столбцов (строк).

Так как из любой системы образующих можно выбрать базис, то ранг набора векторов – это наибольшее количество линейно независимых векторов из данного набора. Так как образы базисных векторов порождают образ оператора, то ранг оператора равен рангу набора, состоящего из образов базисных векторов. Легко понять, что он равен также столбцовому рангу матрицы этого оператора (независимо от выбора базисов). Далее в этом параграфе мы докажем, что строчный и столбцовый ранги матрицы равны.

Лемма 11.2. *Пусть $A \in \mathrm{M}_{m \times n}(F)$.*

1. *Набор столбцов матрицы A линейно независим тогда и только тогда, когда ее столбцовый ранг равен n.*
2. *Набор столбцов матрицы A порождает F^m тогда и только тогда, когда ее столбцовый ранг равен m.*
3. *Набор столбцов матрицы A является базисом в F^m тогда и только тогда, когда ее столбцовый ранг равен $m = n$. В этом случае матрица A обратима.*
4. *Если все строки A линейно независимы, и все ее столбцы обладают тем же свойством, то $m = n$, а матрица A обратима.*

Доказательство. Утверждения (1) и (2) очевидны. Из них следует, что столбцовый ранг равен $m = n$ тогда и только тогда, когда набор столбцов является базисом. В этом случае A является матрицей перехода от стандартного базиса пространства F^m к базису из столбцов матрицы A и, следовательно, является обратимой.

Количество линейно независимых столбцов не может быть больше размерности пространства, откуда $n \leqslant m$. Аналогичное рассуждение для строк доказывает обратное неравенство, т. е. $m = n$. □

Лемма 11.3. *Умножение матрицы на обратимую (слева или справа) не меняет ее столбцовый и строчный ранг.*

Доказательство. Столбцовый ранг матрицы оператора равен рангу оператора, следовательно, не зависит от выбора базиса. Пусть L – оператор умножения на матрицу $A \in \mathrm{M}_{m\times n}(F)$, а $B \in \mathrm{GL}_m(F)$ и $C \in \mathrm{GL}_n(F)$ – обратимые матрицы. Тогда $A = L_e^e$, а BAC – матрица того же оператора в других базисах пространств F^n и F^m. Таким образом, $\mathrm{c.rank}\, A = \mathrm{c.rank}\, BAC$.

Второе утверждение следует из того, что строчный ранг матрицы равен столбцовому рангу транспонированной к ней, а транспонированная к обратимой – обратима. □

Теорема 11.4 (PDQ-разложение). *Для любого линейного отображения $L : U \to V$, где U и V – конечномерные пространства, существуют базисы пространств U и V, в которых матрица отображения L имеет вид $\begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix}$.*

Любая матрица $A \in \mathrm{M}_{m\times n}(F)$ представляется в виде $A = PDQ$, где $P \in \mathrm{GL}_m(F)$, $Q \in \mathrm{GL}_n(F)$, а D записывается в блочном виде $D = \begin{pmatrix} E & 0 \\ 0 & 0 \end{pmatrix}$. При этом размер единичной матрицы в формуле для D равен строчному и столбцовому рангу A.

Доказательство. Дополним базис u ядра отображения L до базиса $w \cup u$ пространства U. По лемме 9.3 $L(w)$ является базисом $\mathrm{Im}\, L$. Дополним его до базиса $L(w) \cup v$ пространства V. Тогда матрица отображения L в базисах $w \cup u$ и $L(w) \cup v$ имеет требуемый вид.

Второе утверждение следует из первого, примененного к матрице оператора умножения на A.

Последнее предложение теоремы вытекает из предыдущей леммы. □

Аналогичная теорема верна с заменой поля на хорошее кольцо (область главных идеалов), а единичной матрицы на диагональную, впрочем, доказательство уже менее банально. Если в качестве кольца взять $\mathbb{Z}$, то это утверждение является ключевым шагом классификации конечнопорожденных абелевых групп.

Лемма 11.5. *Квадратная матрица обратима тогда только тогда, когда ее ранг равен ее размеру.*

Следующее тривиальное утверждение во многих книгах преподносится, как верх математической мысли (справедливости ради, надо сказать, что ранг матрицы в этих книгах определяется, как минорный ранг, см. § 4 главы 5). При нашем определении ранга оно сразу следует из того, что система линейных уравнений $Ax = b$ имеет решение тогда и только тогда, когда b содержится в линейной оболочке столбцов матрицы A.

Теорема 11.6 (Кронекера–Капелли). *Система $Ax = b$ совместна тогда и только тогда, когда ранг матрицы A равен рангу расширенной матрицы $(A\, b)$.*

12. Разложения Брюа и Гаусса

Матрица a называется верхней (нижней) треугольной, если $a_{ij} = 0$ при всех $i > j$ (соотв. $i < j$). Треугольная матрица с 1 на главной диагонали называется унитреугольной. Обозначим множество верхних (нижних) обратимых треугольных матриц через $B = B_n(F)$ (соотв. $B^- = B_n^-(F)$). Аналогично, множества унитреугольных матриц обозначаются через $U = U_n(F)$ и $U^- = U_n^-(F)$. В теории алгебраических групп B и B^- называются противоположными борелевскими подгруппами, а U и U^- – их унипотентными радикалами. Через $W = W_n$ обозначим множество матриц перестановок, т. е. матриц, отличающихся от единичной перестановкой столбцов (группа Вейля).

Лемма 12.1. *Множества W, B, B^-, U и U^- являются подгруппами в $\mathrm{GL}_n(F)$.*

Теорема 12.2 (Разложение Брюа). $\mathrm{GL}_n(F) = BWB$.

Доказательство. Требуется доказать, что любая матрица $a \in \mathrm{GL}_n(F)$ представляется в виде $a = bwc$, для некоторых $b, c \in B$ и $w \in W$. Индукцией по n докажем, что домножая a слева и

справа на верхнетреугольные матрицы можно получить матрицу перестановку. Так как обратная к верхнетреугольной является верхнетреугольной, из этого следует результат.

Пусть i – наибольший индекс, для которого $a_{i1} \neq 0$. Запишем a в виде

$$a = \begin{pmatrix} x & * \\ a_{i1} & z \\ 0 & * \end{pmatrix} \text{ где } x = \begin{pmatrix} a_{11} \\ \vdots \\ a_{i-1\,1} \end{pmatrix}, \text{ а } z = (a_{i2}, \ldots, a_{in}).$$

Домножая a слева на верхнетреугольную матрицу получим матрицу, у которой первый столбец совпадает с i-м столбцом единичной матрицы. После этого, домножая справа на подходящую верхнетреугольную матрицу можем сделать i-ю строку равной первой строке единичной матрицы. Точнее

$$\begin{pmatrix} E & -\frac{x}{a_{i1}} & 0 \\ 0 & \frac{1}{a_{i1}} & 0 \\ 0 & 0 & E \end{pmatrix} \begin{pmatrix} x & * \\ a_{i1} & z \\ 0 & * \end{pmatrix} \begin{pmatrix} 1 & -z/a_{i1} \\ 0 & E \end{pmatrix} = \begin{pmatrix} 0 & f \\ 1 & 0 \\ 0 & g \end{pmatrix}$$

для некоторых матриц f, g. Заметим, что, так как строки получившейся матрицы линейно независимы, то и строки матрицы $\left(\begin{smallmatrix} f \\ g \end{smallmatrix}\right)$ также являются линейно независимыми. Поэтому последняя матрица обратима, и к ней можно применить индукционное предположение.

Следовательно, существуют матрицы $u, v \in B_{n-1}(F)$ такие, что $u \left(\begin{smallmatrix} f \\ g \end{smallmatrix}\right) v \in W_{n-1}$. Пусть $u = \left(\begin{smallmatrix} u^{(1)} & u^{(2)} \\ 0 & u^{(3)} \end{smallmatrix}\right)$, где $u^{(1)} \in B_{i-1}(F)$, а $u^{(3)} \in B_{n-i}(F)$. Тогда

$$\begin{pmatrix} u^{(1)} & u^{(2)} \\ 0 & u^{(3)} \end{pmatrix} \begin{pmatrix} f \\ g \end{pmatrix} \cdot v$$

является матрицей-перестановкой, а, следовательно, и

$$\begin{pmatrix} u^{(1)} & 0 & u^{(2)} \\ 0 & 1 & 0 \\ 0 & 0 & u^{(3)} \end{pmatrix} \begin{pmatrix} 0 & f \\ 1 & 0 \\ 0 & g \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & v \end{pmatrix}$$

– также матрица-перестановка. □

Множество BwB при фиксированном $w \in W$ называется клеткой Брюа.

Упражнение 12.3. Две различные клетки Брюа не пересекаются.

Теорема 12.4 (Разложение Гаусса)**.** $\mathrm{GL}_n(F) = WB^-B$.

В параграфе 8 главы 3 мы докажем, что клетка Брюа BwB содержится в клетке Гаусса wB^-B. В частности, разложение Гаусса является следствием разложения Брюа. Здесь же мы дадим непосредственное доказательство разложения Гаусса. Оно сразу вытекает из следующих трех лемм, которые нам все равно понадобятся в дальнейшем. Сформулируем в двух словах основную идею. Сначала надо выбрать перестановку строк так, чтобы все квадратные подматрицы (любого размера), стоящие в левом верхнем углу, были бы обратимы. При доказательстве возможности этого ключевым замечанием является то, что ранг подматрицы, составленной из первых k столбцов матрицы a, равен k (все столбцы обратимой матрицы линейно независимы), поэтому существует k линейно независимых строк этой подматрицы. Затем матрица приводится к верхнетреугольному виду элементарными преобразованиями с ведущими элементами на главной диагонали.

Главной подматрицей матрицы a порядка k называется подматрица, стоящая на пересечении первых k строк и первых k столбов матрицы a.

Лемма 12.5. *Умножение матрицы на обратимую нижнетреугольную слева и на обратимую верхнетреугольную справа не меняет обратимости главных подматриц.*

Лемма 12.6. *Все главные подматрицы обратимы тогда и только тогда, когда матрица раскладывается в произведение обратимых нижнетреугольной и верхнетреугольной.*

Доказательство. Если матрица является произведением обратимых нижнетреугольной на верхнетреугольную, то из предыдущей леммы следует, что все главные подматрицы обратимы.

Доказательство обратной импликации проведем индукцией по размеру n матрицы a. Для $n = 1$ доказывать нечего. Пусть $a = \left(\begin{smallmatrix} \widetilde{a} & x \\ y & \alpha \end{smallmatrix}\right)$, где $\alpha \in F$, $x \in F^{n-1}$, $y \in {}^{n-1}F$, а $\widetilde{a} \in \mathrm{M}_{n-1}(F)$. Заметим, что по условию все главные подматрицы в $\widetilde{a}$ обратимы. Поэтому к $\widetilde{a}$ применимо индукционное предположение. Запишем $\widetilde{a} = bc$, где $b \in B^-$, $c \in B$. Тогда

$$a = \begin{pmatrix} E & 0 \\ y\widetilde{a}^{-1} & 1 \end{pmatrix} \begin{pmatrix} \widetilde{a} & x \\ 0 & \alpha - y\widetilde{a}^{-1}x \end{pmatrix} = \begin{pmatrix} b & 0 \\ yc^{-1} & 1 \end{pmatrix} \begin{pmatrix} c & b^{-1}x \\ 0 & \alpha - y\widetilde{a}^{-1}x \end{pmatrix}$$

□

Лемма 12.7. *Для любой матрицы $a \in \mathrm{GL}_n(F)$ существует матрица-перестановка w такая, что все главные подматрицы в матрице wa обратимы.*

Доказательство. Проведем доказательство индукцией по n. При $n = 1$ доказывать нечего. Пусть $n > 1$. Ранг подматрицы, составленной из первых $n - 1$ столбцов матрицы a, равен $n - 1$, так как все столбцы обратимой матрицы линейно независимы. Поэтому существует $n-1$ линейно независимых строк этой матрицы. Переставим эти строки на первые $n - 1$ мест. Тогда главная подматрица порядка $n-1$ станет обратимой. Далее используем индукционное предположение для главной подматрицы порядка $n - 1$. □

Глава 3

Начала теории групп

1. Простейшие конструкции

Гомоморфизмом алгебраических структур данного типа называется функция из одной такой структуры в другую, сохраняющая все операции. Для того, чтобы придать этому определению точный смысл, необходимо сначала строго определить, что такое алгебраическая структура, что увело бы нас в дебри науки, называемой универсальной алгеброй. Так что разберемся отдельно с группами, а в соответствующем месте с кольцами.

Определение 1.1. Пусть G с операцией $\star$ и H с операцией $\#$ – группы. Функция $f : G \to H$ называется гомоморфизмом, если $f(a \star b) = f(a)\#f(b)$ для любых $a, b \in G$.

Подструктурой алгебраической структуры называется подмножество структуры, замкнутое относительно всех операций, включая взятие нейтрального и обратного элемента, если аксиомы структуры гарантируют их наличие. Из этого вытекает, что подструктура является структурой того же типа, потому что выполнение свойств (кроме существования) в подмножестве следует из выполнения этих свойств во всем множестве (в этом смысл считать, что группа – это множество с одной 0-арной, одной унарной и одной бинарной операциями; тогда свойства формулируются без квантора существования).

Определение 1.2. Непустое подмножество H группы G называется подгруппой, если

$$a, b \in H \implies ab, a^{-1} \in H.$$

Если $a \in H$, то $a^{-1} \in H$, а, следовательно, и их произведение равное нейтральному элементу лежит в подгруппе H. Ясно, что подгруппа сама являются группой относительно тех же операций, которые заданы в объемлющей группе. Если H – подгруппа в G, то пишут $H \leqslant G$. В любой группе есть две тривиальные подгруппы: сама группа и множество состоящее из одного нейтрального элемента.

Прямым произведением алгебраических структур одного типа называется декартово произведение множеств с покомпонентными операциями. В случае, если одна из операций называется сложением, обычно говорят о прямой сумме, а не о прямом произведении. Надо отметить, что эта терминология не совпадает с терминологией теории категорий, которая приобретает все большую популярность.

Определение 1.3. Пусть G_1 и G_2 – группы с операциями $\star_1$ и $\star_2$ соответственно. Прямое произведение $G = G_1 \times G_2$ – это декартово произведение G_1 и G_2 с операцией $\star$, заданной покомпонентно: $(g_1, g_2) \star (h_1, h_2) = (g_1 \star_1 h_1, g_2 \star_2 h_2)$, где $g_1, h_1 \in G_1$, а $g_2, h_2 \in G_2$.

Аналогично определяется прямое произведение любого (даже не обязательно конечного) семейства групп. Если группы коммутативны, операция обозначена знаком $+$, а их количество конечно, то вместо термина “прямое произведение” обычно употребляют термин “прямая сумма” и обозначают ее знаком $\bigoplus$, например, $G = \bigoplus_{k=1}^{n} G_n$.

Если G_1 и G_2 являются подгруппами в G, то можно определить отображение $\pi : G_1 \times G_2 \to G$ формулой $\pi(x, y) = xy$. Для того, чтобы это отображение было гомоморфизмом, необходимо и достаточно, чтобы $xy = yx$ при всех $x \in G_1$ и $y \in G_2$. Ясно, что π сюръективно $\iff G = G_1G_2$. Инъективность π равносильна тому, что $G_1 \cap G_2 = \{1\}$.

Предложение 1.4. *Отображение $\pi : G_1 \times G_2 \to G$ является изоморфизмом тогда и только тогда, когда*

(1) $xy = yx$ *при всех* $x \in G_1$ *и* $y \in G_2$.
(2) $G = G_1 G_2$.
(3) $G_1 \cap G_2 = \{1\}$.

В этом случае говорят, что G является внутренним прямым произведением подгрупп G_1 и G_2.

Пусть теперь снова G_1, G_2 – произвольные группы. Мономорфизм $\iota_1 : G_1 \to G_1 \times G_2$, $\iota_1(x) = (x, 1)$, позволяет отождествить G_1 с ее образом в прямом произведении. Аналогично, G_2 отождествляется с подгруппой $\{1\} \times G_2$ группы $G_1 \times G_2$ при помощи мономорфизма $\iota_2 : G_2 \to G_1 \times G_2$, $\iota_2(y) = (1, y)$. После этих отождествлений G становится внутренним прямым произведением G_1 и G_2 (на самом деле внутренним прямым произведением $\operatorname{Im}\iota_1$ и $\operatorname{Im}\iota_2$).

Определены также проекции прямого произведения на сомножители: $\pi_i : G_1 \times G_2 \to G_i$, $\pi_1(x, y) = x$, $\pi_2(x, y) = y$. Заметим, что композиции $\pi_i \circ \iota_i$ являются тождественными отображениями.

2. Гомоморфизмы, ядро и образ

Образ гомоморфизма $\varphi : X \to Y$ – это его образ как функции, т. е. $\operatorname{Im}\varphi = \{\varphi(x) \mid x \in X\}$. Ядро гомоморфизма $\operatorname{Ker}\varphi = \varphi^{-1}(e)$. Инъективный гомоморфизм называется мономорфизмом, сюръективный – эпиморфизмом, а биективный – изоморфизмом. Если между двумя группами существует изоморфизм, то они называются изоморфными.

Лемма 2.1. *Если $\varphi : G \to H$ – гомоморфизм групп, то $\varphi(e_G) = e_H$, а $\varphi(x^{-1}) = \varphi(x)^{-1}$ для любого $x \in G$.*

Лемма 2.2. *Пусть $\varphi : G \to H$ – гомоморфизм групп, $g \in G$, а $h = \varphi(g)$. Тогда $\varphi^{-1}(h) = g \operatorname{Ker}\varphi$. Гомоморфизм инъективен тогда и только тогда, когда его ядро состоит из одного элемента.*

Определение 2.3. Подгруппа H группы G называется нормальной, если для любых $g \in G$ и $h \in H$ имеет место включение $g^{-1}hg \in H$. В других обозначениях: $\forall g \in G :\ g^{-1}Hg \subseteq H$. Если H – нормальная подгруппа в G, то пишут $H \trianglelefteq G$.

Заметим, что любая подгруппа абелевой группы является нормальной.

Лемма 2.4. *Подгруппа H нормальна в группе G тогда и только тогда, когда $\forall g \in G$: $gH = Hg$.*

Лемма 2.5. *Образ гомоморфизма групп является подгруппой, а ядро – нормальной подгруппой.*

3. Порождение, циклические группы, порядок элемента

Определение 3.1. Пусть X – подмножество группы G. Подгруппой, порожденной множеством X, называется наименьшая подгруппа в G, содержащая X. Подгруппа, порожденная X, обозначается $\langle X \rangle$. Подгруппа, порожденная одним элементом (точнее, одноэлементным множеством) называется циклической.

Так как пересечение подгрупп снова является подгруппой, то подгруппа, порожденная X, всегда существует. Действительно, это пересечение всех подгрупп, содержащих X.

Лемма 3.2. *$\langle X \rangle$ состоит из всех элементов вида $x_1 \cdots x_k$, где k – некоторое натуральное число, а $x_i \in X \cup X^{-1}$ (здесь, как обычно, $X^{-1} = \{x^{-1} \mid x \in X\}$). Обратите внимание, что здесь элементы $x_1, \ldots, x_k$ не обязательно различны. Если групп абелева, то произведение этих элементов приводится к виду $x_1^{i_1} \cdots x_m^{i_m}$, где $x_1, \ldots, x_m$ различны, а $i_j \in \mathbb{Z}$. В аддитивной записи это превращается в линейную комбинацию элементов из X с целыми коэффициентами.*

Предложение 3.3. *Любая циклическая группа изоморфна аддитивной группе $\mathbb{Z}$ или $\mathbb{Z}_n$.*

Доказательство. По определению циклическая группа $\langle g\rangle = \{g^k \mid k \in \mathbb{Z}\}$. Если $g^k \neq g^m$ при $k \neq m$, то отображение $\mathbb{Z} \to \langle g\rangle$, $k \mapsto g^k$, является изоморфизмом.

Если $g^k = g^m$ при некоторых $k > m$, то $g^{k-m} = 1$. Пусть n – наименьшее натуральное число (не 0) такое, что $g^n = 1$. Любое целое l можно разделить на n с остатком: $l = ns + r$, $0 \leqslant r < n$. Тогда $g^l = g^{ns}g^r = g^r$. Таким образом, $\langle g\rangle = \{1, g, \ldots, g^{n-1}\}$. Нетрудно видеть, что отображение $\mathbb{Z}_n \to \langle g\rangle$, $k \mapsto g^k$, является изоморфизмом. □

Определение 3.4. Пусть g – элемент группы G. Порядок циклической подгруппы, порожденной g, называется порядком элемента g, т. е. $\operatorname{ord} g = |\langle g\rangle|$. Как видно из доказательства предыдущего утверждения, порядок элемента g – это наименьшее натуральное число n такое, что $g^n = 1$.

4. Смежные классы и теорема Лагранжа

Определение 4.1. Пусть $H \leqslant G$. Множества gH и Hg называются левым (соотв. правым) смежными классами по подгруппе H. Множество левых смежных классов обозначается через G/H, а правых – $H\backslash G$ (не путать с $H \setminus G = H \smallsetminus G$; я специально немного опустил H и приподнял G, но обычно это не делается). Элемент смежного класса часто называют его представителем.

Как мы заметили в параграфе 2, левые смежные классы по ядру гомоморфизма совпадают с правыми. Кроме того, из леммы 2.2 следует, что группа разбивается в дизъюнктное объединение смежных классов по ядру. Сейчас мы выясним, какие свойства выживают для смежных классов по произвольным подгруппам.

Определим отношение "сравнимости по модулю H" на множестве G по формуле:
$a \equiv b \mod H \iff a \in bH$. На самом деле мы определили "сравнимость по модулю H слева". Сравнимость по модулю H справа определяется включением $a \in Hb$. Везде, кроме следующей леммы мы будем использовать понятие сравнимости по модулю H, только когда H – нормальная подгруппа в G. В этом случае по лемме 2.4 $gH = Hg$, и сравнимость слева и справа совпадают.

Лемма 4.2. *Сравнимость по модулю H является отношением эквивалентности. Два левых смежных класса либо не пересекаются, либо совпадают.*

Доказательство. Рефлексивность: $a = ae \in aH$. Симметричность: $a \in bH \implies \exists h \in H : a = bh \implies b = ah^{-1} \in aH$. Транзитивность: если $a \in bH$ и $b \in cH$, т. е. $a = bh$ и $b = ch'$ для некоторых $h, h' \in H$, то $a = ch'h \in cH$.

Ясно, что классы сравнимости по модулю H – это левые смежные классы по подгруппе H, откуда вытекает второе утверждение леммы. □

Лемма 4.3. *Любые два смежных класса равномощны, т. е. между ними существует биекция. В частности, если они конечны, то они содержат одинаковое количество элементов.*

Теорема 4.4 (теорема Лагранжа)**.** *Если H – подгруппа конечной группы G, то $|G| = |H| \cdot |G/H|$.*

Предложение 4.5. *Множества G/H и $H\backslash G$ равномощны, т. е. между ними существует биекция.*

Доказательство. Биекция $G/H \to H\backslash G$ задается по правилу $aH \mapsto (aH)^{-1} = Ha^{-1}$ (здесь $(aH)^{-1} = \{(ah)^{-1} \mid h \in H\}$). □

В частности, если количество левых или правых смежных классов конечно, то $|G/H| = |H\backslash G|$. Это число называют индексом подгруппы H в G и обозначают через $|G : H|$ (если количество смежных классов бесконечно, пишут $|G : H| = \infty$). Индекс подгруппы может быть конечен, даже если сама группа бесконечна, например $H = 2\mathbb{Z}$ в $G = \mathbb{Z}$.

5. Факторгруппа и теоремы о гомоморфизме

Предложение 5.1. *Для любой нормальной подгруппы H группы G существует группа F и эпиморфизм $\pi : G \to F$, ядро которого равно H.*

Доказательство. Положим $F = G/H$ и зададим отображение $\pi : G \to F$ по формуле $\pi(x) = xH$. Зададим операцию в F по формуле $(xH) \cdot (yH) = xyH$. Так как H – нормальная подгруппа, то эта операция не зависит от выбора представителей x и y смежных классов xH и yH. Действительно, $xhyh' = xy(y^{-1}hy)h' \in xyH$. Ассоциативность операции следует из ассоциативности операции в группе G, нейтральным элементом является смежный класс $eH = H$, а обратным для xH – смежный класс $x^{-1}H$. Таким образом, F является группой. Тот факт, что π – гомоморфизм, сразу следует из формулы умножения смежных классов. Очевидно, что π сюръективен. Уравнение $\pi(x) = e_{G/H} = H$ равносильно тому, что $x \in H$ (иначе смежные классы $\pi(x) = xH$ и H не совпадают). Следовательно, $\operatorname{Ker}\pi = H$. □

Определение 5.2. Группа G/H, построенная в доказательстве, называется факторгруппой G по H, а отображение π – канонической проекцией или гомоморфизмом редукции по модулю H.

Теорема 5.3 (универсальное свойство факторгруппы)**.** *Пусть $\varphi : G \to H$ гомоморфизм групп, а $N \trianglelefteq G$. Если $\operatorname{Ker}\varphi \geqslant N$, то существует единственный гомоморфизм $\psi : G/N \to H$ такой, что $\varphi = \psi \circ \pi$. Если φ – эпиморфизм, то и ψ – эпиморфизм. Если $\operatorname{Ker}\varphi = N$, то ψ – мономорфизм.*

Доказательство. Пусть $x \in G$. Из равенства $\varphi = \psi \circ \pi$ следует, что

$$\psi(xN) = \varphi(x).$$

Если y – другой представитель смежного класса xN, то $y = xn$ для некоторого $n \in N$, и $\psi(yN) = \varphi(y) = \varphi(x)\varphi(n) = \varphi(x)$, так как $n \in N \leqslant \operatorname{Ker}\varphi$. Таким образом, вынесенная формула корректно определяет отображение. Из определения умножения смежных классов следует, что это отображение является гомоморфизмом. Очевидно, что отображение, удовлетворяющее равенству $\varphi = \psi \circ \pi$, единственно.

Если композиция сюръективна, то левый гомоморфизм (тот, который применяется последним) обязан быть сюръективным. Пусть теперь $\operatorname{Ker}\varphi = N$. Тогда

$$xN \in \operatorname{Ker}\psi \iff x \in \operatorname{Ker}\varphi = N \iff xN = 1_{G/N}.$$

□

Следствие 5.4 (Теорема о гомоморфизме групп)**.** *Пусть $\varphi : G \to H$ – гомоморфизм групп. Тогда $\operatorname{Im}\varphi \cong G/\operatorname{Ker}\varphi$.*

Доказательство. Отображение $\bar{\varphi} : G \to \operatorname{Im}\varphi$ заданное формулой $\bar{\varphi}(x) = \varphi(x)$ является эпиморфизмом, причем его ядро равно $\operatorname{Ker}\varphi$. По предыдущей теореме существует изоморфизм $G/\operatorname{Ker}\varphi \to \operatorname{Im}\varphi$. □

С помощью теоремы о гомоморфизме легко доказать, что $\mathbb{Z}_n \cong \mathbb{Z}/n\mathbb{Z}$.

Лемма 5.5. *Любая подгруппа в $\mathbb{Z}$ совпадает с $n\mathbb{Z}$ для некоторого $n \in \mathbb{N}_0$.*

Доказательство. Пусть $H \leqslant \mathbb{Z}$, а n – наименьшее натуральное число, содержащееся в H (если H не содержит натуральных чисел, то $H = \{0\}$ и можно взять $n = 0$). Ясно, что $n\mathbb{Z} \leqslant H$. Для любого $x \in H$ остаток от деления x на n равен $r = x - kn$ для некоторого $k \in \mathbb{Z}$, и, следовательно, лежит в H. Можно считать, что $0 \leqslant r < n$. Так как в H нет чисел от 1 до $n-1$, то $r = 0$, т. е. $x \in n\mathbb{Z}$. Таким образом, $H \leqslant n\mathbb{Z}$. □

Пусть $G = \langle g \rangle$ – циклическая группа. Зададим отображение $\varphi : \mathbb{Z} \to G$ формулой $\varphi(k) = g^k$. Ясно, что φ – эпиморфизм. Применив к φ теорему о гомоморфизме, получим $G \cong \mathbb{Z}/n\mathbb{Z}$ для некоторого $n \in \mathbb{N}_0$ т. е., фактически, доказательство предложения 3.3.

Будем говорить, что элемент $g \in G$ нормализует подгруппу $B \leqslant G$, если $g^{-1}Bg = B$. Подгруппа $A \leqslant G$ нормализует B, если каждый ее элемент нормализует B. Легко видеть, что в этом случае AB – подгруппа, и B нормальна в AB.

Теорема 5.6 (2-я теорема о гомоморфизме)**.** *Пусть $A, B \leqslant G$. Если A нормализует B, то*

$$\frac{AB}{B} \cong \frac{A}{A \cap B}.$$

Доказательство. Зададим $\varphi : A \to AB/B$ как композицию вложения $A \hookrightarrow AB$ с канонической проекцией $AB \twoheadrightarrow AB/B$. $\varphi(a) = 1_{AB/B} \iff aB = B \iff a \in B$. Поэтому $\operatorname{Ker}\varphi = A \cap B$. Любой элемент группы AB/B имеет вид $abB = aB = \varphi(a)$ для некоторых $a \in A$ и $b \in B$. Следовательно, φ сюръективен. По теореме о гомоморфизме: $A/A \cap B = A/\operatorname{Ker}\varphi \cong AB/B$. □

Теорема 5.7 (3-я теорема о гомоморфизме)**.** *Пусть $K \trianglelefteq G$, а $\pi : G \to G/K$ – канонический гомоморфизм. Тогда отображение $H \mapsto \pi(H)$ является биекцией множества подгрупп в G, содержащих K, на множество подгрупп в G/K. При этом H нормальна в G тогда и только тогда, когда $\pi(H)$ нормальна в G/K и*

$$\frac{G/K}{H/K} \cong \frac{G}{H}.$$

(здесь $\pi(H)$ обозначена за H/K, так как эти группы даже не просто изоморфны, а равны).

Доказательство. Обратным к заданному отображению будет взятие полного прообраза. Очевидно, что $\pi\big(\pi^{-1}(F)\big) = F$ и $\pi^{-1}\big(\pi(H)\big) \supseteq H$. Если $g \in \pi^{-1}\big(\pi(H)\big)$, т. е. $\pi(g) = \pi(h)$ для некоторого $h \in H$, то $gh^{-1} \in \operatorname{Ker}\pi = K \leqslant H$, откуда $g \in hH = H$.

Так как при нашей биекции сопряженные подгруппы переходят в сопряженные, утверждение о нормальности очевидно. Для доказательства изоморфизма зададим гомоморфизм $\pi' : G \to \frac{G/K}{H/K}$ как композицию двух канонических проекций. Композиция сюръекций сюръективна, с другой стороны $\operatorname{Ker}\pi' = \pi^{-1}(H/K) = H$. Теперь результат следует из первой теоремы о гомоморфизме. □

6. Сопряженные элементы, коммутаторы и коммутант

Пусть x, y, z – элементы группы G. Элемент $x^y := y^{-1}xy$ называется (правым) сопряженным к x при помощи y, а ${}^yx = x^{y^{-1}} = yxy^{-1}$ – левым сопряженным к x при помощи y.

Лемма 6.1. *$(xy)^z = x^z y^z$ и ${}^z(xy) = {}^zx \cdot {}^zy$. Функция $c_z : G \to G$, $c_z(x) = {}^zx$, является автоморфизмом группы G.*

${}^{yz}x = {}^y({}^zx)$. Функция $c(z) = c_z$ является гомоморфизмом из группы G в группу $\operatorname{Aut}(G)$ автоморфизмов G с операцией композиции (проверьте, что $\operatorname{Aut}(G)$ является подгруппой в S_G).

Отношение "x сопряжено с y" очевидно является отношением эквивалентности. Классы этой эквивалентности называются классами сопряженных элементов.

Заметим, что для доказательства нормальности подгруппы H в группе G, достаточно проверить условие нормальности только на образующих.

Лемма 6.2. *Пусть $H = \langle X \rangle$ – подгруппа в группе $G = \langle Y \rangle$. Тогда H нормальна в G тогда и только тогда, когда $x^y \in H$ для любых $x \in X$ и $y \in Y \cup Y^{-1}$.*

Доказательство. Если $H \trianglelefteq G$, то включения очевидны. Обратно, пусть $h \in H$, а $g = y_1 \cdots y_m \in G$, где $y_1, \ldots, y_m \in Y \cup Y^{-1}$. Индукцией по m докажем, что $h^g \in H$, откуда будет следовать нормальность. При $m = 0$ это очевидно, так как $g = 1$. Пусть $m \geqslant 1$. По индукционному предположению $h^{y_1 \cdots y_{m-1}} \in H$, следовательно, $h^{y_1 \cdots y_{m-1}} = x_1 \cdots x_n$ для некоторого $n \in \mathbb{N}$ и

$x_1, \dots, x_n \in X \cup X^{-1}$. Тогда $h^g = (x_1 \cdots x_n)^{y_m} = x_1^{y_m} \cdots x_n^{y_m}$. Если, $x_i \in X$, то $x_i^{y_m} \in H$ по условию. Если же $x_i^{-1} = x \in X$, то $x_i^{y_m} = (x^{-1})^{y_m} = (x^{y_m})^{-1} \in H^{-1} = H$. □

Наименьшая нормальная подгруппа группы G, содержащая подгруппу H, называется нормальным замыканием H в G и обозначается H^G. Нетрудно видеть, что она порождена всеми элементами вида h^g, $h \in H$, $g \in G$.

Коммутатором называется элемент $[x, y] = xyx^{-1}y^{-1}$.

Лемма 6.3. *Выполнены следующие коммутаторные формулы.*

(1) $[x, y]^{-1} = [y, x]$.
(2) $[x, yz] = [x, y] \cdot {}^y[x, z]$.

Пусть X, Y – подгруппы в G. Взаимным коммутантом этих подгрупп называется подгруппа, порожденная всеми коммутаторами $[x, y]$, $x \in X$, $y \in Y$. Так как сопряженный с коммутатором является коммутатором, то взаимный коммутант двух нормальных подгрупп является нормальной подгруппой. Коммутант $[G, G]$ группы G обладает следующим универсальным свойством.

Лемма 6.4. *Ядро любого гомоморфизма из G в абелеву группу содержит коммутант. Коммутант группы – это наименьшая нормальная подгруппа, фактор по которой абелев.*

Из леммы вытекает, что любой гомоморфизм в абелеву группу пропускается через факторгруппу $G^{\mathrm{ab}} := G/[G, G]$.

Взаимный коммутант подгрупп, не являющихся нормальными, также нормален, хотя и не во всей группе.

Лемма 6.5. *Пусть X и Y – подгруппы в G. Тогда $[X, Y] \trianglelefteq \langle X \cup Y \rangle$.*

Доказательство. Докажем, что $[x, y]^z \in [X, Y]$ при всех $x \in X$ и $y, z \in Y$. Действительно, по формуле 6.3(2)

$$[x, y]^z = {}^{z^{-1}}[x, y] = [x, z^{-1}]^{-1} \cdot [x, z^{-1}y] \in [X, Y].$$

Аналогично, при $x, z \in X$ и $y \in Y$ имеем

$$\begin{aligned}[x, y]^z &= \big([y, x]^{-1}\big)^z = \Big({}^{z^{-1}}[y, x]\Big)^{-1} = \big([y, z^{-1}]^{-1} \cdot [y, z^{-1}x]\big)^{-1} \\ &= [z^{-1}x, y] \cdot [z^{-1}, y]^{-1} \in [X, Y].\end{aligned}$$

Теперь результат следует из леммы 6. □

Почти то же самое рассуждение показывает, что взаимный коммутант является наименьшей нормальной подгруппой в $\langle X \cup Y \rangle$, содержащей все коммутаторы образующих групп X и Y.

Лемма 6.6. *Пусть S_X и S_Y – множества образующих подгрупп X и Y соответственно. Тогда $[X, Y] = \langle [s, t] \mid s \in S_X,\ t \in S_Y \rangle^{\langle X \cup Y \rangle}$.*

Доказательство. Обозначим правую часть последнего равенства через Z. По предыдущей лемме она содержится в левой. Поэтому достаточно показать, что любой образующий элемент левой группы содержится в Z. Пусть, сначала, $s \in S_X$, а $y = t_1 \dots t_n$, где $t_1, \dots, t_n \in S_Y$. Индукцией по n докажем, что $[s, y] \in Z$. База индукции ($n = 1$) выполнена по определению Z. При $n > 1$ по формуле (2) леммы 6.3 имеем $[s, y] = [s, t_1] \cdot {}^{t_1}[s, t_2 \dots t_n]$. По индукционному предположению $[s, t_2 \dots t_n] \in Z$, следовательно, $[s, y] \in Z$ для любого $s \in S_X$ и $y \in Y$. Замена образующей группы X на любой элемент этой группы происходит аналогично (то, что мы уже доказали, является базой индукции). □

7. Группа унитреугольных матриц

Пусть F – поле или, в большей общности, коммутативное кольцо с 1. Положим

$$U_n^{(k)} = U_n^{(k)}(F) = \{a \in \mathrm{M}_n(F) \mid a_{ii} = 1,\ a_{ij} = 0 \text{ при всех } i \neq j,\ j - i < k\},$$

$U_n = U_n(F) := U_n^{(1)}(F)$ и $U_n^{(k)} = \{1\}$ при $k \geqslant n$.

Лемма 7.1. *Группа $U_n^{(k)}$ порождена трансвекциями $t_{ij}(\alpha)$ по всем $\alpha \in F$ и $j - i \geqslant k$.*

Для доказательства следующих утверждений нам потребуются следующие формулы, которые легко проверить непосредственным вычислением. Пусть $i \neq j \neq h \neq k \neq i$ – натуральные числа. Тогда

$$\begin{aligned} t_{ij}(\alpha)t_{ij}(\beta) &= t_{ij}(\alpha+\beta) \\ [t_{ij}(\alpha), t_{jk}(\beta)] &= t_{ik}(\alpha\beta) \\ [t_{ij}(\alpha), t_{hi}(\beta)] &= t_{hj}(-\alpha\beta) \\ [t_{ij}(\alpha), t_{hk}(\beta)] &= e. \end{aligned} \tag{15}$$

Пусть $H \leqslant G$. Для $g \in G$ и $h \in H$ условие ${}^g h \in H$ равносильно $[g, h] \in H$. Поэтому условие нормальности подгруппы H в G можно переписать в виде: $\forall\, g \in G,\ h \in H :\ [g, h] \in H$ или $[G, H] \leqslant H$.

Лемма 7.2. *Группа $U_n^{(k)}$ нормальна в U_n. Более того, $[U_n^{(k)}, U_n^{(m)}] = U_n^{(k+m)}$.*

Доказательство. Заметим, что плохой случай $[t_{ij}(\alpha), t_{ji}(\beta)]$ не может встретиться в верхнетреугольной группе. Поэтому с помощью формул (15) легко проверить условия леммы 6.2, откуда вытекает нормальность.

Из тех же формул легко видеть, что коммутатор образующих групп $U_n^{(k)}$ и $U_n^{(m)}$ лежит в $U_n^{(k+m)}$. Так как последняя подгруппа нормальна, то по лемме 6.6 и весь взаимный коммутант $[U_n^{(k)}, U_n^{(m)}]$ содержится в этой подгруппе. С другой стороны, каждая образующая группы $U_n^{(k+m)}$ является коммутатором образующих групп $U_n^{(k)}$ и $U_n^{(m)}$, откуда вытекает требуемое равенство. □

Лемма 7.3. *Любой элемент группы U_n единственным образом выражается в виде произведения $\prod\limits_{n \geqslant j > i \geqslant 1} t_{ij}(\alpha_{ij})$ в любом наперед заданном порядке на множестве пар (i, j), $n \geqslant j > i \geqslant 1$.*

Доказательство. Пусть $u \in U_n$. Индукцией по k докажем, что

$$u \in \prod_{1 \leqslant j-i < k} t_{ij}(\alpha_{ij}) \cdot U_n^{(k)},$$

где произведение берется в заданном порядке, а α_{ij} определены единственным образом. При $k = 1$ доказывать нечего. Пусть $k > 1$, а $u \in \prod\limits_{1 \leqslant j-i < k-1} t_{ij}(\alpha_{ij}) \cdot U_n^{(k-1)}$ – представление u, существующее по индукционному предположению.

Легко видеть, что любой элемент a из $U_n^{(k-1)}$ лежит в смежном классе $\prod\limits_{i=1}^{n-k+1} t_{i\,i+k-1}(\alpha_{i\,i+k-1}) U_n^{(k)}$, где $\alpha_{i\,i+k-1} = a_{i\,i+k-1}$. По лемме 7.2 трансвекции $t_{i\,i+k-1}(\alpha_{i\,i+k-1})$ коммутируют с элементами U_n по модулю $U_n^{(k)}$. Поэтому эти трансвекции можно поставить в нужное место произведения $\prod\limits_{1 \leqslant j-i < k-1} t_{ij}(\alpha_{ij})$, чтобы получить требуемое включение.

Если $\prod\limits_{1 \leqslant j-i < k} t_{ij}(\alpha_{ij}) \equiv \prod\limits_{1 \leqslant j-i < k} t_{ij}(\beta_{ij}) \mod U_n^{(k)}$,

то $\prod\limits_{1 \leqslant j-i < k-1} t_{ij}(\alpha_{ij}) \prod\limits_{j-i=k-1} t_{ij}(\alpha_{ij}) \equiv \prod\limits_{1 \leqslant j-i < k-1} t_{ij}(\beta_{ij}) \prod\limits_{j-i=k-1} t_{ij}(\beta_{ij}) \mod U_n^{(k)}$, откуда по индукционному предположению $\alpha_{ij} = \beta_{ij}$ при всех $1 \leqslant j - i < k - 1$. Следовательно, $\prod\limits_{j-i=k-1} t_{ij}(\alpha_{ij}) \equiv$

$\prod_{j-i=k-1} t_{ij}(\beta_{ij}) \mod U_n^{(k)}$. Вычисление показывает, что при $j-i=k-1$ в позиции (i,j) левой части равенства стоит α_{ij}, а правой – β_{ij} (умножение на матрицу из $U_n^{(k)}$ не меняет значения в этих позициях). Таким образом, $\alpha_{ij}=\beta_{ij}$ при всех $1 \leqslant j-i<k$, что доказывает единственность. □

8. Приведенное разложение Брюа и второе доказательство разложения Гаусса

Напомним, что через $W=W_n$ мы обозначаем множество матриц перестановок. Для $w \in W$ положим

$$U_w = \langle t_{ij}(\alpha) \mid 1 \leqslant i < j \leqslant n,\ \alpha \in F,\ t_{ij}(\alpha)^w \in U_n^- \rangle.$$

Напомним, что $B_n = B_n(F)$ обозначает множество обратимых верхнетреугольных, $B_n^- = B_n^-(F)$ – обратимых нижнетреугольных, а $U_n^- = U_n^-(F)$ – нижних унитреугольных матриц размера $n \times n$.

Теорема 8.1 (приведенное разложение Брюа)**.** *Пусть $w \in W$. Тогда $B_nwB_n = U_wwB_n$, следовательно, $\mathrm{GL}_n(F) = U_wwB_n$. При этом разложение данного элемента единственно, т. е. для любого $g \in \mathrm{GL}_n(F)$ существуют единственные $w \in W$, $u \in U_w$ и $b \in B_n$ такие, что $g = uwb$.*

Доказательство. Обозначим через $T_n = T_n(F)$ множество обратимых диагональных матриц. Любая обратимая треугольная матрица однозначно представляется в виде произведения унитреугольной на диагональную, т. е. $B_n = U_nT_n$. Ясно также, что $T_n^w = T_n$. Поэтому $B_nwB_n = U_nT_nwB_n = U_nwT_n^wB_n = U_nwB_n$.

Обозначим $\overline{U}_w = \langle t_{ij}(\alpha) \mid t_{ij}(\alpha)^w \in U_n \rangle$. Тогда по лемме 7.3 $U_n = U_w\overline{U}_w$. Следовательно,

$$B_nwB_n = U_nwB_n = U_w\overline{U}_wwB_n = U_ww\overline{U}_w^wB_n \subseteq U_wwU_nB_n = U_wwB_n.$$

Обратное включение очевидно.

Докажем теперь единственность. Пусть $uwb = u'w'b'$, где $w, w' \in W$, $u \in U_w$, $u' \in U_{w'}$, и $b, b' \in B_n$. Тогда $(w')^{-1}(u')^{-1}uw = b'b^{-1} \in B_n$. Обозначим последнюю матрицу через c. Пусть w соответствует перестановке σ, т. е. $w_{i,\sigma(i)} = 1$ для некоторой перестановки $\sigma \in S_n$, и $w_{ij} = 0$ при $j \neq \sigma(i)$. Пусть, далее, w' соответствует $\sigma' \in S_n$. Тогда у матрицы $(w')^{-1}$ единицы стоят в позициях $(\sigma'(i), i)$. Следовательно, $c_{\sigma'(i)\,\sigma(i)} = \big((u')^{-1}u\big)_{ii} = 1$. Если $\sigma' \neq \sigma$, то найдется такой индекс i, что $\sigma'(i) > \sigma(i)$. Но тогда предыдущее равенство противоречит тому факту, что c – верхнетреугольная матрица. Таким образом, $\sigma' = \sigma$ и $w' = w$.

По определению U_w имеем $(u')^{-1}u \in U_w$ и $c = w^{-1}(u')^{-1}uw \in U_n^-$. Так как $U_n^- \cap B_n = \{e\}$, то $u' = u$ и $b' = b$. □

Заметим, что из последнего рассуждения следует единственность матрицы перестановки в обычном разложении Брюа, т. е. тот факт, что клетки Брюа не пересекаются.

Следствие 8.2. *Любая клетка Брюа содержится в соответствующей клетке Гаусса.*

Доказательство. $B_nwB_n = U_wwB_n = wU_w^wB_n \subseteq wU_n^-B_n = wB_n^-B_n$. □

9. Симметрическая группа

Одним из важных примеров групп является симметрическая группа. Она будет полезна нам как для иллюстрации понятий теории групп, так и в линейной алгебре при изучении антисимметричных форм и определителя матрицы.

Пусть X – множество. Напомним, что множество биекций $X \to X$ с операцией композиции называется симметрической группой множества X и обозначается через S_X. Если $X = \{1, \ldots, n\}$, то S_X обозначается через S_n и называется симметрической группой порядка n.

Ясно, что множество всех функций $X \to X$ является моноидом (нейтральный элемент – тождественное отображение $id(x) = x \forall x \in X$), а S_X является его группой обратимых элементов. Далее будем изучать группу S_n.

Пусть $\sigma \in S_n$. Определим отношение эквивалентности на множестве $\{1, \dots, n\}$ по правилу $i \overset{\sigma}{\sim} j \iff i = \sigma^k(j)$ для некоторого $k \in \mathbb{Z}$ В каждом классе эквивалентности выберем представителя и запишем все элементы класса в виде цикла: $\big(i\,\sigma(i)\,\dots\,\sigma^{m-1}(i)\big)$, где $\sigma^m(i) = i$. Записав все классы эквивалентности в виде циклов получим циклическую запись перестановки σ. Эта запись единственна с точностью до перестановки циклов и выбора первого элемента каждого цикла. Циклы длины 1 обычно не пишут. Набор из длин независимых циклов называется циклическим (или цикленным) типом перестановки. Например, перестановка

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 1 & 4 & 3 & 6 & 7 & 5 \end{pmatrix} = (12)(34)(567)$$

Имеет циклический тип $2 + 2 + 3$. Транспозицией называется цикл длины 2.

Лемма 9.1. $\sigma \circ \big(i_1\,\dots\,i_k\big) \circ \sigma^{-1} = \big(\sigma(i_1)\,\dots\,\sigma(i_k)\big)$.
Следовательно, сопряжение не меняет циклического типа перестановки.

Следствие 9.2. *Класс сопряженных элементов в S_n состоит из всех перестановок данного циклического типа. Количество классов сопряженных элементов равно количеству разбиений числа n в сумму натуральных чисел (порядок слагаемых не важен).*

Определение 9.3. Пусть $\sigma \in S_n$. Инверсией называется пара (i, j), $1 \leqslant i < j \leqslant n$, такая, что $\sigma(i) > \sigma(j)$. Четность количества инверсий называется четностью перестановки σ.

Лемма 9.4. *Любая перестановка записывается в виде произведения транспозиций соседних индексов.*

Доказательство. Если $\sigma \neq \mathrm{id}$, то существует индекс i такой, что $\sigma(i) > \sigma(i+1)$. Тогда в перестановке $\sigma \circ (i\,i+1)$ инверсий на 1 меньше, чем в σ. Далее индукция по числу инверсий. □

Лемма 9.5. *Если перестановка представлена в виде произведения m транспозиций соседних индексов, то ее четность равна четности m.*

Доказательство. Если $\sigma(i) > \sigma(i+1)$, то в перестановке $\sigma \circ (i\,i+1)$ инверсий на 1 меньше, чем в σ, в противном случае – на 1 больше. □

Теорема 9.6. *Отображение $\varepsilon : S_n \to \mathbb{Z}/2\mathbb{Z}$, сопоставляющее перестановке ее четность, удовлетворяет соотношению $\varepsilon(\sigma\tau) = \varepsilon(\sigma) + \varepsilon(\tau) \mod 2$, т. е. является гомоморфизмом групп.*

Следствие 9.7. *Четность перестановки циклического типа $k_1 + \dots + k_m$ равна $k_1 + \dots + k_m - m \mod 2$.*

Доказательство. Так как ε является гомоморфизмом, достаточно доказать, что четность цикла длины k равна $k - 1 \mod 2$. Гомоморфизм в абелеву группу переводит сопряженные элементы в одно и то же. По следствию 9.2 любой цикл длины m сопряжен с циклом $(1\,2\,\dots\,m)$. В этой перестановке легко посчитать количество транспозиций.

Альтернативно, можно представить цикл длины m в виде произведения цикла длины $m-1$ и транспозиции, посчитать количество инверсий в транспозиции (оно будет нечетным) и применить индукцию по m. □

10. Экспонента группы

Порядок произведения двух некоммутирующих элементов группы никак не связан с порядками сомножителей. Но если элементы коммутируют, то такая связь есть.

Лемма 10.1. *Пусть G – группа, $a, b \in G$, и $ab = ba$. Тогда $\operatorname{ord} ab$ делит $\operatorname{lcm}(\operatorname{ord} a, \operatorname{ord} b)$ и делится на $\frac{\operatorname{lcm}(\operatorname{ord} a, \operatorname{ord} b)}{\gcd(\operatorname{ord} a, \operatorname{ord} b)}$. В частности, если $\gcd(\operatorname{ord} a, \operatorname{ord} b) = 1$, то $\operatorname{ord} ab = \operatorname{ord} a \operatorname{ord} b$.*

Доказательство. Если m – общее кратное $\operatorname{ord} a$ и $\operatorname{ord} b$, то $(ab)^m = a^m b^m = 1$, откуда $\operatorname{ord} ab$ делит m.

Пусть $\operatorname{ord} a = kp$, а $\operatorname{ord} b = kq$, где $\gcd(p, q) = 1$. Пересечение $\langle a\rangle \cap \langle b\rangle$ является подгруппой и в $\langle a\rangle$ и в $\langle b\rangle$. По теореме Лагранжа его порядок делит $\operatorname{ord} a$ и $\operatorname{ord} b$, следовательно, делит $\gcd(\operatorname{ord} a, \operatorname{ord} b) = k$. Если $(ab)^n = 1$, то $a^n = b^{-n} \in \langle a\rangle \cap \langle b\rangle$, откуда $a^{nk} = b^{-nk} = 1$. Следовательно, nk делится на pk и на qk, а значит, n делится на p и q. Так как p и q взаимно просты, то n делится на $pq = \frac{\operatorname{lcm}(\operatorname{ord} a, \operatorname{ord} b)}{\gcd(\operatorname{ord} a, \operatorname{ord} b)}$. □

Определение 10.2. Экспонентой (или показателем) группы G называется наименьшее натуральное число d такое, что $g^d = 1$ для любого $g \in G$. Если такого d не существует, то говорят, что экспонента группы равна бесконечности.

Лемма 10.3 (свойства экспоненты группы)**.**

(1) *Экспонента группы равна наименьшему общему кратному порядков ее элементов.*
(2) *Если группа конечна, то ее экспонента делит ее порядок.*
(3) *Экспонента прямого произведения групп $G_1 \times \cdots \times G_m$ равна наименьшему общему кратному экспонент групп $G_1, \ldots, G_m$.*
(4) *Если G – абелева группа конечной экспоненты, то существует элемент, порядок которого равен ее экспоненте.*
(5) *Конечная абелева группа является циклической тогда и только тогда, когда ее экспонента равна ее порядку.*

Доказательство. Все пункты кроме пункта (4) доказываются легко. Пусть $d = p_1^{k_1} \cdots p_m^{k_m}$ – экспонента группы G, где $p_1, \ldots p_m$ – различные простые числа. Тогда существуют элементы $g_1, \ldots, g_m \in G$, порядки которых делятся на $p_1^{k_1}, \ldots, p_m^{k_m}$ соответственно. Ясно, что если $\operatorname{ord} g = rs$, то $\operatorname{ord} g^s = r$. Возводя элементы $g_1, \ldots, g_m$ в подходящие степени, можно считать, что $\operatorname{ord} g_i = p_i^{k_i}$ при всех $i = 1, \ldots, m$.

Пусть теперь G – абелева группа. Тогда по лемме 10.1 для элементов $a, b \in G$ взаимно простых порядков имеет место равенство $\operatorname{ord}(ab) = \operatorname{ord} a \operatorname{ord} b$. Теперь индукцией по m легко доказать, что $\operatorname{ord}(g_1 \cdots g_m) = \operatorname{ord} g_1 \cdots \operatorname{ord} g_m = d$. □

Глава 4

Кольца и алгебры

Мы всегда будем предполагать, что все кольца и алгебры содержат единицу. Однако в параграфах 1–2 наличие единицы практически не дает никакого упрощения. Поэтому в этих параграфах можно рассматривать произвольные ассоциативные кольца. Начиная с некоторого места все кольца и алгебры будут предполагаться коммутативными.

1. Модули, алгебры, гомоморфизмы

Пусть R – кольцо с 1. Аддитивная абелева группа M называется правым R-модулем, если определено умножение $M \times R \to M$, $(x, r) \mapsto xr$, удовлетворяющее свойствам:

(1) $x(rs) = (xr)s$;
(2) $x(r + s) = xr + xs$;
(3) $(x + y)r = xr + yr$;
(4) $x \cdot 1 = x$

для любых $r, s \in R$ и $x, y \in M$.

Аналогично определяется левый R-модуль. Если R коммутативно, то понятия левого и правого R-модулей совпадают. Отображение $\varphi : M \to M'$ называется гомоморфизмом модулей, если $\varphi(x + y) = \varphi(x) + \varphi(y)$ и $\varphi(xr) = \varphi(x)r$.

Понятия ядра, образа, слоя гомоморфизма, а также универсальное свойство фактормодуля и теоремы о гомоморфизме переносятся с векторных пространств без изменений.

Пусть теперь R – коммутативное кольцо с 1. Кольцо A (возможно некоммутативное без 1) назывется R-алгеброй, если A является R-модулем и

$$(ab)r = (ar)b = a(br) \text{ для любых } a, b \in A \text{ и } r \in R.$$

Определение 1.1. Пусть R и A – кольца. Функция $f : R \to A$ называется гомоморфизмом колец, если $f(a + b) = f(a) + f(b)$ и $f(a \cdot b) = f(a) \cdot f(b)$ для любых $a, b \in R$. Для гомоморфизма колец с единицей будем требовать также, чтобы $f(1_R) = 1_A$.[1]

Если A – R-алгебра с 1, то можно определить *структурный гомоморфизм* $\varphi : R \to A$ по формуле $\varphi(r) = 1_A \cdot r$. Заметим, что образ φ лежит в центре кольца A, где $\operatorname{Center} A = \{x \in A \mid ax = xa \, \forall a \in A\}$. Обратно, если A – кольцо с 1, то любой гомоморфизм $\varphi : R \to \operatorname{Center} A$ задает на A структуру R-алгебры по правилу $ar = a\varphi(r)$. Несложно проверить, что эти две конструкции "взаимно обратны", т. е. применив их по очереди, получим то, с чего начинали. Таким образом, задание структуры R-алгебры на кольце с 1 эквивалентно заданию гомоморфизма $R \to A$, образ которого лежит в центре A.

Любое кольцо автоматически является $\mathbb{Z}$-алгеброй, поэтому без потери общности можно рассматривать алгебры вместо колец.

Определение 1.2. Гомоморфизм алгебр – это гомоморфизм модулей, одновременно являющийся гомоморфизмом колец. Пусть A и B – две R-алгебры. Образ гомоморфизма $f : A \to B$ – это его образ как функции. Ядро гомоморфизма $\operatorname{Ker} f = f^{-1}(0)$.

[1] В отличие от гомоморфизма групп сохранение единицы не вытекает из сохранения умножения. Например, отображение $f : \mathbb{Z} \to \mathbb{Z}_6$, заданное равенством $f(x) = 4x \mod 6$ является гомоморфизмом колец, но не является гомоморфизмом колец с 1, несмотря на то, что оба кольца содержат 1.

Инъективный гомоморфизм называется мономорфизмом, сюръективный – эпиморфизмом, а биективный – изоморфизмом. Если между двумя R-алгебрами существует изоморфизм, то они называются изоморфными.

Лемма 1.3. *Пусть $\varphi : A \to B$ – гомоморфизм R-алгебр. Тогда $\varphi(0) = 0$, а $\varphi(-x) = -\varphi(x)$ для любого $x \in A$.*

Если φ – гомоморфизм алгебр с 1, а $x \in A^*$, то $\varphi(x) \in B^*$ и $\varphi(x^{-1}) = \varphi(x)^{-1}$.

Лемма 1.4. *Пусть $\varphi : A \to B$ – гомоморфизм R-алгебр, $x \in A$, а $y = \varphi(x)$. Тогда $\varphi^{-1}(y) = x + \operatorname{Ker}\varphi$.*

Гомоморфизм инъективен тогда и только тогда, когда его ядро равно $\{0\}$.

Определение 1.5. Аддитивная подгруппа I кольца A называется левым (правым) идеалом, если для любых $r \in A$ и $x \in I$ имеет место включение $rx \in I$ (соотв., $xr \in I$). Другими словами, для левого идеала $AI = I$, а для правого – $IA = I$. Двусторонний идеал – это идеал, являющийся и левым, и правым.

Лемма 1.6. *Образ гомоморфизма R-алгебр является R-подалгеброй, а ядро – двусторонним идеалом.*

2. Факторкольцо и теорема о гомоморфизме

Так как гомоморфизм колец является в частности гомоморфизмом аддитивных подгрупп этих колец, то большинство свойств гомоморфизмов сохраняется. Роль нормальной подгруппы выполняет двусторонний идеал.

Теорема 2.1. *Для любого двустороннего идеала I кольца R существует кольцо A и эпиморфизм $\pi : R \to A$, ядро которого равно I.*

Доказательство. Так как I – подгруппа аддитивной группы кольца, то можно рассмотреть факторгруппу R/I. Зададим на ней умножение по формуле $(r + I) \cdot (s + I) = rs + I$, где $r, s \in R$ (это не является равенством множеств, для множеств выполнено только включение $(r + I) \cdot (s + I) \subseteq rs + I$). Если $r + x \in r + I$ и $s + y \in s + I$ – другие представители тех же смежных классов (т. е. $x, y \in I$), то $(r + x)(s + y) = rs + (ry + xs + xy) \in rs + I$. Поэтому определение корректно, т. е. не зависит от выбора представителей смежных классов. Тот факт, что операции в R/I удовлетворяют свойствам кольца, сразу следует из соответствующих свойств кольца R. Наконец, отображение π задается так же, как и для групп, где уже проверено, что оно сохраняет сложение, найдено ядро π и отмечено, что это отображение сюръективно. Осталось проверить, что π сохраняет умножение, но это сразу следует из определения произведения смежных классов. □

Определение 2.2. Кольцо $A = R/I$, построенное в доказательстве, называется факторкольцом R по I, а отображение $\pi = \pi_I$ – канонической проекцией или гомоморфизмом редукции по модулю I.

Для двустороннего идеала I кольца R определено отношение “сравнение по модулю I”, которое в соответствии с обсуждением в параграфе 4 является отношением эквивалентности. Доказательства утверждений настоящего параграфа показывают, что сравнения можно складывать и умножать, например, если $a \equiv b \mod I$, а $c \equiv d \mod I$, то $ac \equiv bd \mod I$. Таким образом, сравнения – это просто другая запись вычислений в факторгруппе или факторкольце.

Доказательства следующих утверждений про кольца очень похожи на доказательства для групп (на самом деле единственное, что надо проверить по сравнению с предыдущими доказательствами, это то, что отображение g сохраняет умножение, т. е. не просто является гомоморфизмом аддитивных групп, а и гомоморфизмом колец).

Теорема 2.3. *Пусть R и R' – кольца, I – двусторонний идеал в R, а $f : R \to R'$ – гомоморфизм. Если $I \subseteq \operatorname{Ker} f$, то существует единственный гомоморфизм $g : R/I \to R'$ такой, что $f = g \circ \pi$. Если $\operatorname{Ker} f = I$, то g инъективен. Если f сюръективен, то и g сюръективен.*

Следствие 2.4 (Теорема о гомоморфизме колец)**.** *Пусть $f : R \to R'$ – гомоморфизм колец. Тогда $\operatorname{Im} f \cong R/\operatorname{Ker} f$.*

3. Примеры идеалов и гомоморфизмов колец

$\mathbb{Z}/n\mathbb{Z} \cong \mathbb{Z}_n$.
$F[t]/(t-1) \cong F$.
$F[t]/(t^2-1) \cong F \oplus F$.
$F[t]/(t^2)$ – кольцо усеченных многочленов или кольцо двойных чисел.
$\mathbb{R}[t]/(t^2+1) \cong \mathbb{C}$.
$F[x,y]/(x,y) \cong F$.
Если $M = (x,y) := xF[x,y] + yF[x,y]$, то $M^n = (x^n, x^{n-1}y, \ldots, y^n)$, где M^n – идеал, порожденный всевозможными произведениями $a_1 \ldots a_n$, где $a_1, \ldots, a_n \in M$. Таким образом в этом кольце есть идеалы, порожденные любым конечным количеством элементов.
Пусть F – поле.

Упражнение 3.1. В кольце $\mathrm{M}_n(F)$ нет нетривиальных двусторонних идеалов.

Пусть G – группа. $FG := \{\sum\limits_{g\in G} \alpha_g g \mid \alpha_g \in F,$ почти все $\alpha_g = 0\}$ – векторное пространство над F с базисом G. Введем умножение на FG по правилу $\sum\limits_{g\in G} \alpha_g g \sum\limits_{h\in G} \beta_h h = \sum\limits_{g,h\in G} \alpha_g\beta_h gh$. Полученная F-алгебра называется групповой алгеброй группы G над полем F.
$I := \{\sum\limits_{g\in G} \alpha_g g \mid \sum\limits_{g\in G} \alpha_g = 0\}$ – пополняющий идеал.[2] $FG/I \cong F$.

Пусть $F = \mathbb{C}$, а $G = S_3$. $\varepsilon : S_3 \to C_2$ – четность перестановки. Здесь $C_2 = \{1, t\}$ – циклическая группа, т. е. $t^2 = 1$. Любой гомоморфизм групп $\varphi : G \to H$ индуцирует гомоморфизм групповых алгебр $\widetilde{\varphi} : FG \to FH$, $\sum\limits_{g\in G} \alpha_g g \mapsto \sum\limits_{g\in G} \alpha_g \varphi g$. $FC_2 = F[t]/(t^2-1) \cong F \oplus F$.

$\operatorname{Ker}\widetilde{\varepsilon} = \{\sum\limits_{g\in S_3} \alpha_g g \mid \sum\limits_{g\in A_3} \alpha_g = \sum\limits_{g\notin A_3} \alpha_g = 0\}$.
$FS_3/\operatorname{Ker}\widetilde{\varepsilon} \cong F \oplus F$.
На самом деле $\operatorname{Ker}\widetilde{\varepsilon} \cong \mathrm{M}_2(\mathbb{C})$ как кольца с 1.

Упражнение 3.2. Найти единицу кольца $\operatorname{Ker}\widetilde{\varepsilon}$.

В главе “Теория представлений” мы узнаем, что $\mathbb{C}S_3 \cong \mathbb{C} \oplus \mathbb{C} \oplus \mathrm{M}_2(\mathbb{C})$.

4. Комплексные числа

Определение 4.1. Факторкольцо $\mathbb{C} = \mathbb{R}[t]/(t^2+1)$ называется полем комплексных чисел (тот факт, что это поле, мы скоро проверим).

Композиция отображений $\mathbb{R} \hookrightarrow \mathbb{R}[t] \twoheadrightarrow \mathbb{C}$ является гомоморфизмом колец с 1, а так как $\mathbb{R}$ – поле, то она инъективна (ее ядро – идеал в $\mathbb{R}$, поэтому оно тривиально). Будем отождествлять элементы поля $\mathbb{R}$ с их образами под действием этого мономорфизма и считать, что $\mathbb{R}$ – подполе в $\mathbb{C}$.

Обозначим через i смежный класс $t + (t^2+1)\mathbb{R}[t]$. Заметим, что $i^2 + 1 = t^2 + 1 + (t^2+1)\mathbb{R}[t] = 0$ (имеется в виду ноль поля $\mathbb{C}$), откуда $i^2 = -1$. Так как в любом смежном классе $p(t) + (t^2+1)\mathbb{R}[t]$ есть единственный многочлен степени $\leqslant 1$ (остаток от деления на t^2+1), то любой элемент поля $\mathbb{C}$ может быть однозначно записан в виде $a + bi$ для некоторых $a, b \in \mathbb{R}$.

[2] Терминология возникшая в гомологической алгебре, в настоящий момент трудно объяснить, что этот идеал пополняет.

Ясно, что сложение определено по правилу

$$(a+bi)+(c+di)=(a+c)+(b+d)i.$$

Учитывая равенство $i^2=-1$ получаем формулу умножения в $\mathbb{C}$:

$$(a+bi)(c+di)=(ac-bd)+(ad+bc)i.$$

Таким образом, наше определение совпадает с классическим. Пусть $x,y\in\mathbb{R}$, а $z=x+iy$. Тогда $x=\operatorname{Re}z$ называется вещественной частью, а $y=\operatorname{Im}z$ – мнимой частью числа z. Число $\overline{z}=x-iy$ называется комплексно сопряженным к z. Из определения сразу следует, что $z+\overline{z}$, $z\overline{z}\in\mathbb{R}$, а также $z\in\mathbb{R}\iff z=\overline{z}$. Как мы узнаем позже, из неприводимости многочлена t^2+1 в $\mathbb{R}[t]$ следует, что $\mathbb{C}$ является полем. Однако нетрудно явно найти мультипликативный обратный к любому ненулевому элементу. Для этого достаточно просто домножить числитель и знаменатель на сопряженное к знаменателю:

$$\frac{1}{a+bi}=\frac{a-bi}{(a+bi)(a-bi)}=\frac{a}{a^2+b^2}+\frac{-b}{a^2+b^2}i.$$

Следующее утверждение проверяется непосредственно.

Лемма 4.2. *Отображение $\mathbb{C}\to\mathbb{C}$, отображающее z в $\overline{z}$ является автоморфизмом поля $\mathbb{C}$.*

Так как комплексное число $a+bi$ однозначно определяется парой вещественных чисел (a,b), то его удобно изображать точкой на плоскости с координатами (a,b) или ее радиус-вектором. Ясно, что сумма комплексных чисел изображается суммой векторов, соответствующих слагаемым. Произведение также имеет некоторый геометрический смысл. Назовем модулем комплексного числа длину вектора, который его изображает, а его аргументом – тригонометрический угол между положительным направлением вещественной оси и этим вектором.

Другими словами,

$$|a+bi|=\sqrt{a^2+b^2},$$
$$\operatorname{Arg}(a+bi)=\arg(a+bi)+2\pi\mathbb{Z},\ \text{где}$$
$$\arg(a+bi)=\begin{cases}\operatorname{arctg}(b/a), & a>0\\ \operatorname{arctg}(b/a)+\pi, & a<0\\ \pi/2, & a=0,\ b>0\\ -\pi/2, & a=0,\ b<0\end{cases}.$$

Обратите внимание, что тригонометрический угол определен с точностью до целого кратного 2π, т. е. принимает значения в аддитивной группе $\mathbb{R}/2\pi\mathbb{Z}$. Пусть $r=|a+bi|$, а $\varphi=\operatorname{Arg}(a+bi)$. Из определения синуса и косинуса следует, что $a=r\cos\varphi$, $b=r\sin\varphi$ и, следовательно,

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

Правая часть последнего равенства называется тригонометрической формой комплексного числа.

Перемножая комплексные числа в тригонометрической форме, и используя формулы для синуса и косинуса суммы, получим:

$$zw=|z|\cdot|w|\big(\cos(\operatorname{Arg}z+\operatorname{Arg}w)+i\sin(\operatorname{Arg}z+\operatorname{Arg}w)\big).$$

Другими словами, при перемножении комплексных чисел их модули перемножаются, а аргументы складываются. Геометрически это означается, что при умножении на w вектор, изображающий z, растягивается в $|w|$ раз и поворачивается на угол $\operatorname{Arg}w$. Из последней формулы для целого n получаем:

$$z^n=|z|^n\big(\cos(n\operatorname{Arg}z)+i\sin(n\operatorname{Arg}z)\big).$$

Последняя формула называется формулой Муавра.

Единственность представления комплексного числа в тригонометрической форме и формулу для произведения в тригонометрической форме можно выразить следующим образом:

$$\mathbb{C}^* \cong \mathbb{R}^*_{>0} \times \mathbb{R}/2\pi\mathbb{Z}. \tag{26}$$

Учитывая, что $\ln : \mathbb{R}^*_{>0} \to \mathbb{R}$ является изоморфизмом, получим следующий результат.

Предложение 4.3. $\mathbb{C}^* \cong \mathbb{R} \times \mathbb{R}/2\pi\mathbb{Z}$.

Доказательство. Отображения $z \mapsto (\ln|z|, \operatorname{Arg} z)$ и $(r, x) \mapsto e^r(\cos x + i \sin x)$ являются взаимно обратными гомоморфизмами. □

Так как $\mathbb{R}/2\pi\mathbb{Z} \cong \mathbb{R}/\mathbb{Z}$, можно получить более короткую запись: $\mathbb{C}^* \cong \mathbb{R} \times \mathbb{R}/\mathbb{Z}$, но она менее интуитивна (соответствует замене единицы измерения углов).

В курсе математического анализа вы узнаете, что тригонометрические функции и экспонента раскладываются в степенные ряды (ряды Тэйлора) следующим образом:

$$e^z = \sum_{n=0}^{\infty} \frac{z^n}{n!}, \qquad \sin t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k+1}}{(2k+1)!}, \qquad \cos t = \sum_{k=0}^{\infty} (-1)^k \frac{t^{2k}}{(2k)!}.$$

При этом из теории функций комплексной переменной известно, что если две функции $\mathbb{C} \to \mathbb{C}$ раскладываются в ряды, которые сходятся во всей комплексной плоскости, и совпадают на вещественной оси, то они равны. Поэтому указанные ряды разумно принять за определения комплексной экспоненты, синуса и косинуса. Подставляя $z = it$ в формулу для экспоненты, получим $e^{it} = \cos t + i \sin t$. Так как свойства экспоненты следуют из правила умножения рядов, для $t, u \in \mathbb{R}$ имеем

$$e^{u+it} = e^u(\cos t + i \sin t),$$

причем $e^u = |e^{u+it}|$, а $t \in \operatorname{Arg}(e^{u+it})$. В этих терминах предложение 4.3 можно переписать в виде:

$$\mathbb{C}^* \cong \mathbb{C}/(2\pi i\mathbb{Z}),$$

где изоморфизм справа налево задается экспонентой. Естественно, что обратную функцию называют логарифмом. Обратите внимание, что комплексный логарифм принимает значения не в $\mathbb{C}$, а в аддитивной группе $\mathbb{C}/(2\pi i\mathbb{Z})$.

Рассмотрим уравнение

$$z^n = w, \text{ где } w \in \mathbb{C}.$$

Пусть $z = re^{i\varphi}$, $w = se^{i\psi}$, где $r, s \in \mathbb{R}_{>0}$, а $\varphi, \psi \in \mathbb{R}/2\pi\mathbb{Z}$. Используя изоморфизм (26) получаем

$$z^n = w \iff r^n = s \,\&\, n\varphi = \psi \in \mathbb{R}/2\pi\mathbb{Z} \iff r = \sqrt[n]{s} \,\&\, \varphi = \frac{\psi}{n} + \frac{2\pi k}{n},$$

где $k \in \mathbb{Z}/n\mathbb{Z}$. Таким образом,

$$z = \sqrt[n]{|w|} e^{i\frac{\psi+2\pi k}{n}}, \text{ где } k \in \mathbb{Z}/n\mathbb{Z}.$$

Решения уравнения $z^n = w$ называются корнями из w. Обратите внимание, что символ $\sqrt[n]{w}$ обозначает множество всех корней из w. В частности, если $w = 1$, то $\sqrt[n]{1}$ является множеством всех корней из 1.

Так как возведение в n-ую степень является гомоморфизмом $\mathbb{C}^* \to \mathbb{C}^*$, то $\mu_n := \sqrt[n]{1} = \{e^{\frac{2\pi k}{n}i} \mid k \in \mathbb{Z}/n\mathbb{Z}\}$ является его ядром и, следовательно, подгруппой. Ясно, что μ_n циклическая порядка n. Формула $k \mapsto e^{\frac{2\pi k}{n}i}$ задает изоморфизм $\mathbb{Z}/n\mathbb{Z} \to \mu_n$. Образующие этой подгруппы называются первообразными корнями из 1. Другими словами, первообразный корень из 1 – это элемент группы $\mathbb{C}^*$ порядка n, в отличии от корней из 1 не являющихся первообразными, порядок которых делит n, но не равен ему. Число $e^{\frac{2\pi k}{n}i}$ является первообразным корнем из 1 тогда и только тогда, когда $k \in (\mathbb{Z}/n\mathbb{Z})^*$.

5. Порождение

Определение 5.1. Левым, правым или двусторонним идеалом, порожденным подмножеством X кольца R, называется наименьший левый, правый или двусторонний идеал, содержащий X. Левый (правый) идеал, порожденный подмножеством X, обозначается $\sum\limits_{x\in X} xR$ (соотв., $\sum\limits_{x\in X} Rx$). Если R – коммутативное кольцо, то идеал, порожденный подмножеством $X \subseteq R$, иногда обозначается (X). Левый или правый идеал, порожденный одним элементом, называется главным идеалом.

Так как пересечение идеалов снова является идеалом, то идеал, порожденный X, всегда существует. Действительно, это пересечение всех идеалов, содержащих X.

Лемма 5.2. *Левый (правый, двусторонний) идеал кольца R, порожденный X, состоит из всевозможных сумм элементов вида rx (соотв., xs, rxs), где $r, s \in R$, а $x \in X$.*

Пусть A – алгебра над коммутативным кольцом R. Подалгеброй, порожденным множеством X, называется наименьшая R-подалгебра в A, содержащая X. Для коммутативных алгебр подалгебра, порожденная X, чаще всего обозначается $R[X]$. Это обозначение совпадает с обозначением кольца многочленов с множеством переменных X, потому что его элементы и есть многочлены от элементов множества X с коэффициентами из R.

6. Прямая сумма алгебр

Пусть A и B – алгебры над коммутативным кольцом R. Прямой суммой $A \oplus B$ называется декартово произведение $A \times B$ с покомпонентными операциями:

$$(a,b)+(a',b')=(a+a',b+b'), \quad (a,b)(a',b')=(aa',bb'), \quad (a,b)r=(ar,br),$$

где $a, a' \in A$, $b, b' \in B$, $r \in R$. Нулем является элемент $(0,0)$. Если A и B – алгебры с 1, то $(1,1)$ – единица алгебры $A \oplus B$. В этом случае элемент $e=(1,0)$ удовлетворяет равенствам

$$e^2 = e \qquad \text{и} \qquad ex = xe \qquad \text{для любого } x \in A \oplus B.$$

Пусть C – кольцо. Элемент $e \in C$ называется идемпотентом, если $e^2 = e$. Идемпотент, лежащий в центре кольца, т. е. коммутирующий со всеми элементами кольца, называется центральным идемпотентом. Идемпотент, не равный 0 и 1, называется нетривиальным.

Предложение 6.1. *Алгебра C с единицей раскладывается в прямую сумму алгебр с единицей тогда и только тогда, когда она содержит нетривиальный центральный идемпотент.*

Доказательство. Пусть $e = e^2$ – нетривиальный идемпотент. Положим $f = 1 - e \neq 0$. Тогда $ef = 0$, откуда $Ce \neq C$ и $fC \neq C$. Кроме того, f также является идемпотентом. Если e лежит в центре, то и f обладает тем же свойством. Тогда $ece = ce^2 = ce$, $fcf = cf^2 = cf$, а $fce = 0$ для любого $c \in C$. Таким образом, e является единицей в кольце Ce, а f – в Cf. $c = ce + cf$, откуда $C = Ce + Cf$. Если $x = ce = c'f$, то $x = xe = c'fe = 0$, т. е. $Ce \cap Cf = \{0\}$. Также как и в теории групп, отсюда следует, что $C \cong Ce \oplus Cf$. Легко проверить, что этот изоморфизм сохраняет произведение и, следовательно, является изоморфизмом колец. Обратная импликация уже была доказана выше. □

Определение 6.2. Элемент a кольца R называется (левым) делителем нуля, если существует $b \in R \setminus \{0\}$ такой, что $ab = 0$. Кольцо называется областью целостности, если там нет нетривиальных делителей нуля (тривиальный делитель нуля – это ноль).

Заметим, что (левый) делитель нуля не может быть обратим (слева). Любое поле является областью целостности. Как мы узнаем позже, коммутативное кольцо является областью целостности тогда и только тогда, когда оно вкладывается в поле. Как мы видели, нетривиальный идемпотент всегда является делителем нуля. Таким образом, прямая сумма колец с 1 не может быть областью целостности.

7. Евклидовы кольца

В этом параграфе и далее все кольца коммутативны и имеют 1, если не оговорено противное.

Определение 7.1. Пусть R – область целостности. Предположим, что задана функция $f : R \to \mathbb{N} \cup \{-\infty\}$ обладающая следующими свойствами:

(1) $f(0) < f(r)$ для любого $r \in R \setminus \{0\}$.
(2) Для любых элементов a и $b \neq 0$ кольца R существуют $q, r \in R$ такие, что $a = bq + r$ и $f(r) < f(b)$.

Тогда R называется евклидовым кольцом с евклидовой нормой f.

Кольцо целых чисел является евклидовым кольцом с евклидовой нормой "модуль числа", а кольцо многочленов $F[t]$ над полем F – с нормой "степень многочлена" (степень 0 по определению считается равной $-\infty$).

Определение 7.2. Кольцо R называется кольцом главных идеалов, если любой идеал в R является главным (напомним, что это означает, что он имеет вид aR для некоторого $a \in R$). Область главных идеалов (ОГИ) – это область целостности, в которой любой идеал главный.

Теорема 7.3. *Евклидово кольцо является ОГИ.*

Доказательство. Пусть I – нетривиальный идеал в R. Возьмем ненулевой элемент $b \in I$ с наименьшей возможной евклидовой нормой. Пусть $a \in I$. Тогда существуют $q, r \in R$ такие, что $a = bq + r$ и $f(r) < f(b)$. Элемент $r = a - bq$ принадлежит I и его норма меньше, чем норма b. Следовательно, он должен быть равен нулю. Мы доказали, что произвольный элемент из I делится на b, поэтому $I \subseteq bR$. Обратное включение следует из того, что $b \in I$. □

Примеры евклидовых колец: $\mathbb{Z}$, $F[x]$, $\mathbb{Z}[i]$, $\mathbb{Z}[\sqrt{-2}]$. Кольцо $\{\frac{a+b\sqrt{-19}}{2} \mid a, b \in \mathbb{Z},\ a + b \in 2\mathbb{Z}\}$ является кольцом главных идеалов, но не является евклидовым. Доказательство обоих фактов нетривиально.

8. Китайская теорема об остатках

Пусть R – кольцо, а I и J – идеалы в R. Легко проверить, что сумма $I+J = \{a+b \mid a \in I,\ b \in J\}$ является идеалом, причем это наименьший идеал, содержащий $I \cup J$. В отличие от этого обычное произведение множеств I и J, т. е. $\{ab \mid a \in I,\ b \in J\}$ в общем случае не является идеалом, потому что не замкнуто относительно сложения. Поэтому произведением идеалов будем называть идеал IJ, порожденный элементами ab по всем $a \in I$ и $b \in J$. Другими словами,

$$IJ = \{\sum_{i=1}^{k} a_i b_i \mid k \in \mathbb{N},\ a_i \in I,\ b_i \in J\}.$$

Заметим, что по определению идеала $IJ \subseteq I \cap J$.

Определение 8.1. Идеалы I и J кольца R называются взаимно простыми, если $I + J = R$.

Лемма 8.2. *Если I и J взаимно простые идеалы, то $I \cap J = IJ$.*

Доказательство. Пусть $x \in I \cap J$. Так как I и J взаимно просты, то существуют $a \in I$ и $b \in J$ такие, что $a + b = 1$. Тогда $x = xa + xb \in (I \cap J)I + (I \cap J)J \subseteq IJ$. □

Теорема 8.3. *Если идеалы I и J взаимно просты, то $R/IJ \cong R/I \oplus R/J$.*

Доказательство. Естественный гомоморфизм $R \to R/I \oplus R/J$, $x \mapsto (x + I, x + J)$ имеет ядро $I \cap J = IJ$. Осталось доказать, что он сюръективен. Пусть $a + b = 1$ для некоторых $a \in I$ и $b \in J$. Тогда очевидно, что $xb + ya$ является прообразом элемента $(x + I, y + J)$. □

Лемма 8.4. *Если идеал J взаимно прост с каждым из идеалов $I_1, \ldots, I_n$, то он взаимно прост с их произведением.*

Доказательство. $R = J + I_1 = J + I_1 R = J + I_1(J + I_2) \subseteq (J + I_1 J) + I_1 I_2 \subseteq J + I_1 I_2$. Далее по индукции. □

Следствие 8.5 (китайская теорема об остатках)**.** *Если идеалы $I_1, \dots, I_n$ попарно взаимно просты, то*

$$R/(I_1 \cdots I_n) \cong R/I_1 \oplus \cdots \oplus R/I_n.$$

На самом деле, это была качественная формулировка К.Т.О. Количественная формулировка включает в себя формулу для вычисления элемента из $R/(I_1 \cdots I_n)$ по набору сравнений $x \equiv y_k \mod I_k$, $k = 1, \dots, n$.

Следствие 8.6. *Пусть идеалы $I_1, \dots, I_n$ попарно взаимно просты. Если $x \equiv y_k \mod I_k$, $k = 1, \dots, n$, то*

$$x \equiv \sum_{k=1}^{n} y_k c_k \mod I_1 \dots I_n, \text{ где } c_k \in (\prod_{j \neq k} I_j) \cap (1 + I_k)$$

(такие элементы c_k существуют, так как I_k взаимно прост с произведением остальных идеалов по лемме 8.4).

Замечание 8.7. Если R некоммутативно, то IJ надо заменить на $IJ + JI$.

9. Простые и максимальные идеалы

Определение 9.1. Собственный идеал I называется максимальным, если он не содержится ни в каком другом собственном идеале.

Лемма 9.2. *Для любого собственного идеала существует максимальный идеал, содержащий его.*

Доказательство. Пусть $I \triangleleft R$. Объединение линейно упорядоченного (по включению) набора собственных идеалов в R, содержащих I, является идеалом, содержащим I. Так как собственный идеал не содержит 1, то это объединение – собственный идеал. По лемме Цорна в множестве собственных идеалов, содержащих I, существует максимальный. Ясно, что он является максимальным и среди всех собственных идеалов. □

Определение 9.3. Собственный идеал I называется простым, если $ab \in I$ влечет $a \in I$ или $b \in I$.

Легко видеть, что кольцо является областью целостности тогда и только тогда, когда $\{0\}$ – простой идеал.

Лемма 9.4. *Прообраз простого идеала – простой. Прообраз максимального идеала при эпиморфизме – максимальный.*

Следствие 9.5. *Идеал I простой тогда и только тогда, когда R/I – область целостности.*
Идеал I максимальный тогда и только тогда, когда R/I – поле.
Любой максимальный идеал является простым.

10. Простые и неприводимые элементы

Определение 10.1. Элементы $a, b \in R$ называются ассоциированными, если $aR = bR$.
Необратимый элемент $a \in R$ называется неприводимым, если из равенства $a = bc$ следует, что b или c ассоциирован с a.

Заметим, что если aR – максимальный из собственных главных идеалов, то a неприводим. Обратное неверно, что показывает пример $p = 0$ в области целостности. В области целостности это – единственный пример, как показывает следующее утверждение.

Лемма 10.2. *Пусть R область целостности, а $c \in R$ – ненулевой необратимый элемент.*

(1) *Элементы $a, b \in R$ ассоциированы тогда и только тогда, когда $a = b\varepsilon$ для некоторого $\varepsilon \in R^*$.*
(2) *c неприводим, если он не раскладывается в произведение необратимых элементов.*
(3) *c неприводим тогда и только тогда, когда cR максимальный в множестве главных идеалов.*

Доказательство. 1. Пусть $aR = bR$. Если $a = 0$, то и $b = 0$ и утверждение очевидно. Иначе $a = b\varepsilon$ и $b = a\delta$ для некоторых $\varepsilon, \delta \in R$, откуда $a = a\delta\varepsilon$. Так как R область целостности, то можно сокращать на ненулевой элемент, следовательно, $1 = \delta\varepsilon$, т. е. $\varepsilon \in R^*$. Обратное утверждение очевидно.

2. Пусть $c = xy$. Если x или y обратим, то другой ассоциирован с c. Обратно, если, скажем, x ассоциирован с c, то из первого абзаца доказательства следует, что y обратим.

3. $cR \subseteq xR \iff c = xy \iff x \in R^* \vee y \in R^* \iff xR = R \vee xR = cR$. □

Определение 10.3. Необратимый элемент p кольца R называется простым, если pR – простой идеал. Другими словами, p простой $\iff (p|ab \implies p|a \vee p|b)$.

Лемма 10.4. *Любой простой элемент неприводим. Обратное, вообще говоря, неверно.*

Доказательство. Если p – простой, то

$$ab = p \implies ab \in pR \implies a \in pR \text{ или } b \in pR \implies aR = pR \text{ или } bR = pR,$$

т. е. p ассоциирован с a или с b.

Контрпример к обратному утверждению дает, например, кольцо $\mathbb{Z}[\sqrt{-3}]$, в котором $2 \cdot 2 = (1 + \sqrt{-3})(1 - \sqrt{-3})$. Действительно, элемент этого кольца обратим тогда и только тогда, когда квадрат модуля равен 1. Так как квадрат модуля каждого сомножителя равен 4, то он может раскладываться в произведение необратимых элементов только с квадратами модулей равными 2, а таких элементов в нашем кольце нет. Поэтому все сомножители являются неприводимыми, но ни один не является простым. □

Лемма 10.5. *Пусть R – область главных идеалов, а $p \in R \setminus \{0\}$. Тогда следующие условия эквивалентны.*

(1) *pR – максимальный идеал.*
(2) *pR – простой идеал (т. е. p – простой элемент).*
(3) *p неприводим.*

Доказательство. Импликации (1) $\implies$ (2) $\implies$ (3) доказаны в следствии 9.5 и лемме 10.4. Если p неприводим, то по лемме 10.2 pR максимальный в множестве собственных главных идеалов, а так как любой идеал в R является главным, то и в множестве всех собственных идеалов. □

11. Нетеровы кольца и разложение на неприводимые

Предложение 11.1. *Пусть R – (не обязательно коммутативное) кольцо с 1. Следующие условия эквивалентны.*

(1) *Любое линейно упорядоченное (по включению) множество правых идеалов содержит наибольший элемент.*
(2) *Любая возрастающая цепочка правых идеалов обрывается.*
(3) *Любой правый идеал порожден конечным числом элементов.*
(4) *Любой подмодуль конечно порожденного правого модуля конечнопорожден.*

Естественно, слово “правый” можно заменить на “левый”, или убрать совсем, если кольцо коммутативно.

Определение 11.2. Кольцо, удовлетворяющее условиям предыдущего предложения, называется нетеровым (справа).

Доказательство. Импликации (1) $\Longrightarrow$ (2) и (4) $\Longrightarrow$ (3) очевидны.

(2) $\Longrightarrow$ (3). Предположим, что идеал I не может быть порожден конечным числом элементов. По индукции найдем элементы $r_k \in I$, $k \in \mathbb{N}$ такие, что $r_{k+1} \notin I_k := \sum_{h=1}^{k} r_h R$. Тогда цепочка идеалов I_k строго возрастает и по условию (2) должна оборваться, т. е. для некоторого индекса m получим $I_m = I_k$ при всех $k > m$. Это, однако, противоречит построению I_k.

(3) $\Longrightarrow$ (1). Пусть $\mathscr{X}$ – линейно упорядоченное множество идеалов, а I – объединение всех идеалов из $\mathscr{X}$. По (3) I порожден конечным числом элементов $r_1, \ldots, r_n$. Каждый из r_k лежит в каком-то идеале $I_k \in \mathscr{X}$. Так как $\mathscr{X}$ линейно упорядочено, то среди I_k найдется наибольший, скажем, I_m. Тогда $r_1, \ldots, r_n \in I_m$, откуда $I_m = I$ – наибольший элемент в $\mathscr{X}$.

(3) $\Longrightarrow$ (4). Пусть $M = \langle x_1, \ldots, x_m \rangle_R$ – конечнопорожденный правый R-модуль, а $N \leqslant M$. Зададим эпиморфизм свободного R-модуля R^m на M по формуле $a \mapsto \sum_{i=1}^{m} x_i a_i$. Пусть N' – полный прообраз N под действием этого эпиморфизма. Ясно, что из конечнопорожденности N' следует конечнопорожденность N. Индукцией по m докажем, что N' конечнопорожден. Пусть $\pi : R^m \to R$ – проекция на последнюю компоненту, т. е. $\pi(a) = a_m$. Образ подмодуля – подмодуль, поэтому $\pi(N')$ правый идеал в R. По (3) этот идеал конечнопорожден, скажем, множеством $\{\alpha_1, \ldots, \alpha_n\}$. Выберем прообразы $y_i \in \pi^{-1}(\alpha_i) \in N'$. Тогда $N' = \langle y_1, \ldots, y_n \rangle + (N' \cap \operatorname{Ker} \pi)$. По индукционному предположению $N' \cap \operatorname{Ker} \pi$ конечнопорожден, как подмодуль свободного модуля $\operatorname{Ker} \pi \cong R^{m-1}$. Следовательно, и N' конечнопорожден. □

Теорема 11.3. *Любой необратимый элемент нетерова кольца раскладывается в произведение неприводимых.*

Доказательство. Пусть $r = r_1 \in R$ – необратимый элемент. Если r приводим, то существуют $r_2, r_3 \in R$ такие, что $r_1 = r_2 r_3$, $r_1 R \subsetneqq r_2 R$ и $r_1 R \subsetneqq r_3 R$. По индукции для каждого приводимого r_i найдем r_{2i} и r_{2i+1} такие, что $r_i = r_{2i} r_{2i+1}$, $r_i R \subsetneqq r_{2i} R$ и $r_i R \subsetneqq r_{2i+1} R$. Получим бинарное дерево. Каждая ветка этого дерева конечна за счет нетеровости кольца R. Следовательно, все дерево конечно (иначе строим бесконечную ветку, выбирая ребро, на котором висит бесконечное поддерево). Очевидно, что листья дерева неприводимы. По индукции нетрудно доказать, что r равно произведению всех листьев. □

12. Факториальность колец главных идеалов

Возвращаемся к изучению коммутативных колец.

Определение 12.1. Область целостности R называется факториальным кольцом (UFD), если любой ненулевой необратимый элемент раскладывается в произведение неприводимых единственным образом. Единственность понимается в следующем смысле: если $\prod_{i=1}^{m} p_i$ ассоциировано с $\prod_{j=1}^{n} q_j$ для некоторых неприводимых элементов $p_i, q_j \in R$, то $m = n$ и существует перестановка $\sigma \in S_n$ такая, что p_i ассоциирован с $q_{\sigma(i)}$ для всех $i = 1, \ldots, n$.

Задача настоящего параграфа – доказать, что область главных идеалов является факториальным кольцом.

Лемма 12.2. *Пусть R – область целостности, в которой любой элемент раскладывается в произведение неприводимых. Кольцо R является факториальным тогда и только тогда, когда каждый неприводимый элемент порождает простой идеал.*

Доказательство. Пусть $\varepsilon p_1 \cdots p_n = \theta q_1 \cdots q_m$, где все элементы p_k и q_k неприводимы, а ε, θ – обратимы. Индукцией по $\min(m, n)$ докажем, что $m = n$ и существует перестановка $\sigma \in S_n$ такая, что p_k ассоциирован с $q_{\sigma(k)}$ для всех k от 1 до n. База индукции: если $m = 0$, то правая часть обратима, поэтому $n = 0$.

Индукционный переход. По условию идеал p_nR простой. Поэтому найдется ℓ такое, что $q_\ell \in p_nR$. Так как q_ℓ неприводим, а R – область целостности, то по лемме 10.2 $q_\ell = \delta p_n$, где δ обратимо. Подставляя это в исходное равенство и сокращая на p_n получим $\varepsilon p_1 \cdots p_{n-1} = \theta\delta q_1 \cdots q_m/q_\ell$. По индукционному предположению $n-1 = m-1$ и существует биекция $\tau : \{1,\dots,n-1\} \to \{1,\dots,m\} \setminus \{\ell\}$ такая, что p_k ассоциирован с $q_{\tau(k)}$ для всех k от 1 до $n-1$. Положив $\sigma(k) = \tau(k)$ при всех k от 1 до $n-1$, а $\sigma(n) = \ell$, получаем результат.

Обратно, пусть p – неприводимый элемент, а $ab = pc \in pR$. Раскладывая a, b, c на неприводимые множители получим $a_1 \dots a_i b_1 \dots b_j = pc_1 \dots c_h$. По единственности разложения один из a_k или b_k ассоциирован с p, а из этого следует, что a или b делится на p. □

Так как область главных идеалов нетерова (любой идеал порожден конечным числом элементом), то из предыдущей леммы, теоремы 11.3 и леммы 10.5 непосредственно вытекает следующий факт.

Теорема 12.3. *Область главных идеалов является факториальным кольцом.*

В будущем мы докажем, что кольцо многочленов над факториальным кольцом также является факториальным, так что области главных идеалов – это далеко не все факториальные кольца.

Кольцо $\mathbb{Z}[\sqrt{-3}]$ не является факториальным кольцом. Действительно, $(1+\sqrt{-3})(1-\sqrt{-3}) = 2 \cdot 2$ является примером неоднозначного разложения на неприводимые множители.

13. Наибольший общий делитель

Определение 13.1. Пусть $a, b \in R$. Элемент d кольца R называется наибольшим общим делителем элементов a и b, если он делит и a, и b, и делится на любой другой общий делитель a и b.

Другими словами, d – наибольший общий делитель, если dR – наименьший главный идеал, содержащий a и b. В этом нет ничего удивительного, потому что отношение делимости на множестве элементов кольца и отношение "$\subseteq$" на множестве главных идеалов совпадают, т. е. x делится на y тогда и только тогда, когда $xR \subseteq yR$.

Наибольший общий делитель a и b обозначается через $\gcd(a,b)$. Как следует из последней формулировки, $\gcd(a,b)$ определен с точностью до ассоциированности. Заметим, что идеал содержит a и b тогда и только тогда, когда он содержит идеал $aR + bR$.

НОД не всегда существует. Задействуя пример неоднозначного разложения на множители в $\mathbb{Z}[\sqrt{-3}]$ легко показать, что 4 и $2(1+\sqrt{-3})$ делятся на 2 и на $1+\sqrt{-3}$, но не делятся одновременно ни на какое их общее кратное.

Предложение 13.2. *У любых двух элементов факториального кольца R существует наибольший общий делитель.*

Доказательство. Если $a = 0$, то $\gcd(a,b) = b$. Пусть $a = \delta p_1^{k_1} \dots p_n^{k_n} \neq 0$, и $b = \varepsilon p_1^{m_1} \dots p_n^{m_n} \neq 0$, где $k_i, m_i \in \mathbb{N}_0$, а $\delta, \varepsilon \in R^*$, – разложение элементов $a, b \in R$ на неприводимые. Тогда $\gcd(a,b) = p_1^{\min(k_1,m_1)} \dots p_n^{\min(k_n,m_n)}$. □

Теорема 13.3 (о линейном представлении НОД.)**.** *Пусть R – кольцо главных идеалов. Для любых $a, b \in R$ существуют $x, y \in R$ такие, что $ax + by = \gcd(a,b)$.*

Доказательство. Идеал $aR + bR$ является минимальным идеалом, содержащим a и b, а по условию он является главным. Таким образом, $aR + bR = dR$, и по определению НОД $d = \gcd(a,b)$. □

Следствие 13.4. *Пусть R – кольцо главных идеалов. Идеалы aR и bR являются взаимно простыми, если у элементов a и b нет необратимых общих делителей (такие элементы называются взаимно простыми).*

Для нахождения НОД в евклидовом кольце используется алгоритм Евклида. Он использует следующую лемму.

Лемма 13.5. *Для любых $a, b, c \in R$ имеет место равенство* $\gcd(a, b) = \gcd(a - bc, b)$.

Доказательство. Ясно, что $a-bc$ и b содержатся в идеале $aR+bR$, поэтому $(a-bc)R+bR \subseteq aR+bR$. С другой стороны, $a = (a-bc)+bc \in (a-bc)R+bR$, откуда следует обратное включение. Так как $(a - bc)R + bR = aR + bR$, то и наименьший главный идеал, содержащий эти идеалы, одинаковый. □

Алгоритм Евклида используется для нахождения наибольшего общего делителя двух элементов a, b евклидова кольца R, а также его линейного представления (обратный ход).

Алгоритм Евклида. Обозначим $r_0 = a$ и $r_1 = b$ и положим $i = 1$.

(1) Разделить r_{i-1} на r_i с остатком: $r_{i-1} = r_i q_i + r_{i+1}$.
(2) Если $r_{i+1} \neq 0$, то увеличить i и вернуться к первому шагу.
(3) Если на k-ом круге $r_{k+1} = 0$, то $\gcd(a, b) = r_k$.

Действительно, так как $r_{i+1} = r_{i-1} - r_i q_i$, то по предыдущей лемме $\gcd(r_{i-1}, r_i) = \gcd(r_{i+1}, r_i)$, а $\gcd(r_k, 0) = r_k$.

Обратный ход алгоритма Евклида.

(1) $\gcd(a, b) = r_k = r_{k-2}x_{k-2} + r_{k-1}y_{k-2}$, где $x_{k-2} = 1$, а $y_{k-2} = -q_{k-1}$.
(2) Подставляя $r_{i+1} = r_{i-1} - r_i q_i$ в равенство $\gcd(a, b) = r_i x_i + r_{i+1} y_i$ получаем выражение $\gcd(a, b) = r_{i-1}x_{i-1} + r_i y_{i-1}$, где $x_{i-1} = y_i$, а $y_{i-1} = x_i - y_i q_i$.
(3) В результате цикла по $i = k - 2, \ldots, 1$ получаем $\gcd(a, b) = r_0 x_0 + r_1 y_0 = ax_0 + by_0$.

Аналогичное НОД понятие с обращением включений – это наименьшее общее кратное (НОК).

Определение 13.6. Пусть $a, b \in R$. Элемент c кольца R называется наименьшим общим кратным элементов a и b, если он делится на a и на b, и делит любое другое общее кратное a и b.

Другими словами, c – наименьшее общее кратное, если cR – наибольший главный идеал, содержащийся в $aR \cap bR$. Наименьшее общее кратное элементов a и b обозначается через $\operatorname{lcm}(a, b)$.

Лемма 13.7. *Если R – факториальное кольцо, $a, b \in R \setminus \{0\}$, то* $\operatorname{lcm}(a, b) = \frac{ab}{\gcd(a,b)}$.

Доказательство. Пусть $a = p_1^{k_1} \ldots p_n^{k_n}$, а $b = p_1^{m_1} \ldots p_n^{m_n}$, где $k_i, m_i \in \mathbb{N}_0$, – разложение элементов $a, b \in R$ на неприводимые. Тогда $\gcd(a, b) = p_1^{\min(k_1,m_1)} \ldots p_n^{\min(k_n,m_n)}$, а $\operatorname{lcm}(a, b) = p_1^{\max(k_1,m_1)} \ldots p_n^{\max(k_n,m_n)}$. Результат следует теперь из равенства $\max(k_i, m_i) = k_i + m_i - \min(k_i, m_i)$. □

14. Локализация

Идея состоит в том, чтобы обратить некоторый набор элементов универсальным образом. Заметим, что обратимый элемент не может быть делителем нуля. Поэтому, если мы хотим обратить хоть один делитель нуля, то все элементы, которые в произведении с ним дают 0, должны обратиться в 0. Заметим также, что если два элемента стали обратимыми, то обратимым стало и их произведение. Так как 1 уже обратима, то включение 1 в наше множество ничего не изменит. Поэтому мы будем говорить только об обращении элементов из некоторого мультипликативно замкнутого подмножества с 1. Другими словами, наше множество всегда будет мультипликативным моноидом, содержащим 1 кольца. Коротко такое множество называется мультипликативным.

В анализе и алгебраической геометрии рассматривают F-алгебру A (аналитических или полиномиальных) функций на многообразии. Множество M функций, обращающихся в 0 в данной точке x, является максимальным идеалом, как ядро гомоморфизма $A \to F$, $f \mapsto f(x)$. Локализация в мультипликативном подмножестве $A \setminus M$ отражает свойства многообразия в окрестности точки x, отсюда произошел этот термин.

Определение 14.1. Пусть S – мультипликативное подмножество кольца R. Локализацией кольца R в S называется кольцо $S^{-1}R$ вместе с локализационным гомоморфизмом $\lambda_S : R \to S^{-1}R$ удовлетворяющее следующим свойствам.

(1) Для любого $s \in S$ элемент $\lambda_S(s)$ обратим в $S^{-1}R$.
(2) Для любого гомоморфизма $\varphi : R \to A$, при котором $\varphi(s) \in A^*$ для всех $s \in S$, существует единственный гомоморфизм $\psi : S^{-1}R \to A$ такой, что $\psi \circ \lambda_S = \varphi$.

Это определение ничего не говорит о существовании локализации. На самом деле она всегда существует и, как и все универсальные конструкции, единственна с точностью до единственного изоморфизма.

Определим отношение "$\sim$" на множестве $R \times S$ по следующему правилу:

$$(r_1, s_1) \sim (r_2, s_2) \iff \exists s \in S : \ ss_2r_1 = ss_1r_2.$$

Проверим, что "$\sim$" является отношением эквивалентности. Рефлексивность и симметричность очевидны. Пусть $(r_1, s_1) \sim (r_2, s_2) \sim (r_3, s_3)$, т. е. $sr_1s_2 = sr_2s_1$ и $s'r_2s_3 = s'r_3s_2$ для некоторых $s, s' \in S$. Домножая первое равенство на $s's_3$, а второе – на ss_1, получим $s'ss_2r_1s_3 = s'sr_2s_1s_3 = ss's_2r_3s_1$. Так как $ss's_2 \in S$, то $(r_1, s_1) \sim (r_3, s_3)$, что доказывает транзитивность.

Положим $S^{-1}R = R \times S/\sim$. Класс эквивалентности, содержащий (r, s) обозначается $\frac{r}{s}$. Определим отображение $\lambda_S : R \to S^{-1}R$ формулой $\lambda_S(r) = \frac{r}{1}$.

Теорема 14.2. *Пусть S – мультипликативное подмножество кольца R. Определим операции на $S^{-1}R$ следующими формулами:*

$$\frac{r_1}{s_1} \cdot \frac{r_2}{s_2} = \frac{r_1r_2}{s_1s_2} \qquad \text{и} \qquad \frac{r_1}{s_1} + \frac{r_2}{s_2} = \frac{s_1r_2 + s_2r_1}{s_1s_2}.$$

Тогда $S^{-1}R$ является локализацией кольца R в мультипликативном подмножестве S с локализационным гомоморфизмом λ_S.

Доказательство. Докажем, что наше определение сложения и умножения не зависит от выбора представителей классов эквивалентности. Пусть $\frac{r_1'}{s_1'} = \frac{r_1}{s_1}$ и $\frac{r_2'}{s_2'} = \frac{r_2}{s_2}$, т. е. $sr_1s_1' = sr_1's_1$ и $s'r_2s_2' = s'r_2's_2$ для некоторых $s, s' \in S$. Перемножая последние равенства получаем $ss'r_1s_1'r_2s_2' = ss'r_1's_1r_2's_2$, откуда $\frac{r_1r_2}{s_1s_2} = \frac{r_1'r_2'}{s_1's_2'}$. Далее,

$$ss'(r_1s_2 + r_2s_1)s_1's_2' = ss'(r_1s_2s_1's_2' + r_2s_1s_1's_2') = ss'(r_1's_2s_1s_2' + r_2's_1s_1's_2) = ss'(r_1's_2' + r_2's_1')s_1s_2,$$

что доказывает равенство $\frac{r_1s_2+r_2s_1}{s_1s_2} = \frac{r_1's_2'+r_2's_1'}{s_1's_2'}$.

Непосредственно проверяется, что заданные операции коммутативны, ассоциативны, и выполнена дистрибутивность. Проверим для примера ассоциативность сложения (самое длинное вычисление).

$$\left(\frac{r_1}{s_1} + \frac{r_2}{s_2}\right) + \frac{r_3}{s_3} = \frac{r_1s_2 + r_2s_1}{s_1s_2} + \frac{r_3}{s_3} = \frac{r_1s_2s_3 + r_2s_1s_3 + r_3s_1s_2}{s_1s_2s_3}$$
$$\frac{r_1}{s_1} + \left(\frac{r_2}{s_2} + \frac{r_3}{s_3}\right) = \frac{r_1}{s_1} + \frac{r_2s_3 + r_3s_2}{s_2s_3} = \frac{r_1s_2s_3 + r_2s_3s_1 + r_3s_2s_1}{s_1s_2s_3}.$$

Нейтральным элементом по сложению является $\frac{0}{1} = \frac{0}{s}$, обратным к $\frac{r}{s}$ – $\frac{-r}{s}$. Мультипликативно нейтральным является $\frac{1}{1} = \frac{s}{s}$. Сразу видно, что λ_S – гомоморфизм. Для $s \in S$ имеем $\lambda_S(s) \cdot \frac{1}{s} = \frac{s}{1} \cdot \frac{1}{s} = \frac{s}{s} = 1$, так что первое свойство локализации выполнено.

Пусть теперь $\varphi : R \to A$ – гомоморфизм из второго свойства определения 14.1. Определим отображение $\psi : S^{-1}R \to A$ равенством $\psi(\frac{r}{s}) = \varphi(r)\varphi(s)^{-1}$. Если $\frac{r'}{s'} = \frac{r}{s}$, то $s''r's = s''rs'$ для некоторого $s'' \in S$, и $\varphi(s'')\varphi(r')\varphi(s) = \varphi(s'')\varphi(r)\varphi(s')$. Домножая на $\varphi(s'')^{-1}\varphi(s)^{-1}\varphi(s')^{-1}$ получаем $\varphi(r')\varphi(s')^{-1} = \varphi(r)\varphi(s)^{-1}$, что доказывает корректность определения ψ. Учитывая, что $\varphi(1) = 1$, из определения сразу следует, что $\varphi = \psi \circ \lambda_S$. Легко проверить, что ψ является гомоморфизмом.

Равенство $\varphi = \psi \circ \lambda_S$ однозначно задает, $\psi(\frac{r}{1}) = \varphi(r)$. Так как ψ должен быть гомоморфизмом, то

$$\varphi(r) = \psi(\frac{r}{1}) = \psi(\frac{r}{s} \cdot \frac{s}{1}) = \psi(\frac{r}{s}) \cdot \varphi(s).$$

Учитывая, то $\varphi(s)$ по условию обратимо, получаем $\psi(\frac{r}{s}) = \varphi(r)\varphi(s)^{-1}$. Таким образом, ψ однозначно определяется своими свойствами. □

Приведем несколько часто используемых примеров мультипликативных подмножеств и локализаций в них.

(1) Для $s \in R$ положим $\langle s \rangle = \{s^n \mid n \in \mathbb{N}_0\}$. Локализация $\langle s \rangle^{-1}R$ обозначается через R_s и называется главной локализацией в элементе s (по аналогии с главным идеалом).

Упражнение 14.3. Докажите, что $R_s \cong R[t]/(st - 1)$.

(2) Если P – простой идеал кольца R, то $R \setminus P$ является мультипликативным подмножеством. В этом случае локализация $R_P := (R \setminus P)^{-1}R$ называется локализацией кольца R в простом идеале P.

R_P является локальным кольцом, т. е. кольцом с единственным максимальным идеалом (равносильно: множество необратимых элементов является аддитивной подгруппой).

(3) S – множество всех элементов в R, не являющихся делителями 0 (сам 0 является делителем 0). Тогда $S^{-1}R$ называется полным кольцом частных кольца R. Это максимальная локализация, для которой гомоморфизм локализации инъективен.

(4) $R = K[x]$, где K – кольцо, S – множество унитальных многочленов.

15. Поле частных и разложение на простейшие дроби

Если R – область целостности, то $\{0\}$ является простым идеалом. Локализация в этом идеале, очевидно, будет полем, которое называется полем частных кольца R. Другими словами, поле частных – это полное кольцо частных области целостности. Локализационный гомоморфизм в этом случае – универсальное вложение R в поле в следующем смысле.

Лемма 15.1. *Пусть R – область целостности, а $S = R \setminus \{0\}$. Тогда $F = S^{-1}R$ является полем, а гомоморфизм локализации $\lambda_S : R \rightarrowtail F$ инъективен. При этом λ_S удовлетворяет следующему универсальному свойству: для любого поля K и мономорфизма $\varphi : R \rightarrowtail K$ существует единственный мономорфизм $\psi : F \rightarrowtail K$ такой, что $\varphi = \psi \circ \lambda_S$.*

Обычно мы отождествляем элементы R с их образами в поле частных F под действием гомоморфизма локализации и считаем, что $R \subseteq F$.

Пусть теперь R – область главных идеалов. В этом случае поле частных кольца R аддитивно порождено дробями, знаменатели которых являются степенями неприводимых элементов (первая часть доказательства следующей теоремы). Для евклидовых колец можно еще ограничить евклидову норму числителя.

Определение 15.2. Пусть R – евклидово кольцо с евклидовой нормой f, а F – его поле частных. Простейшей дробью называется элемент $\frac{r}{s^n} \in F$, где $r, s \in R$, s – неприводим, и $f(r) < f(s)$.

Теорема 15.3. *Пусть R – евклидово кольцо с нормой f, а F – его поле частных. Любой элемент из F представляется в виде суммы элемента из R и простейших дробей.*

Доказательство. Разложим сначала $\frac{a}{bc}$, $a, b, c \in R$, $\gcd(b, c) = 1$ в сумму дробей со знаменателями b и c. По теореме о линейном представлении НОД существуют такие $x, y \in R$, что $1 = bx + cy$. Тогда $\frac{a}{bc} = \frac{abx+acy}{bc} = \frac{ax}{c} + \frac{ay}{b}$. По индукции легко доказать, что любая дробь со знаменателем $p_1^{k_1} \cdots p_m^{k_m}$, где $p_1, \ldots, p_m \in R$ – неприводимые элементы, раскладывается в сумму дробей $\sum \frac{r_i}{p_i^{k_i}}$ (заметим, что до сих пор мы пользовались только тем, что R – кольцо главных идеалов).

Для завершения доказательства осталось показать, что любая дробь $\frac{r}{p^k}$, $r, p \in R$, p неприводим, раскладывается в сумму простейших и элемента из R. Докажем это индукцией по k. При $k = 0$ наша дробь лежит в R и доказывать нечего. Пусть $k > 0$. Разделим с остатком r на p: $r = sp + q$, где $f(q) < f(p)$. Тогда $\frac{r}{p^k} = \frac{s}{p^{k-1}} + \frac{q}{p^k}$. Вторая дробь является простейшей, а первая раскладывается в сумму простейших и элемента из R по индукционному предположению. □

16. Многочлены от одной переменной

Пусть F – коммутативное кольцо с 1.

Определение 16.1. Многочленом p от одной переменной t над F называется конечный набор элементов поля $(\alpha_0, \dots, \alpha_n)$, записанный в виде $p(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_n t^n$. При этом $n = \deg p$ называется степенью многочлена p.

Многочлены складываются покомпонентно (при этом отсутствующие компоненты считаются равными 0), а перемножаются по правилу свертки: коэффициент при t^k произведения pq, где $q(t) = \beta_0 + \beta_1 t + \dots + \beta_m t^m$, равен

$$\sum_{0 \leqslant i \leqslant n,\ 0 \leqslant j \leqslant m,\ i+j=k} \alpha_i \beta_j.$$

Множество всех многочленов с операциями сложения и умножения является алгеброй над F и обозначается $F[t]$. Пусть A – другая алгебра над F. Определим полиномиальную функцию $\tilde{p} : A \to A$ формулой $\tilde{p}(a) = \alpha_0 + \alpha_1 a + \dots + \alpha_n a^n$. Допуская вольность речи, будем обозначать полиномиальную функцию тем же символом, что и многочлен, т. е. писать p вместо $\tilde{p}$.

Последнее соглашение общепринято, но не так безобидно, как кажется на первый взгляд. Например, если $A = F = \mathbb{F}_2$, то $p(t) = t$ и $q(t) = t^2$ являются разными многочленами, но задают одну и ту же полиномиальную функцию. Как мы увидим, это невозможно для бесконечного поля F.

Пусть A – алгебра над F, а $a \in A$. Операции над многочленами специально определены так, чтобы отображение

$$\varepsilon_a : F[t] \to A, \qquad \varepsilon_a(p) = p(a)$$

являлось гомоморфизмом F-алгебр. Он называется гомоморфизмом подстановки или вычисления значения в точке a.

Предложение 16.2 (универсальное свойство кольца многочленов)**.** *Для любого $a \in A$ существует единственный гомоморфизм F-алгебр $\varepsilon : F[t] \to A$, отображающий t в a.*

Доказательство. Существование: $\varepsilon = \varepsilon_a$.

Из того, что ε гомоморфизм F-алгебр мы знаем образы всех элементов из F под действием ε, а $\varepsilon(t)$ задано. Так как ε сохраняет сумму и произведение, а любой многочлен получается из t и элементов кольца F при помощи этих операций, образ любого многочлена под действием ε определен однозначно. □

Далее F является полем.

Теорема 16.3. *Кольцо многочленов $F[t]$ над полем F является евклидовым кольцом с евклидовой нормой* deg.

Заметим, что в отличии от целых чисел с евклидовой нормой “модуль”, деление с остатком в кольце многочленов с евклидовой нормой deg единственно.

Теорема 16.4 (теорема Безу)**.** *Пусть $\alpha \in F$, а $p \in F[t]$, где F – поле. Остаток от деления многочлена p на $t - \alpha$ равен $p(\alpha)$.*

Элемент α является корнем многочлена p тогда и только тогда, когда p делится на $t - \alpha$.

Многочлен степени n не может иметь больше, чем n корней.

Следующее утверждение вытекает из последней фразы предыдущего и критерия цикличности абелевой группы. Оно, в часности, играет важную роль в классификации конечных полей.

Теорема 16.5. *Любая конечная подгруппа мультипликативной группы поля циклическая.*

Доказательство. Пусть F – поле, $G \leqslant F^*$, а $|G| = n$. Пусть $k = \exp G$. Это означает, что многочлен $t^k - 1$ имеет $\geqslant n$ корней (все элементы группы G – его корни). По теореме 16.4 $n \leqslant k$. С другой стороны, по лемме 10.3 n делится на k, откуда $n = k$. По той же самой лемме группа G циклическая. □

В случае, когда мы рассматриваем сравнения в кольце $F[t]$ по модулю многочленов первой степени, китайская теорема об остатках превращается в интерполяционную формулу Лагранжа. Конечно, эту формулу легко проверить и непосредственно, связь ее с китайской теоремой об остатках скорее позволяет лучше понять доказательство самой китайской теоремы.

Теорема 16.6. *Пусть $t_0, y_0, \ldots, t_n, y_n \in F$, причем $t_i \neq t_j$ при $i \neq j$. Существует единственный многочлен p степени не выше n, удовлетворяющий условиям $p(t_i) = y_i$ для любого $i = 0, \ldots, n$. Этот многочлен можно найти по формуле*

$$p(t) = \sum_{i=0}^{n} y_i \frac{\prod\limits_{j \neq i}(t - t_j)}{\prod\limits_{j \neq i}(t_i - t_j)}.$$

Доказательство. По теореме Безу условия $p(t_i) = y_i$ равносильны условиям $p \equiv y_i \mod (t - t_i)$. По китайской теореме об остатках существует единственный по модулю $w(t) = \prod\limits_{i=0}^{n}(t - t_j)$ многочлен, удовлетворяющий этим сравнениям. Единственный многочлен степени, не превосходящей n, – это остаток от деления любого такого многочлена на w. Формулу легко проверить непосредственно, или получить, применив конструктивное доказательство китайской теоремы об остатках. □

Другой, итерационный, способ решить задачу интерполяции называется интерполяцией по Ньютону. На k-ом шаге строится многочлен степени $\leqslant k$, удовлетворяющий первым $k + 1$ условиям. На нулевом шаге положим $p_0(t) = y_0$. Предположим, что построен многочлен p_k, удовлетворяющий условиям $\deg p_k \leqslant k$ и $p_k(t_i) = y_i$ для любого $i = 0, \ldots, k$. Будем искать p_{k+1} в виде $p_{k+1}(t) = p_k(t) + \lambda(t - t_0) \cdots (t - t_k)$. Первые $k + 1$ условий выполнены независимо от значения λ. Поэтому λ можно найти из условия $p_{k+1}(t_{k+1}) = y_{k+1}$. Очевидно, что тогда все требования к p_{k+1} будут выполнены.

17. Формальная производная и кратность корня

Так как поле F произвольно, то невозможно определить производную многочлена средствами дифференциального исчисления. Однако понятие формальной производной оказывается почти ничем не хуже. По некоторым соображениям нам будет удобно определить формальную производную для многочленов с коэффициентами в произвольном коммутативном кольце с единицей.

Определение 17.1. Пусть R – коммутативное кольцо с единицей. Формальной производной многочлена $p(t) = a_n t^n + \cdots + a_1 t + a_0 \in R[t]$ называется многочлен $p'(t) = a_n n t^{n-1} + \cdots + a_1$ (здесь натуральное число n понимается, как сумма n единиц кольца R, в частности, оно может оказаться равным нулю).

Лемма 17.2. *Формальная производная удовлетворяет всем обычным свойствам производной, т. е. для любых $p, q \in R[t]$ и $\alpha \in R$ имеют место равенства:*

(1) $(p + q)' = p' + q'$, $(\alpha p)' = \alpha p'$;
(2) $(pq)' = p'q + pq'$;
(3) $(p \circ q)' = (p' \circ q) \cdot q'$.

Приведем 2 доказательства этой леммы: непосредственное и "методом общего элемента". Непосредственное доказательство в данном случае может быть даже проще, но "метод общего элемента" при небольшой привычке к нему позволяет вообще не думать о доказательстве подобного рода утверждений.

Непосредственное доказательство. Линейность формальной производной очевидна. Учитывая это, второе свойство достаточно проверить для одночленов:

$$(t^n \cdot t^m)' = (m+n)t^{m+n-1} = nt^{n-1}t^m + mt^n t^{m-1} = (t^n)' \cdot t^m + t^n \cdot (t^m)'.$$

Аналогично, последнее свойство достаточно проверить для случая, когда $p = t^n$. В этом случае доказательство можно провести индукцией по n, используя свойство 2. База индукции, $n = 1$, очевидна. При $n > 1$, используя индукционное предположение, имеем

$$(q^n)' = (q \cdot q^{n-1})' = q' \cdot q^{n-1} + q \cdot (n-1)q^{n-2}q' = nq^{n-1}q'.$$

□

Для доказательства свойств производной методом общего элемента нам потребуются 2 леммы. Первую из них мы примем пока без доказательства (доказательство использует сведения про расширения полей, которые мы изучим позже).

Лемма 17.3. *Для любого натурального n кольцо многочленов от n переменных над $\mathbb{Z}$ вкладывается (т. е. изоморфно подкольцу) в $\mathbb{R}$.*

Следующее утверждение – это обобщение универсального свойства кольца многочленов на многочлены нескольких переменных.

Лемма 17.4. *Пусть R коммутативное кольцо с 1, $n \in \mathbb{N}$, а $c_1, \dots, c_n \in R$. Существует единственный гомоморфизм $\varphi : \mathbb{Z}[z_1, \dots, z_n] \to R$ такой, что $\varphi(z_k) = c_k$ при всех $k = 1, \dots, n$.*

Еще нам понадобится определение гомоморфизма колец многочленов, индуцированного гомоморфизмом колец коэффициентов. Пусть $\varphi : A \to B$ – гомоморфизм колец. Обозначим через $\hat{\varphi} : A[t] \to B[t]$ гомоморфизм, заданный формулой $\hat{\varphi}\left(\sum_{k=0}^{l} c_k t^k\right) = \sum_{k=0}^{l} \varphi(c_k)t^k$ и назовем его гомоморфизмом, индуцированным φ.

Доказательство леммы 17.2 методом общего элемента. Я не смогу сейчас объяснить, что такое "общий элемент для некоторой задачи", придется ограничиться определением общего элемента для нашей конкретной задачи – доказательства свойств формальной производной. Зафиксируем степени m и n многочленов p и q. Тогда свойства производной включают в себя элементы $a_0, \dots, a_n, b_0, \dots, b_m$ (коэффициенты многочленов p и q) и α кольца R. При этом между этими элементами нет никаких соотношений, кроме тех, которые следуют из свойств коммутативного кольца с единицей. Наиболее общая ситуация, когда такие элементы существуют – кольцо многочленов $P = \mathbb{Z}[x_0, \dots, x_n, y_0, \dots, y_m, \beta]$ от $n + m + 3$ переменных. Обозначим $f(t) = \sum_{k=0}^{n} x_k t^k \in P[t]$ и $g(t) = \sum_{k=0}^{m} y_k t^k \in P[t]$. Набор (f, g, β) и будет общим элементом для нашей задачи.

По лемме 17.4 для любого кольца R, элемента $\alpha \in R$ и многочленов $p, q \in R[t]$ степеней не превосходящих n и m, соответственно, существует единственный гомоморфизм $\varphi : P \to R$ такой, что $\varphi(x_k) = a_k$, $\varphi(y_k) = b_k$ и $\varphi(\beta) = \alpha$. Ясно, что эти равенства равносильны равенствам $\hat{\varphi}(f) = p$, $\hat{\varphi}(g) = q$ и $\varphi(\beta) = \alpha$. Эти свойства и означают, что P – универсальное кольцо, а (f, g, β) общий элемент для нашей задачи.

Легко видеть, что дифференцирование коммутирует с гомоморфизмами, т. е. для гомоморфизма колец $\psi : A \to B$ и многочлена $h \in A[t]$ имеет место равенство $\hat{\psi}(h)' = \hat{\psi}(h')$. Поэтому выполнение свойств формальной производной для $\beta \in P$ и многочленов $f, g \in P[t]$ влечет выполнение этих свойств для их образов при любом гомоморфизме.

Учитывая сказанное выше, выполнение свойств формальной производной для конкретного $\beta \in P$ и конкретных многочленов $f, g \in P[t]$ влечет выполнение этих свойств для любых многочленов

степеней не выше, чем n и m, а так как n и m произвольные, то и вообще для любых многочленов над любым коммутативным кольцом.

По лемме 17.3 кольцо P вкладывается в поле вещественных чисел, а над полем вещественных чисел свойства производной известны из математического анализа. Так как свойства сохраняются при изоморфизме, то можно считать, что $P \subseteq \mathbb{R}$, поэтому свойства производной выполнены для общего элемента, а, значит, и в общем случае. □

Единственная неприятность, с которой можно столкнуться, используя формальную производную над полем ненулевой характеристики, – это то, что она может оказаться равной нулю для многочлена ненулевой степени. Например, при $F = \mathbb{F}_p$ производная многочлена $t^p - 1$ тождественно равна нулю.

Определение 17.5. Число $\alpha \in F$ имеет кратность k в многочлене $f \in F[t]$, если k – наибольшее неотрицательное целое, для которого f делится на $(t-\alpha)^k$. Используя теорему Безу можно переформулировать это определение следующим образом: α имеет кратность k в f, если $f(t) = (t-\alpha)^k g(t)$, причем $g(\alpha) \neq 0$.

Ясно, что α имеет кратность больше 0 в f тогда и тогда, когда α – корень f. Корни первой кратности называются простыми корнями, а корни кратности не меньше 2 – кратными. С помощью формальной производной легко искать кратные корни многочлена, это опирается на следующее утверждение.

Лемма 17.6. *Пусть α – корень многочлена f кратности k. Кратность α в f' не меньше $k-1$. Если $k \neq 0$ в поле F, то α имеет кратность ровно $k-1$ в f', в частности, $k = 1 \iff f'(\alpha) \neq 0$.*

Доказательство. По условию $f(t) = (t-\alpha)^k g(t)$, причем $g(\alpha) \neq 0$. По свойствам дифференцирования

$$f'(t) = k(t-\alpha)^{k-1} g(t) + (t-\alpha)^k g'(t) = (t-\alpha)^{k-1}\big(kg(t) + (t-\alpha)g'(t)\big).$$

Сразу видно, что f' делится на $(t-\alpha)^{k-1}$. Если $k \neq 0$ в поле F, то $kg(\alpha) + (\alpha-\alpha)g'(\alpha) = kg(\alpha) \neq 0$. □

18. Основная теорема алгебры

При построении поля комплексных чисел мы присоединили корень многочлена t^2+1, но оказывается, что присоединились корни всех многочленов!

Определение 18.1. Поле F называется алгебраически замкнутым, если любой многочлен из $F[t]$ степени $\geqslant 1$ имеет хотя бы один корень в F.

Лемма 18.2. *Если поле F алгебраически замкнуто, то любой многочлен из $F[t]$ раскладывается на множители степени $\leqslant 1$.*

Теорема 18.3 (Основная теорема алгебры)**.** *Поле комплексных чисел алгебраически замкнуто.*

Существует по крайней мере 4 существенно различных доказательств этой теоремы.

- Прямое аналитическое доказательство, которое будет приведено ниже.
- Топологическое доказательство или “дама с собачкой”
- Доказательство из теории функций комплексного переменного с помощью теоремы Лиувилля.
- Доказательство в рамках теории Галуа, которое мы изучим в соответствующей главе.

Прямое аналитическое доказательство использует следующую лемму.

Лемма 18.4. *Пусть X – компактное топологическое пространство. Непрерывная функция $f : X \to \mathbb{R}$ достигает своего наименьшего значения.*

Доказательство основной теоремы алгебры. Пусть $p(z) = a_n z^n + \dots + a_0 \in \mathbb{C}[z]$.

$$|p(z)| \geqslant |a_n| \cdot |z|^n - |a_{n-1}| \cdot |z|^{n-1} - \dots - |a_0| \to \infty \text{ при } |z| \to \infty.$$

Следовательно, существует $r \in \mathbb{R}$ такой, что $|p(c)| > |p(0)|$ при $|c| > r$. Значит

$$\inf_{c \in \mathbb{C}} |p(c)| = \inf_{|c| \leqslant r} |p(c)|.$$

По предыдущей лемме существует $w \in \mathbb{C}$ такой, что $|p(w)| = \inf p$. Тогда многочлен $q(z) = p(z + w)$ принимает наименьшее значение в 0. Предположим, что $\inf |q(z)| = |q(0)| > 0$. Тогда $f(z) = q(z)/q(0)$ имеет вид

$$f(z) = 1 + b_m z^m + \dots + b_n z^n, \text{ где } 1 \leqslant m \leqslant n \text{ и } b_m \neq 0,$$

а наименьшее значение f равно $f(0) = 1$.

$$|f(z) - 1 - b_m z^m| = |b_{m+1} z^{m+1} + \dots + b_n z^n| \leqslant |b_{m+1}| \cdot |z|^{m+1} + \dots + |b_n| \cdot |z|^n = o(|z|^m)$$

при $|z| \to 0$. Выберем такое $\varepsilon > 0$, что $|f(z) - 1 - b_m z^m| < |z|^m \cdot |b_m|/2$ при $|z| < \varepsilon$. Подберем аргумент числа $z_0 \in \mathbb{C}$ так, чтобы $b_m z_0^m$ было бы отрицательным вещественным числом. Выберем модуль z_0 из двух условий: $|z_0| < \varepsilon$ и $|b_m z_0^m| = \alpha < 1$. Тогда

$$|f(z_0)| \leqslant |1 + b_m z_0^m| + |f(z_0) - 1 - b_m z_0^m| \leqslant 1 - \alpha + \alpha/2 < 1.$$

Противоречие показывает, что $q(0) = p(w) = 0$. □

19. Вещественные многочлены

Следствие 19.1. *Пусть $g \in \mathbb{C}[t]$, а $w \in \mathbb{C}$. Обозначим через $\overline{g}$ многочлен, коэффициенты которого сопряжены с коэффициентами многочлена g.*

1. $\overline{g(w)} = \overline{g}(\overline{w})$.
2. *Если $g \in \mathbb{R}[t]$, то $\overline{g(w)} = g(\overline{w})$.*
3. *Пусть $g \in \mathbb{R}[t]$, а $w \in \mathbb{C}$. Кратность w в g равна кратности $\overline{w}$ в g. В частности, если $g(w) = 0$, то $g(\overline{w}) = 0$.*

Доказательство. Первые 2 утверждения очевидно следуют из леммы 4.2.

Обозначим через k кратность w в g. Тогда $g(t) = (t - w)^k f(t)$, где $f \in \mathbb{C}[t]$. Взяв комплексно сопряженные к обеим частям равенства получим $g(t) = \overline{g}(t) = (t - \overline{w})^k \overline{f}(t)$, причем $\overline{f}(\overline{w}) = \overline{f(w)} \neq 0$, что и означает, что кратность $\overline{w}$ в g равна k. □

Следствие 19.2. *Любой многочлен степени $\geqslant 3$ из кольца $\mathbb{R}[x]$ приводим. Следовательно, любой многочлен над $\mathbb{R}$ раскладывается на множители степени $\leqslant 2$.*

Доказательство. Пусть $p \in \mathbb{R}[t] \subseteq \mathbb{C}[t]$, $\deg p \geqslant 3$. По основной теореме алгебры он имеет комплексный корень w. Если $w \in \mathbb{R}$, то по теореме Безу p делится на $t - w$. В противном случае по лемме 4.2 $p(\overline{w}) = 0$. Так как в этом случае $t - w$ и $t - \overline{w}$ взаимно просты, а по теореме Безу p делится на каждый из этих многочленов, то он делится и на их произведение $(t - w)(t - \overline{w}) = t^2 - (w + \overline{w})t + w\overline{w}$, которое имеет вещественные коэффициенты. В обоих случаях p делится на многочлен степени 1 или 2, и степень частного не меньше 1.

Второе утверждение следует из факториальности кольца многочленов. □

20. Лемма Гаусса

Везде в этом параграфе R обозначает факториальную область целостности, а F – ее поле частных.

Лемма Гаусса – это набор утверждений про кольцо многочленов $R[t]$, каждое из которых в разных текстах называется “лемма Гаусса”. Из леммы Гаусса следует, в частности, факториальность $R[t]$.

Пусть $p \in R$ – неприводимый элемент, а $f \in R[t]$. Обозначим через $v_p(f)$ наибольшее целое d такое, что f делится на p^d в кольце $R[t]$. Наибольший общий делитель коэффициентов многочлена f называется *содержанием* этого многочлена и обозначается через $\operatorname{cont} f$. Многочлен называется *примитивным*, если его содержание равно 1.

Лемма 20.1. *Любой многочлен $f \in F[t]$ представляется в виде αg, где $\alpha \in F$, а g – примитивный многочлен из $R[t]$.*

Пусть A – кольцо, I – идеал в A, а $I[t]$ – идеал кольца $A[t]$, состоящий из многочленов с коэффициентами из I.

Лемма 20.2. *$A[t]/I[t] \cong (A/I)[t]$. Если I – простой идеал, то $A[t]/I[t]$ – область целостности.*

Доказательство. Определим очевидный эпиморфизм $A[t] \to (A/I)[t]$ и посчитаем его ядро.
Кольцо многочленов над областью целостности не имеет делителей нуля, так как коэффициент при старшей степени произведения равен произведению старших коэффициентов. □

Лемма 20.3. *Произведение примитивных многочленов примитивно.*

Доказательство. Предположим, что $fg = ph$ для некоторых $f, g, h \in R[t]$ и $p \in R$, причем f и g примитивны, а p неприводим. Тогда образ fg в факторкольце $R[t]/pR[t]$ равен нулю. Так как R факториально, а p неприводим, то pR – простой идеал, и по предыдущей лемме $R[t]/pR[t]$ – область целостности. Следовательно, образ одного из многочленов f и g равен нулю. Таким образом, f или g лежит в $pR[t]$, т. е. делится на p. Но тогда и все коэффициенты этого многочлена делятся на p, что противоречит его примитивности. □

Как утверждение, так и доказательство, легко обобщается на многочлены от нескольких переменных.

Следствие 20.4. *Пусть $f, g \in R[t]$, а p – неприводимый элемент кольца R. Тогда $v_p(fg) = v_p(f) + v_p(g)$, и $\operatorname{cont} fg = \operatorname{cont} f \operatorname{cont} g$. В частности, если fg делится на p, то или f, или g делится на p.*

Теорема 20.5 (Гаусса)**.** *Предположим, что многочлен $f \in R[t]$ примитивен. Тогда f неприводим в $R[t]$ тогда и только тогда, когда он неприводим в $F[t]$.*

Доказательство. Предположим, что f приводим в $R[t]$, т. е. $f = uv$ для некоторых необратимых элементов $u, v \in R[t]$. Так как f примитивен, то $u, v \notin R$. Но тогда $f = uv$ – разложение на необратимые в $F[t]$, т. е. f приводим в $F[t]$.
Обратно, пусть f приводим в $F[t]$, т. е. $f = uv$ для некоторых многочленов $u, v \in F[t]$ ненулевой степени. Приводя коэффициенты многочленов u, v к общему знаменателю получим $uv = \tilde{u}\tilde{v}/r$ для некоторых $\tilde{u}, \tilde{v} \in R[t]$ и $r \in R$. Домножая на r получаем $rf = \tilde{u}\tilde{v}$. Тогда $\operatorname{cont}(rf) = r \operatorname{cont} f = r = \operatorname{cont}(\tilde{u}\tilde{v}) = \operatorname{cont} \tilde{u} \operatorname{cont} \tilde{v}$. Таким образом, $f = \frac{\tilde{u}}{\operatorname{cont} \tilde{u}} \cdot \frac{\tilde{v}}{\operatorname{cont} \tilde{v}}$ – разложение f на необратимые множители в $R[t]$. □

Теорема 20.6 (о рациональных корнях многочлена)**.** *Пусть $f(t) = a_n t^n + \dots + a_0 \in R[t]$. Тогда корнями f в F могут быть только элементы вида $\frac{c}{d}$, где c – делитель свободного члена a_0, а d – делитель старшего коэффициента a_n.*

Доказательство. Пусть $f = h \cdot (t - c/d)$, где $h \in F[t]$, а $\gcd(c, d) = 1$. По лемме 20.1 $h = \frac{r}{s} g$ для некоторых $r, s \in R$ и примитивного многочлена $g \in R[t]$. Тогда $dsf = rg \cdot (dt - c)$. По лемме 20.4 $ds \operatorname{cont} f = r$, откуда $b = \frac{r}{ds} \in R$. Приравнивая старшие коэффициенты в равенстве $f = bg \cdot (dt - c)$ получаем $a_n = bg_{n-1}d$, Аналогично, приравнивая свободные члены, имеем $a_0 = -bg_0c$ (здесь g_0 и g_{n-1} – соответствующие коэффициенты многочлена g). □

Пусть $A \subseteq B$ – коммутативные кольца. Элемент $b \in B$ называется целым над A, если он является корнем унитального многочлена с коэффициентами из A. Множество $\operatorname{Int}_B A$, состоящее из элементов B, целых над A, называется целым замыканием A в B. Кольцо A называется целозамкнутым в B, если $\operatorname{Int}_B A = A$. Область целостности A называется целозамкнутой, если она целозамкнута в своем поле частных. Следующее утверждение является частным случаем теоремы о рациональных корнях многочлена.

Следствие 20.7. *Факториальная область целостности целозамкнута.*

Теорема 20.8. *Кольцо многочленов над факториальным кольцом факториально.*

Доказательство. Докажем вначале существование разложения на неприводимые. Пусть $f \in R[x] \setminus \{0\}$ – многочлен наименьшей степени, не раскладывающийся в произведение неприводимых. Ясно, что его можно представить в виде $f = \tilde{f} \cdot \operatorname{cont} f$, где $\tilde{f}$ примитивен. Так как R факториально, то $\operatorname{cont} f$ равен произведению неприводимых. Так как f не является произведением неприводимых, то $\tilde{f}$ приводим, скажем $\tilde{f} = gh$ для некоторых необратимых $g, h \in R[t]$. Так как $1 = \operatorname{cont} \tilde{f} = \operatorname{cont} g \operatorname{cont} h$, то степени g и h больше нуля, следовательно, $\deg g$, $\deg h < \deg \tilde{f} = \deg f$. По минимальности $\deg f$ многочлены g и h раскладываются в произведение неприводимых, но тогда и f обладает этим свойством.

Для доказательства единственности по лемме 12.2 достаточно доказать, что каждый неприводимый элемент в $R[t]$ прост. Пусть u – неприводимый элемент кольца $R[t]$. Если $\deg u = 0$, т. е. $u \in R$, то по лемме 20.2 $R[t]/uR[t]$ – область целостности. Иначе, u примитивен. По теореме Гаусса u неприводим в $F[t]$. Кольцо $F[t]$ евклидово, следовательно факториальное, поэтому $F[t]/uF[t]$ – область целостности. Ядро композиции $R[t] \to F[t] \to F[t]/uF[t]$ равно $R[t] \cap uF[t]$, поэтому $R[t]/(R[t] \cap uF[t])$ вкладывается в $F[t]/uF[t]$ и, следовательно, является областью целостности. Пусть $v = uw$ для $v \in R[t]$ и $w \in F[t]$. По лемме 20.1 $w = \frac{r}{s}\tilde{w}$ для некоторых $r, s \in R$ и примитивного $\tilde{w} \in R[t]$. Тогда $sv = ru\tilde{w}$, откуда $s \operatorname{cont} v = r \operatorname{cont} u \operatorname{cont} \tilde{w} = r$. Следовательно, $r/s \in R$, откуда $v \in uR[t]$. Таким образом, $R[t] \cap uF[t] = uR[t]$, следовательно, $R[t]/uR[t]$ – область целостности, что и требовалось. □

21. Экспонента мультипликативной группы кольца вычетов

Разберем подробнее кольцо целых чисел. Так как $\mathbb{Z}$ является евклидовым кольцом, то оно является областью главных идеалов. По лемме 10.5 любой ненулевой простой идеал является максимальным, откуда $\mathbb{Z}/p\mathbb{Z}$ является полем тогда и только тогда, когда p – простое число.

Если числа $n_1, \ldots, n_\ell$ – попарно взаимно простые, то имеет место китайская теорема об остатках:

$$\mathbb{Z}/(n_1 \cdots n_\ell \mathbb{Z}) \cong \mathbb{Z}/n_1\mathbb{Z} \oplus \cdots \oplus \mathbb{Z}/n_\ell\mathbb{Z}.$$

Определение 21.1. Порядок мультипликативной группы $(\mathbb{Z}/n\mathbb{Z})^*$ обозначается $\varphi(n)$. Функция $\varphi : \mathbb{N} \to \mathbb{N}$ называется функцией Эйлера.

Лемма 21.2. *Образ числа $m \in \mathbb{Z}$ обратим в кольце $\mathbb{Z}/n\mathbb{Z}$, если и только если $\gcd(m, n) = 1$. Таким образом, $\varphi(n)$ равна количеству чисел от 0 до $n-1$, взаимно простых с n.*

Доказательство. Пусть $\bar{m}$ – образ m в $\mathbb{Z}/n\mathbb{Z}$. Элемент $\bar{m}$ обратим тогда и только тогда, когда существует $\bar{x} \in \mathbb{Z}/n\mathbb{Z}$ такой, что $\bar{m}\bar{x} = 1$. Если $x \in \mathbb{Z}$ прообраз $\bar{x}$, то последнее условие можно переписать в виде $mx \in 1 + n\mathbb{Z}$, другими словами, идеалы $m\mathbb{Z}$ и $n\mathbb{Z}$ взаимно просты. А это по следствию 13.4 означает, что m взаимно просто с n.

Второе утверждение очевидно следует из первого. □

Лемма 21.3. *Если кольцо R с единицей (не обязательно коммутативное) является прямой суммой колец $R_1 \oplus \cdots \oplus R_k$, то $R^* \cong R_1^* \times \cdots \times R_k^*$. Если R^* конечна, то $|R^*| = |R_1^*| \cdots |R_k^*|$.*

Теорема 21.4. *Если* $\gcd(a,b)=1$, *то* $\varphi(ab)=\varphi(a)\varphi(b)$.
Если p – *простое число, а* $k\in\mathbb{N}$, *то* $\varphi(p^k)=p^k-p^{k-1}$.

Пусть $p_1,\dots,p_\ell$ – *различные простые числа,* $k_1,\dots,k_\ell\in\mathbb{N}$, *а* $n=\prod\limits_{i=1}^{\ell}p_i^{k_i}$. *Тогда*

$$\varphi(n)=\prod_{i=1}^{\ell}(p_i^{k_i}-p_i^{k_i-1})=n\prod_{i=1}^{\ell}\frac{p_i-1}{p_i}.$$

Теорема 21.5 (теорема Эйлера). *Если* a *взаимно просто с* n, *то* $a^{\varphi(n)}\equiv 1 \mod n$.

Частный случай этой теоремы для простого n называется малой теоремой Ферма. На самом деле несложно получить более точный результат. Определим функцию Кармайкла $\lambda:\mathbb{N}\to\mathbb{N}$ равенством $\lambda(n)=\exp(\mathbb{Z}/n\mathbb{Z})^*$.

Теорема 21.6. *Пусть* $p_1,\dots,p_\ell$ – *различные простые числа, а* $k_1,\dots,k_\ell\in\mathbb{N}$. *Тогда*

$$\lambda\Big(\prod_{i=1}^{\ell}p_i^{k_i}\Big)=\operatorname*{lcm}_{1\leqslant i\leqslant\ell}\lambda(p_i^{k_i}).$$

Доказательство. Пусть $n=\prod\limits_{i=1}^{\ell}p_i^{k_i}$. По китайской теореме об остатках $\mathbb{Z}/n\mathbb{Z}=\bigoplus\limits_{i=1}^{\ell}\mathbb{Z}/p_i^{k_i}\mathbb{Z}$. По лемме 21.3 $(\mathbb{Z}/n\mathbb{Z})^*=\mathop{\times}\limits_{i=1}^{\ell}(\mathbb{Z}/p_i^{k_i}\mathbb{Z})^*$, а по пункту 3 леммы 10.3 экспонента группы $(\mathbb{Z}/n\mathbb{Z})^*$ равна наименьшему общему кратному экспонент групп $(\mathbb{Z}/p_i^{k_i}\mathbb{Z})^*$. □

Экспонента группы всегда делит ее порядок, так что из теоремы сразу следует, что если $\gcd(a,n)=1$, то $a^{\varphi'(n)}\equiv 1 \mod n$, где $\varphi'(n)$ – наименьшее общее кратное чисел $\varphi(p_i^{k_i})=p_i^{k_i}-p_i^{k_i-1}$. Следующая теорема утверждает, что функция Кармайкла часто равна этому числу.

Теорема 21.7 (Кармайкла). *Группа* $(\mathbb{Z}/p^k\mathbb{Z})^*$ *циклическая для любого простого* $p\neq 2$ *и при* $p=2$, $k\leqslant 2$. *При* $k\geqslant 3$ *экспонента группы* $(\mathbb{Z}/2^k\mathbb{Z})^*$ *равна* 2^{k-2} *(т. е. в 2 раза меньше ее порядка).*

Доказательство. Пусть p – нечетное простое число. Поле $\mathbb{Z}/p\mathbb{Z}$ является факторкольцом кольца $\mathbb{Z}/p^k\mathbb{Z}$ по идеалу, порожденному p. Каноническая проекция $\mathbb{Z}/p^k\mathbb{Z}\to\mathbb{Z}/p\mathbb{Z}$ индуцирует гомоморфизм мультипликативных групп $(\mathbb{Z}/p^k\mathbb{Z})^*\to(\mathbb{Z}/p\mathbb{Z})^*$, который сюръективен, так как все числа от 1 до $p-1$ взаимно просты с p^k и, следовательно, из классы вычетов обратимы. По теореме 16.5 существует элемент $a\in\mathbb{Z}/p\mathbb{Z}$ порядка $p-1$. Порядок любого его прообраза в $(\mathbb{Z}/p^k\mathbb{Z})^*$ делится на $p-1$. Таким образом, экспонента группы $(\mathbb{Z}/p^k\mathbb{Z})^*$ делится на $p-1$. Докажем теперь, что при $k\geqslant 2$ она делится также и на p^{k-1}.

Для этого докажем, что элемент $1+p$ имеет порядок p^{k-1} в этой группе. Точнее, индукцией по k докажем, что $(1+p)^{p^{k-1}}=1+p^ky$, где y не делится на p. При $k=1$ это очевидно. Пусть $k\geqslant 2$. По индукционному предположению $(1+p)^{p^{k-2}}=1+p^{k-1}x$, где x не делится на p.

$$(1+p)^{p^{k-1}}=(1+p^{k-1}x)^p=1+p\cdot p^{k-1}x+\sum_{i=2}^{p}C_p^ip^{i(k-1)}x^i.$$

Так как C_p^i делится на p, то каждое слагаемое суммы делится на $p^{1+2(k-1)}=p^{k+1}p^{k-2}$. Так как $k\geqslant 2$, то сумма равна $p^{k+1}z$, а $(1+p)^{p^{k-1}}=1+p^ky$, где $y=x+pz$ не делится на p.

Таким образом, порядок элемента $1+p$ в группе $(\mathbb{Z}/p^k\mathbb{Z})^*$ делит p^{k-1}. Заменяя k на $k-1$ и учитывая, что y не делит p, видим, что этот порядок не делит p^{k-2} и, следовательно, равен p^{k-1}.

При $p=2$ и $k\geqslant 2$ аналогичным образом докажем, что $(1+4z)^{2^{k-2}}=1+2^ky$, где y имеет ту же четность, что и z (нам нужно только $k\geqslant 3$, но это верно и при $k=2$, которое удобно сделать

базой индукции). Итак, при $k \geqslant 3$ по индукционному предположению $(1+4z)^{2^{k-3}} = 1+2^{k-1}x$, где $x \equiv z \mod 2$.

$$(1+4z)^{2^{k-2}} = (1+2^{k-1}x)^2 = 1+2\cdot 2^{k-1}x + 2^{2k-2}x^2 = 1+2^k(x+2^{k-2}x^2).$$

Так как $k \geqslant 3$, то $y = x + 2^{k-2}x^2 \equiv x \equiv z \mod 2$. Так же, как и в первой части доказательства заключаем, что при нечетном z порядок $1+4z$ в группе $(\mathbb{Z}/2^k\mathbb{Z})^*$ равен $k-2$.

С другой стороны, при любом t имеем $(1+2t)^2 = 1+4z$, где $z = t+t^2$ четно. Поэтому

$$(1+2t)^{2^{k-2}} = (1+4z)^{2^{k-3}} = 1+2^{k-1}y \equiv 1 \mod 2^k,$$

так как y четно. Следовательно, порядок любого элемента группы $(\mathbb{Z}/2^k\mathbb{Z})^*$ делит 2^{k-2}. □

Таким образом, две предыдущие теоремы дают точную версию теоремы Эйлера.

22. О простых числах

В этом параграфе приводятся несколько утверждений, полезных, в частности, для приложений в RSA-шифровании. Читателю рекомендуется прочитать идею RSA хотя бы в Википедии `https://ru.wikipedia.org/wiki/RSA`.

Теорема 22.1. *Пусть R – нетерова область целостности с конечной мультипликативной группой или кольцо многочленов над областью целостности. Тогда в нем существует бесконечно много неприводимых элементов.*

Доказательство. Пусть $p_1, \ldots, p_m$ – все неприводимые элементы кольца R. Тогда произведение $p_1^k \cdots p_m + 1$ не делится ни на один неприводимый для любого $k \in \mathbb{N}$. Все такие элементы различны, так как R – область целостности. Если мультипликативная группа кольца R конечна, то все такие элементы не могут быть обратимы, однако по теореме 11.3 необратимый элемент обязан делится на неприводимый, противоречие.

Конечная область целостности K является полем (если $|K| = n$, а $x \in K \smallsetminus \{0\}$, то $x^0, \ldots, x^n$ не могут быть все различны; $x^k = x^m$ при $k \neq m$ влечет $x^{k-m} = 1$, откуда x обратим). Кольцо многочленов над конечным полем – нетерова область целостности с конечной мультипликативной группой. Если же базовое кольцо бесконечно, то многочлены $t-a$ неприводимы для всех a из базового кольца, а таких уже бесконечно много. □

Теорема 22.2 (Теорема Дирихле о простых в арифметической прогрессии)**.** *Пусть $R = \mathbb{Z}$ или $R = \mathbb{F}_q[t]$. Если a и b взаимно простые элементы кольца R, то множество $a+bR$ содержит бесконечно много простых элементов (заметим, что в таком кольце R неприводимость элемента совпадает с его простотой).*

Теорема 22.3 (распределение простых чисел)**.** *Обозначим через $\pi(n)$ количество простых чисел от 2 до n. Тогда $\frac{\pi(n)}{n/\ln n} \to 1$ при $n \to \infty$.*

Практически важной задачей является нахождение больших простых чисел. В соответствии с предыдущей теоремой простых числе достаточно много. Поэтому, если алгоритм тестирования числа n на простоту требует $g(n)$ операций, то нахождение ближайшего к n простого числа статистически требует $g(n)\ln n$ операций.

Таким образом, важно иметь тесты числа на простоту порядка $\ln^k n$ для небольших k. Все (известные мне) такие тесты являются вероятностными и строятся следующим образом. Для (нечетного) $n > 1$ определяется некоторое подмножество $T = T(n) \subseteq \mathbb{Z}/n\mathbb{Z} \smallsetminus \{0\}$, которое совпадает со всем множеством в случае простого n. Если $a \in T(n)$, то говорят, что n проходит тест по основанию a. Отношение $|T(n)|/(n-1)$ для данного не простого n показывает, насколько хорош наш тест для данного n. На самом деле, можно оценить $\sup |T(n)|/(n-1)$ по всем непростым n. Если этот супремум близок к 1, то тест плохой и применять его в коммерческих целях нецелесообразно. Те не простые числа n, для которых $T = (\mathbb{Z}/n\mathbb{Z})^*$, называются псевдопростыми для данного теста.

В следующей таблице собраны основные вероятностные тесты. Если для теста существуют псевдопростые числа, то $\sup |T|/n$ по всем непростым n равен (или близок к) 1. В последнем столбике таблицы приведена оценка для этого супремума по всем n, кроме простых и псевдопростых.

Название	Псевдопростые числа	$\sup \frac{\lvert T\rvert}{\varphi(n)}$
Ферма	561, 1105, 1729,...	$\leqslant 1/2$
Эйлера	1729,2465,...	$\leqslant 1/2$
Соловея–Штрассена	$\varnothing$	$\leqslant 1/2$
Миллера–Рабина	$\varnothing$	$\leqslant 1/4$

23. Описание тестов на простоту

Для теста Ферма в качестве множества T берется

$$F(n) := \{x \in \mathbb{Z}/n\mathbb{Z} \smallsetminus \{0\} \mid x^{n-1} \equiv 1 \mod n\}.$$

Число n является псевдопростым числом Ферма, если порядок любого элемента группы $(\mathbb{Z}/n\mathbb{Z})^*$ делит $n-1$, что равносильно условию $\exp(\mathbb{Z}/n\mathbb{Z})^* | (n-1)$. Для $561 = 3 \cdot 11 \cdot 17$ имеем $\lambda(561) = \operatorname{lcm}(2, 10, 16) = 80$ делит $560 = 561 - 1$.

Множество $F(n) = F(n) \cap (\mathbb{Z}/n\mathbb{Z})^*$ – является ядром гомоморфизма $(\mathbb{Z}/n\mathbb{Z})^* \to (\mathbb{Z}/n\mathbb{Z})^*$, $x \mapsto x^{n-1}$. Если n не является простым или псевдопростым, то это собственная подгруппа. Ее индекс не меньше 2, таким образом, $\frac{|F(n)|}{n} \leqslant \frac{|F(n)|}{\varphi(n)} \leqslant \frac{1}{2}$.

Лемма 23.1. *Псевдопростое число Ферма свободно от квадратов и не равно произведению двух простых.*

Доказательство. Достаточно доказать, что если n делится на квадрат или равно произведению двух простых, то $n-1$ не делится на экспоненту группы $G = (\mathbb{Z}/n\mathbb{Z})^*$. Если n делится на квадрат, то $n = p^k m$ для некоторого простого p, натурального $k > 1$ и натурального m, не делящегося на p. Тогда экспонента группы $G = (\mathbb{Z}/n\mathbb{Z})^*$ равна $\operatorname{lcm}(\lambda(p^k), \lambda(m))$ и, следовательно, делится на p, в то время как $n-1$ не делится на p.

Если $n = pq$, где p и q – различные простые числа, то экспонента группы G равна $\operatorname{lcm}(p-1, q-1)$. Если $pq-1$ делится на эту экспоненту, то оно делится на $p-1$ и на $q-1$. Тогда $q-1 = (pq-1)-q(p-1)$ делится на $p-1$ и, аналогично, $p-1$ делится на $q-1$, что невозможно. □

Тест Эйлера тестирует только нечетные числа (какой смысл тестировать на простоту четные числа?). В качестве T берется $E(n) := \{a \mid a^{\frac{n-1}{2}} = \pm 1\}$. Ясно, что $E(n) \subseteq F(n)$, так что $|E(n)| \leqslant n/2$, если n не является простым или псевдопростым Ферма. Если n – простое, то в поле $\mathbb{Z}/n\mathbb{Z}$ равенство $a^{n-1} = 1$ равносильно равенству $a^{\frac{n-1}{2}} = \pm 1$. Таким образом, все простые числа проходят тест Эйлера по любому основанию.

При $n = 1729 = 7 \cdot 13 \cdot 19$ имеем $\lambda(1729) = \operatorname{lcm}(6, 12, 18) = 36$, а $(1729-1)/2 = 864 = 36 \cdot 24$. Таким образом, $a^{\frac{n-1}{2}} = 1$ для любого $a \in (\mathbb{Z}/n\mathbb{Z})^*$.

Предложение 23.2. *Нечетное n является псевдопростым числом Эйлера тогда и только тогда, когда $\frac{n-1}{2}$ делится на $\lambda(n)$.*

Для псевдопростых Ферма $n = 561$ и $n = 1105$ имеем $|E(n)| = \varphi(n)/2$.

Доказательство. Если для не простого n число $\frac{n-1}{2}$ не делится на $\lambda(n)$, то $\frac{n-1}{2}$ не делится на $\lambda(p^k)$ для некоторого простого p такого, что $n = p^k m$, где m не делится на p. $(\mathbb{Z}/n\mathbb{Z})^* \cong (\mathbb{Z}/p^k\mathbb{Z})^* \times (\mathbb{Z}/m\mathbb{Z})^*$, причем существует $a \in (\mathbb{Z}/p^k\mathbb{Z})^*$ такое, что $a^{(n-1)/2} \neq 1$. Тогда $(a, 1)^{(n-1)/2} = (a^{(n-1)/2}, 1) \neq \pm(1, 1)$.

Обратно, если $\frac{n-1}{2}$ делится на $\lambda(n)$, то $a^{\frac{n-1}{2}} = 1$ в группе $(\mathbb{Z}/n\mathbb{Z})^*$. При этом n не простое, так как для простых чисел $\lambda(n) = n-1$ не делит $\frac{n-1}{2}$.

При $n = 561 = 3 \cdot 11 \cdot 17$ заметим, что $\frac{n-1}{2} = 280$ делится на $3-1$ и на $11-1$, но не делится на $17-1$. Поэтому a^{280} сравнимо с 1 по модулям 3 и 11, и с ± 1 по модулю 17 (здесь a взаимно просто с n). Так как возведение в степень – гомоморфизм, то прообразы элементов равны по мощности. Таким образом, ровно половина элементов $a \in (\mathbb{Z}/n\mathbb{Z})^*$ перейдет в 1. Доказательство для $n = 1105$ аналогично. □

Тест Миллера–Рабина, также как и тест Эйлера, проводится только для нечетных чисел. Пусть $n - 1 = 2^k m$, где m нечетно. Тогда

$$MR(n) := \{a \in \mathbb{Z}/n\mathbb{Z} \mid a^m = \pm 1 \text{ или } \exists j < k : a^{2^j m} = -1\}.$$

Теорема 23.3. *Число n проходит тест Миллера–Рабина по любому основанию a, взаимно простому с n, тогда и только тогда, когда n простое*

Доказательство. Если n простое, то $\mathbb{Z}/n\mathbb{Z}$ – поле. Заметим, что в любом поле уравнение $x^2 = 1$ имеет ровно два решения: $x = \pm 1$. Так как $a^{n-1} = 1$, то $a^{\frac{n-1}{2}} = \pm 1$. Если $\frac{n-1}{2} = m$ нечетно или $a^{\frac{n-1}{2}} = -1$, то тест выполнен. Иначе $a^{\frac{n-1}{4}} = \pm 1$, и т. д.

Обратно, пусть n не простое, но $MR(n) = (\mathbb{Z}/n\mathbb{Z})^*$. Так как $MR(n) \subseteq F(n)$, то по лемме 23.1 n свободно от квадратов, т. е. $n = p_1 \cdot \ldots \cdot p_\ell$ для различных простых $p_1, \ldots, p_\ell$. Пусть j наибольшее целое, при котором $2^j m$ не делится на $\lambda(n) = \operatorname{lcm}(p_1 - 1, \ldots, p_\ell - 1)$ (такое j существует, так как нечетное m не делится на четное $\lambda(n)$). Предположим для определенности, что $2^j m$ не делится на $p_1 - 1$. Тогда существует $a \in (\mathbb{Z}/p_1\mathbb{Z})^*$ такое, что $a^{2^j m} \neq 1$. Тогда $(a, 1, \ldots, 1)^{2^j m} \neq \pm 1$, откуда $(a, 1, \ldots, 1)^{2^i m} \neq \pm 1$ при любом $i \leqslant j$. С другой стороны, максимальность j означает, что при $i > j$ $b^{2^i m} = 1$ для любого $b \in (\mathbb{Z}/n\mathbb{Z})^*$. Таким образом, по основанию элемента из $(\mathbb{Z}/n\mathbb{Z})^*$, соответствующего элементу $(a, 1, \ldots, 1) \in (\mathbb{Z}/p_1\mathbb{Z})^* \times \cdots \times (\mathbb{Z}/p_\ell\mathbb{Z})^*$, число n не проходит тест Миллера–Рабина. □

Упражнение 23.4. Найдите порядок множеств $F(n)$ и $E(n)$ для любого нечетного n.

Упражнение 23.5. Докажите, что $|MR(n)| \leqslant \frac{\varphi(n)}{4}$, если n не простое.

Упражнение 23.6. Приведите пример числа n, для которого $|MR(n)| = \frac{\varphi(n)}{4}$.

Определение 23.7. Пусть p – нечетное простое число, а $a \in \mathbb{N}$ не делится на p. Символом Лежандра $\left(\frac{a}{p}\right)$ называется число 1, если $a \in (\mathbb{F}_p^*)^2$ и -1 в противном случае. Другими словами, $\left(\frac{a}{p}\right) = \pm 1$ и $\left(\frac{a}{p}\right) \equiv a^{\frac{p-1}{2}} \mod p$.

Пусть $n = p_1 \cdots p_m$, где p_k – нечетные простые числа (не обязательно различные), а $a \in \mathbb{N}$ взаимно просто с N. Символом Якоби называется число $\left(\frac{a}{n}\right) = \left(\frac{a}{p_1}\right) \cdots \left(\frac{a}{p_n}\right)$.

Следующие 2 утверждения дают возможность эффективно вычислять символ Якоби.

Предложение 23.8 (свойства символа Якоби). *Пусть n – нечетное число. Тогда*

$$\left(\frac{1}{n}\right) = 1; \qquad \left(\frac{-1}{n}\right) = (-1)^{\frac{n-1}{2}}; \qquad \left(\frac{2}{n}\right) = (-1)^{\frac{n^2-1}{8}}.$$

Теорема 23.9 (квадратичный закон взаимности). *Если a и n – взаимно простые нечетные числа, то $\left(\frac{a}{n}\right)\left(\frac{n}{a}\right) = (-1)^{\frac{a-1}{2} \cdot \frac{n-1}{2}}$.*

Пусть n – нечетное число. Положим

$$SS(n) = \{a \in (\mathbb{Z}/n\mathbb{Z})^* \mid \left(\frac{a}{n}\right) \equiv a^{\frac{n-1}{2}} \mod n\}.$$

Из определения символа Лежандра следует, что $SS(n) = (\mathbb{Z}/n\mathbb{Z})^*$ при условии, что n простое.

Теорема 23.10. *Если n не простое, то существует не более $\varphi(n)/2$ элементов $a \in (\mathbb{Z}/n\mathbb{Z})^*$, для которых $\left(\frac{a}{n}\right) = a^{\frac{n-1}{2}}$ (равенство в группе $(\mathbb{Z}/n\mathbb{Z})^*$).*

Доказательство последней теоремы оставляется читателю в качестве упражнения.

Глава 5

Определители

1. Полилинейные и антисимметричные формы.

Сейчас мы определим "форму объема" $f : \underbrace{V \times \cdots \times V}_{n \text{ раз}} \to \mathbb{R}$, которая n векторам n-мерного евклидова пространства V сопоставляет объем параллелепипеда, построенного по этим векторам. Ясно, что если два вектора совпадают, то параллелепипед вырожден и объем равен 0. Нетрудно показать, что при умножении любого из векторов на (положительное) число λ объем умножается на λ, т. е.

$$f(v_1, \dots, \lambda v_i, \dots, v_n) = \lambda f(v_1, \dots, v_i, \dots, v_n).$$

При сложении векторов имеет место формула

$$f(v_1, \dots, v_i + v_i', \dots, v_n) = \pm f(v_1, \dots, v_i, \dots, v_n) \pm f(v_1, \dots, v_i, \dots, v_n).$$

Если убрать $\pm$ и разрешить умножать на отрицательные числа, то последние две формулы будут очень похожи на линейность по каждому аргументу. Поэтому разумно рассматривать форму ориентированного объема, который может быть отрицательным. Тогда можно считать, что f линейна по каждому аргументу и удовлетворяет равенству

$$f(\dots, x, \dots, x, \dots) = 0.$$

Заметьте, что мы зафиксировали порядок аргументов функции f, от которого теперь будет зависеть знак. Действительно,

$$\begin{aligned} 0 = f(\dots, x+y, \dots, x+y, \dots) =& f(\dots, x, \dots, x, \dots) + f(\dots, y, \dots, y, \dots) + \\ & f(\dots, x, \dots, y, \dots) + f(\dots, y, \dots, x, \dots), \end{aligned}$$

откуда $f(\dots, x, \dots, y, \dots) = -f(\dots, y, \dots, x, \dots)$.

Мы привели мотивировку для изучения антисимметричных функций, линейных по каждому аргументу. Далее в этом параграфе мы изучаем такие функции над любым коммутативным кольцом вместо поля $\mathbb{R}$. Пусть $V \cong R^n$ – свободный модуль над коммутативным кольцом R.

Определение 1.1. Отображение $f : \underbrace{V \times \cdots \times V}_{m \text{ раз}} \to R$ называется *полилинейной* (точнее, *m-линейной*) *формой*, если оно линейно по каждому аргументу, т.е. для любых $a, b \in V$ и $\lambda \in R$ выполнены следующие равенства

$$\begin{aligned} f(\dots, a+b, \dots) &= f(\dots, a, \dots) + f(\dots, b, \dots), \\ f(\dots, \lambda a, \dots) &= \lambda f(\dots, a, \dots) \end{aligned}$$

Заметим, что если среди аргументов полилинейного отображения f есть 0, то f принимает значение 0. Это сразу следует из линейности по каждому аргументу. Пусть $v = (v_1, \dots, v_n)$ – базис модуля V. Тогда полилинейная форма полностью определяется m-мерным массивом своих значений на базисных векторах. Точнее, выполнено следующее утверждение.

Лемма 1.2. *Пусть f – m-линейная форма на V, а $x = (x_1, \dots, x_m) \in {}^mV$. Положим $A = \big((x_1)_v, \dots, (x_m)_v\big)$, так что $x = vA$. Тогда*

$$f(x_1, \dots, x_m) = \sum_{i_1, \dots, i_m = 1}^{n} f(v_{i_1}, \dots, v_{i_m}) a_{i_1 1} \dots a_{i_m m}.$$

Доказательство. Для доказательства достаточно разложить каждый x_k в линейную комбинацию базисных векторов: $x_k = \sum_{i_k=1}^{n} v_{i_k} a_{i_k k}$ после чего вынести знаки суммирования и константы за знак отображения f, что можно сделать по определению полилинейности. □

Прежде, чем формулировать следствие последней леммы для антисимметричных форм, изучим два различных определения антисимметричности.

Определение 1.3. Пусть X – множество. Функция $f : X \times X \to R$ называется антисимметричной, если для любых $x, y \in X$ выполнены следующие условия:

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

Условия (1) и (2) редко бывают независимыми. Что из чего следует и при каких условиях, изложено в следующей лемме.

Лемма 1.4. *Если 2 не является делителем нуля в R, то* (1) $\Longrightarrow$ (2).
Если же X – R-модуль, а форма f билинейна, то (2) $\Longrightarrow$ (1).

Доказательство. Подставляя в (1) $y = x$, получим $2f(x, x) = 0$, что влечет $f(x, x) = 0$, как только 2 не является делителем нуля. Второе утверждение вытекает из следующего вычисления:

$$0 = f(x + y, x + y) = f(x, x) + f(x, y) + f(y, x) + f(y, y) = f(x, y) + f(y, x),$$

откуда $f(x, y) = -f(y, x)$. □

Полилинейная форма называется антисимметричной, если она обращается в ноль, как только два ее аргумента равны. Пусть теперь в лемме 1.2 $m = n = \dim V$, а форма f антисимметрична. Напомним, что $\varepsilon(\sigma)$ обозначает четность перестановки $\sigma \in S_n$, см. определение 9.3.

Лемма 1.5. *Пусть $f : \underbrace{V \times \cdots \times V}_{n \text{ раз}} \to R$ – полилинейная антисимметричная форма, $v = (v_1, \ldots, v_n)$ – базис модуля V, $x = (x_1, \ldots, x_m) \in {}^m V$, а $A \in M_n(R)$ такова, что $x = vA$. Тогда*

$$f(x_1, \ldots, x_n) = f(v_1, \ldots, v_n) \sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} \prod_{i=1}^{n} a_{\sigma(i)\, i}.$$

Доказательство. Так как f антисимметрична, то $f(v_{i_1}, \ldots, v_{i_n}) = 0$ как только $i_k = i_\ell$ при $k \neq \ell$. Таким образом, суммирование в формуле из леммы 1.2 достаточно вести по всем наборам различных индексов $(i_1, \ldots, i_n)$. Пусть $\sigma : \{1, \ldots, n\} \to \{1, \ldots, n\}$ – функция, заданная равенством $\sigma(k) = i_k$. Так как $i_k \neq i_l$ при $k \neq l$, а область определения σ совпадает с ее множеством значений, то σ – биекция, т. е. $\sigma \in S_n$. Заметим, что за счет антисимметричности $f(v_{\sigma(1)}, \ldots, v_{\sigma(n)}) = (-1)^{\varepsilon(\sigma)} f(v_1, \ldots, v_n)$. С учетом этого формула из леммы 1.2 превращается в доказываемое равенство. □

Определение 1.6. Ненулевая антисимметричная n-линейная форма на n-мерном векторном пространстве называется формой объема.

В заключении параграфа сформулируем еще одно полезное свойство полилинейных антисимметричных отображений.

Лемма 1.7. *Пусть f – полилинейное антисимметричное отображение. Тогда его значение не меняется при первом элементарном преобразовании с аргументами, т. е. при любом $\alpha \in R$*

$$f(\ldots, v_i, \ldots, v_j, \ldots) = f(\ldots, v_i + v_j\alpha, \ldots, v_j, \ldots).$$

2. Определение определителя

Определение 2.1. Определитем матрицы $A \in \mathrm{M}_n(R)$ называется число

$$\det A = \sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} a_{\sigma(1)\,1} \dots a_{\sigma(n)\,n}.$$

Мы уже доказали в предыдущем параграфе, что любая форма объема пропорциональна определителю. Как ни странно, из этого не следует, что сам определитель является полилинейной и антисимметричной формой. К счастью, это так, иначе на свете не существовало бы ни одной формы объема!

Лемма 2.2. *Определитель является полилинейной антисимметричной формой столбцов матрицы, а* $\det E = 1$.

Доказательство. Если в формуле из определения 2.1 зафиксировать все элементы матрицы A кроме элементов k-ого столбца, то получится линейная комбинация элементов k-ого столбца, т. е. $\det A = ba_{*k}$ для некоторой строки $b \in {}^nR$. Ясно, что это выражение линейно по a_{*k}, что и доказывает полилинейность. Утверждение $\det E = 1$ очевидно.

Для доказательства антисимметричности предположим, что $a_{*i} = a_{*j}$. Разложим симметрическую группу в объединение смежных классов по знакопеременной: $S_n = A_n \sqcup A_n\tau$, где в качестве τ можно взять любую нечетную перестановку. Для наших целей удобно взять транспозицию $\tau = (ij)$. Тогда

$$\det A = \sum_{\sigma \in A_n} \prod_{k=1}^{n} a_{\sigma(k)\,k} - \sum_{\rho \in A_n\tau} \prod_{k=1}^{n} a_{\rho(k)\,k}. \tag{27}$$

Заменяя $\rho = \sigma\tau$ получим

$$\sum_{\rho \in A_n\tau} \prod_{k=1}^{n} a_{\rho(k)\,k} = \sum_{\sigma \in A_n} \prod_{k=1}^{n} a_{\sigma\tau(k)\,k} = \sum_{\sigma \in A_n} a_{\sigma(j)\,i} a_{\sigma(i)\,j} \prod_{k \neq i,j} a_{\sigma(k)\,k}$$

Учитывая, что $a_{\sigma(j)\,i} = a_{\sigma(j)\,j}$ и $a_{\sigma(i)\,j} = a_{\sigma(i)\,i}$, эта сумма совпадает с первой суммой в формуле (27), откуда $\det A = 0$. □

Следствие 2.3. *Пусть f форма объема на V, v – базис V, а $A \in \mathrm{M}_n(R)$. Тогда*

$$f(vA) = f(v) \det A.$$

Если f – форма объема на R^n, то $f(A) = f(E) \cdot \det A$ для любой матрицы A размера $n \times n$ (здесь матрица отождествляется с набором своих столбцов).

Множество форм объема на данном свободном модуле является свободным модулем ранга 1.

Лемма 2.4. *Пусть f – форма объема на V.*

1. *Если набор векторов v является базисом, то $f(v) \neq 0$. Если R – поле, то верно и обратное.*
2. *Если u и v два базиса модуля V, то $f(u) = f(v) \det C_{v \to u}$.*
3. *Пусть R – поле. Определитель квадратной матрицы не равен нулю тогда и только тогда, когда ее строки (столбцы) линейно независимы.*

Доказательство. Так как $f \neq 0$, то найдется набор векторов $x \in {}^nV$, для которых $f(x) \neq 0$. Если v – базис, то по следствию 2.3 $f(x) = f(v) \det A$, следовательно, $f(v) \neq 0$.

Обратно, если v не базис, а R – поле, то один из элементов выражается в виде линейной комбинации остальных, скажем, $v_i = \sum_{j \neq i} v_j \alpha_j$. По лемме 1.7 в выражении $f(v_1, \dots, v_n)$ можно заменить v_i на $v_i - \sum_{j \neq i} v_j \alpha_j$, т. е. на 0. Следовательно, $f(v) = f(\dots, 0, \dots) = 0$.

Второе утверждение непосредственно следует из следствия 2.3.

Третье утверждение следует из первого, так как определитель является формой объема на R^n, а n элементов n-мерного пространства являются базисом тогда и только тогда, когда они линейно независимы. □

Напомним, что для линейного отображения $L : V \to V$ и набора $x = (x_1, \dots, x_n) \in {}^nV$ выражение $L(x)$ обозначает строку $\big(L(x_1), \dots, L(x_n)\big)$.

Лемма 2.5. *Пусть $L : V \to V$ – эндоморфизм модуля V, f форма объема на V, а v – базис V.*

- *Фунция $f_L : V \times \dots \times V \to R$, заданная формулой $f_L\big(x\big) = f\big(L(x)\big)$ является формой объема или тождественно равна нулю.*
- *Отношение $f_L(v_1, \dots, v_n)/f(v_1, \dots, v_n)$ не зависит от выбора формы объема и базиса и равно определителю матрицы L_v.*

Доказательство. Для доказательства пункта (1) достаточно проверить, что форма f_L полилинейна и антисимметрична, что не составляет труда. Заметим, что по лемме 2.4 $f(v) \neq 0$, так что частное из пункта (2) всегда имеет смысл. Если u – другой базис пространства V, то по той же лемме

$$\frac{f_L(u)}{f(u)} = \frac{f_L(v)\det C_{v\to u}}{f(v)\det C_{v\to u}} = \frac{f_L(v)}{f(v)}.$$

Если g – другая форма объема, то $g_L\big(v\big) = g\big(L(v)\big) = g\big(v\big)\det L_v$ в соответствии со следствием 2.3. □

Определение 2.6. Определителем эндоморфизма $L : V \to V$ называется коэффициент изменения формы объема, т. е. отношение $f_L(v)/f(v)$ из леммы 2.5.

3. Свойства определителя

Определение 3.1. Пусть B – матрица размера $n \times n$, а i и j – индексы от 1 до n. Обозначим через $M^{(ij)}$ или $M^{(ij)}(B)$ матрицу, полученную из B вычеркиванием i-ой строки и j-ого столбца. *Минором* в позиции (i, j) матрицы B называется число $M_{ij} = \det M^{(ij)}$. *Алгебраическим дополнением* позиции (i, j) матрицы B, называется число $A_{ij} = (-1)^{i+j}M_{ij}$. В том случае, когда хочется явно указать, для какой матрицы вычисляется алгебраическое дополнение (минор), его обозначают через $A_{ij}(B)$ (соотв. $M_{ij}(B)$).

Пусть B – матрица размера $n \times n$, а j – индекс от 1 до n.

Теорема 3.2.

(1) *Определитель матрицы с нулевым столбцом (строкой) равен нулю. Определитель матрицы с пропорциональными столбцами (строками) равен нулю.*
(2) *Определитель не изменяется при первом преобразовании Гаусса. Общий множитель столбца (строки) выносится за знак определителя. При транспозиции столбцов (строк) матрицы ее определитель меняет знак.*
(3) *Определитель композиции эндоморфизмов равен произведению их определителей. Определитель произведения квадратных матриц равен произведению их определителей.*
(4) $\det B = \det B^{\intercal}$. *Поэтому все свойства, сформулированные для столбцов матрицы B верны и для ее строк.*
(5) *Определитель клеточно треугольной матрицы равен произведению определителей диагональных блоков. В частности, определитель треугольной матрицы равен произведению диагональных элементов.*
(6) $\det B = \sum_{i=1}^{n} b_{ji}A_{ji} = \sum_{i=1}^{n} b_{ij}A_{ij}$.

Доказательство. Свойства (1)–(2) вытекают того, что определитель является формой объема (лемма 2.2) и леммы 1.7.

Пусть R – поле, а $L, M : V \to V$ – линейные операторы. Если M необратим, то и $L \circ M$ необратим. Необратимый оператор отображает базис в линейно зависимую систему векторов, от которого любая форма объема равна нулю. Следовательно, в этом случае обе части равенства $\det L \circ M = \det L \cdot \det M$ равны нулю. В противном случае, M отображает базис v в базис $M(v)$. Для формы объема f имеем

$$\det L \circ M = \frac{f\Big(L\big(M(v)\big)\Big)}{f\big(v\big)} = \frac{f\Big(L\big(M(v)\big)\Big)}{f\big(M(v)\big)} \cdot \frac{f\Big(M(v)\Big)}{f\big(v\big)} = \det L \cdot \det M.$$

Композиции операторов соответствует произведение матриц, а определитель оператора равен определителю его матрицы. Так как любая матрица является матрицей какого-то оператора, то из этого следует второе утверждение пункта (3).

Пусть теперь $P = \mathbb{Z}[x_{11}, y_{11}, \ldots, x_{nn}, y_{nn}]$ – кольцо многочленов от $2n^2$ переменных, а $X, Y \in \mathrm{M}_n(P)$ – матрицы с элементами x_{ij} и y_{ij} соответственно. Из доказанного следует, что равенство $\det(XY) = \det X \det Y$ имеет место над полем частных кольца P, а, следовательно, и над самим кольцом P.

Пусть теперь R – произвольное кольцо, а $\varphi : P \to R$ – единственный гомоморфизм, отображающий x_{ij} в a_{ij}, а y_{ij} в b_{ij}. Применяя φ к обеим частям равенства $\det(XY) = \det X \det Y$, получаем $\det(AB) = \det A \det B$.

Для доказательства свойства (4) применим формулу из определения 2.1 к матрице $A^\intercal$:

$$\det A^\intercal = \sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} a_{1\,\sigma(1)} \ldots a_{n\,\sigma(n)}.$$

Переставив сомножители $a_{1\,\sigma(1)}, \ldots, a_{n\,\sigma(n)}$ в соответствии с перестановкой σ^{-1} получим:

$$\begin{aligned}\det A^\intercal &= \sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} a_{\sigma^{-1}(1),\sigma(\sigma^{-1}(1))} \ldots a_{\sigma^{-1}(n),\sigma(\sigma^{-1}(n))} = \\ &= \sum_{\sigma \in S_n} (-1)^{\varepsilon(\sigma)} a_{\sigma^{-1}(1),1} \ldots a_{\sigma^{-1}(n),n}.\end{aligned}$$

Так как ε является гомоморфизмом $S_n \to \mathbb{Z}_2$, то $\varepsilon(\sigma) = \varepsilon(\sigma^{-1})$. С другой стороны, отображение $S_n \to S_n$, заданное правилом $\sigma \mapsto \sigma^{-1}$, биективно (оно обратно самому себе). Поэтому в последней сумме σ^{-1} пробегает все множество S_n. Таким образом,

$$\det A^\intercal = \sum_{\sigma^{-1} \in S_n} (-1)^{\varepsilon(\sigma^{-1})} a_{\sigma^{-1}(1),1} \ldots a_{\sigma^{-1}(n),n} = \det A.$$

Для доказательства свойства (5) рассмотрим сначала случай $A = \left(\begin{smallmatrix} E & * \\ 0 & E \end{smallmatrix}\right)$. Так как эта матрица легко получается из единичной с помощью серии первых преобразований Гаусса, то ее определитель равен 1.

Рассмотрим теперь n-форму f на R^n, сопоставляющую квадратной матрице B число $f(B) = \det \left(\begin{smallmatrix} B & * \\ 0 & E \end{smallmatrix}\right)$. Легко проверить, что f – полилинейная антисимметричная форма. По следствию 2.3 $f(B) = \det B \cdot f(E)$, что равно $\det B$ в соответствии с первым абзацем доказательства. Из свойства (4) легко вывести теперь, что $\det \left(\begin{smallmatrix} B & 0 \\ * & E \end{smallmatrix}\right)$ также равен $\det B$.

В качестве следующего шага доказательства зафиксируем квадратную матрицу B и рассмотрим m-форму g на R^m, заданную формулой $g(C) = \det \left(\begin{smallmatrix} B & 0 \\ * & C \end{smallmatrix}\right)$, где C – квадратная матрица $m \times m$. Снова очевидно, что g – полилинейная антисимметричная форма, и по следствию 2.3 $g(C) = \det C \cdot g(E)$, а $g(E) = \det B$ по предыдущему абзацу доказательства.

Наконец, пусть

$$A = \begin{pmatrix} A^{(1)} & 0 & 0 \\ * & \ddots & 0 \\ * & * & A^{(k)} \end{pmatrix}.$$

Докажем индукцией по k, что $\det A = \det A^{(1)} \dots \det A^{(k)}$. При $k = 1$ доказывать нечего. При $k > 1$ обозначим $C = A^{(k)}$ и

$$B = \begin{pmatrix} A^{(1)} & 0 & 0 \\ * & \ddots & 0 \\ * & * & A^{(k-1)} \end{pmatrix}.$$

По индукционному предположению $\det B = \det A^{(1)} \dots \det A^{(k-1)}$, а по предыдущему абзацу доказательства $\det A = \det B \cdot \det C = \det A^{(1)} \dots \det A^{(k)}$.

Таким образом, свойство доказано для нижних клеточно треугольных матриц. Доказательство для верхних клеточно треугольных матриц легко следует теперь из свойства (4).

Докажем теперь формулу $\det B = \sum_{i=1}^{n} b_{ij} A_{ij}$. По линейности по j-му столбцу $\det B$ раскладывается в сумму определителей

$$\det \begin{pmatrix} b_{11} & \cdots & b_{1\,j-1} & 0 & b_{1\,j+1} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{i-1\,1} & \cdots & b_{i-1\,j-1} & 0 & b_{i-1\,j+1} & \cdots & b_{i-1\,n} \\ b_{i\,1} & \cdots & b_{i\,j-1} & b_{ij} & b_{i\,j+1} & \cdots & b_{in} \\ b_{i+1\,1} & \cdots & b_{i+1\,j-1} & 0 & b_{i+1\,j+1} & \cdots & b_{i+1\,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ b_{n\,1} & \cdots & b_{n\,j-1} & 0 & b_{n\,j+1} & \cdots & b_{n\,n} \end{pmatrix}$$

Производя циклическую перестановку $(1\,2\,\dots\,i)$ строк и циклическую перестановку $(1\,2\,\dots\,j)$ столбцов последней матрицы, получим

$$(-1)^{i-1+j-1} \det \begin{pmatrix} b_{ij} & b_{i\,1} & \cdots & b_{i\,j-1} & b_{i\,j+1} & \cdots & b_{in} \\ 0 & b_{11} & \cdots & b_{1\,j-1} & b_{1\,j+1} & \cdots & b_{1n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & b_{i-1\,1} & \cdots & b_{i-1\,j-1} & b_{i-1\,j+1} & \cdots & b_{i-1\,n} \\ 0 & b_{i+1\,1} & \cdots & b_{i+1\,j-1} & b_{i+1\,j+1} & \cdots & b_{i+1\,n} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & b_{n\,1} & \cdots & b_{n\,j-1} & b_{n\,j+1} & \cdots & b_{n\,n} \end{pmatrix}$$

что по свойству (5) равно $b_{ij} A_{ij}(B)$. Разложение по строке легко вывести из разложения по столбцу при помощи свойства (4). □

4. Формула для элементов обратной матрицы, формулы Крамера и минорный ранг

В этом параграфе мы выведем формулы для элементов обратной матрицы через алгебраические дополнения и определитель исходной. Первым шагом является следующее предложение.

Предложение 4.1. *Сумма произведений элементов столбца (строки) матрицы на алгебраические дополнения другого столбца (строки) равна нулю. Точнее, если $j \neq k$, то*

$$\sum_{i=1}^{n} b_{ij} A_{ik} = \sum_{i=1}^{n} b_{ji} A_{ki} = 0$$

Доказательство. Заменим k-й столбец матрицы B на j-й, оставив все остальное без изменений, т.е. рассмотрим матрицу $\tilde{B}$ с элементами $\tilde{b}_{im} = b_{im}$ при $m \neq k$ и $\tilde{b}_{ik} = b_{ij}$. В полученной матрице будет два одинаковых столбца, следовательно, ее определитель будет равен нулю. С другой стороны, заметим, что алгебраические дополнения элементов k-го столбца не зависят от элементов этого столбца, поэтому $A_{ik}(\tilde{B}) = A_{ik}(B)$. Раскладывая, по свойству 3.2(6), определитель матрицы $\tilde{B}$ по k-ому столбцу, получим $0 = \det \tilde{B} = \sum_{i=1}^{n} \tilde{b}_{ik} A_{ik} = \sum_{i=1}^{n} b_{ij} A_{ik}$. Доказательство второго равенства (для строк) совершенно аналогично. □

Определение 4.2. Матрица называется *невырожденной*, если она квадратная, а ее определитель обратим. Квадратная матрица с нулевым определителем называется *вырожденной*.

Лемма 4.3. *Если матрица A обратима, то она невырождена.*

Доказательство. Мы уже доказывали, что обратимые матрицы обязательно квадратные. Если A^{-1} и A – квадратные, то по свойству 3.2(3) имеем $1 = \det E = \det(A^{-1}A) = \det(A^{-1}) \cdot \det A$. □

Таким образом, $\det_n : \mathrm{GL}_n(R) \to R^*$ является гомоморфизмом групп.

Определение 4.4. Пусть B – матрица размера $n \times n$. Присоединенной к B называется матрица B^{ad}, транспонированная к матрице из алгебраических дополнений матрицы B, т.е. элемент матрицы B^{ad} в позиции (i, j) равен $A_{ji}(B)$.

Теорема 4.5. *Если $B \in \mathrm{M}_n(R)$, то*

$$BB^{\mathrm{ad}} = B^{\mathrm{ad}}B = E \det B.$$

В частности, если матрица B невырождена, то она обратима, и

$$B^{-1} = \frac{1}{\det B} B^{\mathrm{ad}}.$$

Доказательство. Положим $D = B \cdot B^{\mathrm{ad}}$. Элемент матрицы B^{ad} в позиции (k, j) равен A_{jk}. Получаем: $d_{ij} = \sum_{k=1}^{n} b_{ik}A_{jk}$. При $i = j$, по свойству 3.2(6), $d_{ij} = \det B$, а при $i \neq j$, по предложению 4.1, $d_{ij} = 0$. Таким образом, $D = E \det B$. Аналогично, $B^{\mathrm{ad}}B = E \det B$. Второе утверждение сразу вытекает из первого. □

Теорема 4.6 (Формулы Крамера)**.** *Пусть $A \in \mathrm{GL}_n(R)$, а $b \in R^n$. Обозначим через Δ определитель матрицы A, а через Δ_i – определитель матрицы, полученной из A заменой i-ого столбца столбцом b. Система линейных уравнений $Ax = b$ имеет единственное решение $x_i = \frac{\Delta_i}{\Delta}$.*

Доказательство. Умножая равенство $Ax = b$ слева на A^{-1} получим $x = A^{-1}b$. Подставляя формулу для обратной матрицы из теоремы 4.5, получим $x = \frac{1}{\Delta}A^{\mathrm{ad}}b$ или $x_i = \frac{1}{\Delta}\sum_{k=1}^{n} A_{ki}b_k$. Осталось заметить, что по свойству 3.2(6) $\sum_{k=1}^{n} A_{ki}b_k = \Delta_i$. □

Если R – поле, а $A \in \mathrm{M}_n(R)$ вырождена, то оператор $R^n \to R^n$ умножения на матрицу A необратим. Следовательно, он не инъективен и не сюръективен, т. е. система имеет либо ни одного, либо бесконечно много решений.

Далее, до конца главы R является полем. В этом случае следующее определение ранга матрицы более общепринято, чем строчный или столбцовый ранг.

Определение 4.7. Минорным рангом матрицы $A \in \mathrm{M}_{m,n}(R)$ называется наибольший размер квадратной подматрицы, определитель которой не равен нулю.

Теорема 4.8. *Над полем минорный ранг матрицы равен ее строчному (столбцовому) рангу.*

Доказательство. Обозначим через k минорный ранг A, а через r – ее строчный ранг. По определению минорного ранга существует квадратная подматрица размера $k \times k$, определитель которой не равен нулю. По лемме 2.4 строки этой подматрицы линейно независимы, а значит, линейно независимы и k строк матрицы A, в которых стоит выбранная подматрица. Следовательно, $k \leqslant r$.

Обратно, возьмем подматрицу $r \times n$, состоящую из линейно независимых строк матрицы A. По теореме 11.4 главы 2 столбцовый ранг этой подматрицы также равен r, следовательно, в ней найдется r линейно независимых столбцов. Наконец, по лемме 2.4 квадратная матрица с линейно независимыми столбцами имеет ненулевой определитель, откуда $k \geqslant r$. □

Глава 6

Собственные числа и жорданова форма

1. Собственные числа и вектора

В этой главе V обозначает векторное пространство над полем F. Мы будем искать базис, в котором матрица фиксированного оператора $L : V \to V$ выглядит наиболее просто. В частности, этот параграф посвящен ситуации, когда эта матрица диагональна.

Определение 1.1. Оператор называется диагонализуемым, если существует базис, в котором его матрица диагональна. Матрица называется диагонализуемой, если диагонализуем оператор умножения на эту матрицу.

Из определения матрицы оператора следует, что если

$$L_u = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

то базисные вектора базиса $u = (u_1, \ldots, u_n)$ удовлетворяют условию $L(u_k) = \lambda_k u_k$. Это наблюдение является мотивировкой для следующего определения.

Определение 1.2. Ненулевой вектор $x \in V$ называется собственным вектором оператора $L : V \to V$, соответствующим числу $\lambda \in F$, если

$$L(x) = x\lambda.$$

При этом λ называется собственным числом.

Столбец $b \in F^n$ называется собственным вектором матрицы $A \in \mathrm{M}_n(F)$, соответствующим числу $\lambda \in F$, если он является собственным вектором оператора умножения на эту матрицу, т. е. $Ab = b\lambda$.

Ясно, что вектор $x \in V$ является собственным вектором оператора L тогда и только тогда, когда столбец x_v является собственным вектором матрицы L_v (где v – произвольный базис пространства V). Из этого следует также, что собственные числа оператора L и его матрицы в любом базисе совпадают.

Предложение 1.3. *Число $\lambda \in F$ является собственным числом оператора $L : V \to V$ тогда и только тогда, когда* $\det(L - \lambda \cdot \mathrm{id}) = 0$.

Доказательство. Уравнение $L(x) = x\lambda$ равносильно уравнению $(L - \lambda \cdot \mathrm{id})(x) = 0$, которое имеет ненулевое решение тогда и только тогда, когда $\det(L - \lambda \cdot \mathrm{id}) = 0$. Действительно, определитель равен нулю тогда и только тогда, когда оператор не обратим, а это равносильно тому, что он имеет ненулевое ядро. □

Если A – матрица оператора L в каком-то базисе, то $\det(L - \lambda \cdot \mathrm{id}) = \det(A - \lambda E)$. Матрица $A - tE$, принадлежит матричному кольцу $M_n(F[t])$ над кольцом многочленов $F[t]$.[1] Поэтому ее определитель является многочленом. Из формулы для определителя следует, что степень этого

[1]Если быть абсолютно строгим, то она принадлежит кольцу $M_n(F)[t]$ многочленов с матричными коэффициентами, но эти кольца очевидным образом изоморфны и мы отождествляем их.

многочлена равна n (моном старшей степени возникает из произведения диагональных элементов и равен $(-1)^n t^n$).

Определение 1.4. Многочлен $\det(L - t \cdot \mathrm{id})$ называется характеристическим многочленом оператора L и обозначается через χ_L. Характеристический многочлен матрицы A равен $\det(A - tE)$.

Как следует из предложения 1.3, собственные числа матрицы или оператора A и только они являются корнями характеристического многочлена.

Так как у многочлена n-ой степени не может быть больше, чем n корней, то у оператора в n-мерном пространстве может быть максимум n собственных чисел. Оказывается, что если их ровно n, то оператор диагонализуем. Это вытекает из следующего утверждения.

Теорема 1.5. *Собственные вектора, соответствующие различным собственным числам, линейно независимы.*

Доказательство. Пусть $x_1, \ldots, x_m$ – собственные вектора оператора L, соответствующие различным собственным числам $\lambda_1, \ldots, \lambda_m$, т. е. выполнены равенства $L(x_k) = \lambda_k x_k$. Проведем доказательство индукцией по m. При $m = 1$ утверждение следует из того, что собственный вектор по определению не равен 0.

Пусть $m > 1$, и

$$\sum_{k=1}^{m} x_k \alpha_k = 0.$$

Применяя к этому равенству оператор L, получим

$$\sum_{k=1}^{m} L(x_k)\alpha_k = \sum_{k=1}^{m} x_k \lambda_k \alpha_k = 0,$$

а умножая его на λ_m:

$$\sum_{k=1}^{m} x_k \lambda_m \alpha_k = 0.$$

Вычитая последнее равенство из предпоследнего, имеем:

$$\sum_{k=1}^{m} x_k(\lambda_k - \lambda_m)\alpha_k = \sum_{k=1}^{m-1} x_k(\lambda_k - \lambda_m)\alpha_k = 0.$$

По индукционному предположению набор из $m-1$ собственного вектора линейно независим, поэтому $(\lambda_k - \lambda_m)\alpha_k = 0$ при всех $k = 1, \ldots, m-1$. Так как $\lambda_k \neq \lambda_m$ при $k \neq m$, то $\alpha_k = 0$ при всех $k = 1, \ldots, m-1$. Подставляя эти значения в исходную формулу, получаем $x_m \alpha_m = 0$, откуда $\alpha_m = 0$. □

Если у многочлена n-й степени нет n корней даже с учетом кратности, то это можно исправить, расширяя базовое поле. Следующее определение посвящено ситуации, когда характеристический многочлен имеет кратные корни.

Определение 1.6. Алгебраической кратностью собственного числа называется его кратность в характеристическом многочлене.

Собственным подпространством, соответствующим собственному числу λ, называется ядро оператора $L - \lambda \cdot \mathrm{id}$. Другими словами, собственное подпространство – это множество собственных векторов, соответствующих данному собственному числу, дополненное нулем.

Геометрической кратностью собственного числа называется размерность собственного подпространства.

Лемма 1.7. *Геометрическая кратность собственного числа не превосходит его алгебраической кратности.*

Доказательство. Обозначим через $k = \dim \mathrm{Ker}(L - \lambda \cdot \mathrm{id})$ геометрическую кратность собственного числа λ. Выберем базис $(u_1, \dots, u_k)$ пространства $\mathrm{Ker}(L - \lambda \cdot \mathrm{id})$ и дополним его до базиса $u = (u_1, \dots, u_n)$ всего пространства V. Так как $L(u_i) = u_i\lambda$ при всех $i = 1, \dots, k$, то первые k столбцов матрицы L_u совпадают с соответствующими столбцами матрицы λE. Поэтому $\chi_L = \det(L_u - tE)$ делится на $(\lambda - t)^k$, следовательно, алгебраическая кратность λ не меньше k. □

В следующей теореме собраны различные условия диагонализуемости оператора.

Теорема 1.8. *Пусть $L : V \to V$ – оператор на n-мерном пространстве V.*

(1) *L диагонализуем тогда и только тогда, когда существует базис из его собственных векторов (такой базис называется собственным базисом оператора).*
(2) *L диагонализуем тогда и только тогда, когда V равно прямой сумме собственных подпространств.*
(3) *Если существует n различных собственных чисел оператора L, то он диагонализуем (это условие не является необходимым).*
(4) *Предположим, что поле F алгебраически замкнуто. Оператор L диагонализуем тогда и только тогда, когда геометрическая кратность каждого собственного числа равна его алгебраической кратности.*
(5) *Если характеристический многочлен не имеет кратных корней, а поле F алгебраически замкнуто, то оператор диагонализуем.*

Доказательство. **1.** Это утверждение обсуждалось в самом начале параграфа.

2. Если V равно прямой сумме собственных подпространств, то объединение базисов этих подпространств является базисом пространства V, состоящим из собственных векторов. Обратно, каждый собственный вектор лежит в каком-то собственном подпространстве. Поэтому если существует базис из собственных векторов, то V является суммой собственных подпространств. Тот факт, что сумма прямая, следует из теоремы о линейной независимости собственных векторов.

3. Пусть оператор L имеет n различных собственных чисел. Выберем по одному собственному вектору для каждого собственного числа. По теореме о линейной независимости собственных векторов они линейно независимы, а так как их число равно размерности пространства, то они образуют базис.

4. Так как поле алгебраически замкнуто, то сумма алгебраических кратностей собственных чисел равна степени характеристического многочлена, которая равна размерности пространства. Если геометрические кратности равны алгебраическим, то сумма размерностей собственных подпространств равна размерности пространства. Из теоремы о линейной независимости собственных векторов следует, что сумма собственных подпространств является прямой. По следствию о размерности прямой суммы, размерность суммы собственных подпространств равна сумме их размерностей, т. е. размерности пространства. Теперь диагонализуемость оператора следует из пункта 2.

Обратно, если L диагонализуем, то n равно сумме геометрических кратностей и сумме алгебраических кратностей. Так как геометрические кратности не превосходят алгебраических, то они должны быть равны.

5. Если характеристический многочлен не имеет кратных корней, то над замкнутым полем он имеет n различных корней, и утверждение следует из пункта 3. □

2. Жорданова форма и теорема Гамильтона–Кэли

Предыдущая теорема намекает на то, что даже над алгебраически замкнутым полем существуют недиагонализуемые операторы. В этом параграфе мы сформулируем утверждение о том, что можно и их привести к довольно простому виду, называемому жордановой формой оператора.

Определение 2.1. Обозначим через $J = J_n$ матрицу размера $n \times n$ с единицами во всех позициях $(k, k+1)$, $k = 1, \dots, n-1$, и остальными нулями. Жордановым блоком называется матрица $\lambda E + J$. Жордановой матрицей называется блочно диагональная матрица с жордановыми блоками по диагонали.

Теорема 2.2. *Пусть V – конечномерное векторное пространство над алгебраически замкнутым полем F, а $L : V \to V$ – линейный оператор. Тогда существует базис и пространства V такой, что матрица L_u является жордановой. Она называется жордановой формой оператора L и определена единственным образом с точностью до перестановки блоков.*

Доказательство этой теоремы будет дано после изучении модулей над областями главных идеалов, а сейчас мы выведем некоторые следствия этой теоремы. Для этого нам понадобится несколько простых утверждений.

Лемма 2.3. *Матрица J_n^m имеет 1 во всех позициях $(k, k+m)$, $1 \leqslant k \leqslant n-m$.*
Если $N \in \mathrm{M}_m(F)$ – верхнетреугольная матрица с нулями по главной диагонали, то $N^m = 0$.

Следующая лемма позволяет во многих случаях заменять матрицу на ее жорданову форму.

Лемма 2.4. *Отображение $A \mapsto C^{-1}AC$ является автоморфизмом матричного кольца, следовательно, для любого многочлена p: $p(C^{-1}AC) = C^{-1}p(A)C$.*

Следующее утверждение показывает, как работать с блочно-диагональными матрицами. Обозначим через $\mathrm{diag}(A^{(1)}, \dots, A^{(m)})$ диагональную матрицу с квадратными диагональными блоками $A^{(1)}, \dots, A^{(m)}$.

Лемма 2.5. *Множество блочно диагональных матриц с квадратными диагональными блоками фиксированного размера является подкольцом матричного кольца, изоморфным прямой сумме матричных колец соответствующих размеров. В частности, для любого многочлена p*

$$p\big(\mathrm{diag}(A^{(1)}, \dots, A^{(m)})\big) = \mathrm{diag}\big(p(A^{(1)}), \dots, p(A^{(m)})\big).$$

Так как жорданова форма имеет место только над замкнутым полем, мы должны научиться вкладывать любое поле в алгебраически замкнутое. Доказательство этого факта будет дано в главе 13.

Предложение 2.6. *Любое поле F вкладывается в алгебраически замкнутое поле $\overline{F}$.*

Первое применение жордановой формы – доказательство следующего утверждения.

Теорема 2.7 (теорема Гамильтона–Кэли)**.** *Пусть A – квадратная матрица или линейный оператор на конечномерном пространстве. Тогда $\chi_A(A) = 0$.*

Доказательство. Ясно, что доказывать можно только для случая матриц. Пусть $A \in \mathrm{M}_n(F)$. По предыдущему предложению можно считать, что $A \in \mathrm{M}_n(\overline{F})$, и по теореме 2.2 существует $C \in \mathrm{M}_n(\overline{F})$ такая, что $A = C^{-1}A'C$, где A' – жорданова форма матрицы A (матрица оператора умножения на A в жордановом базисе). По лемме 2.4 равенство $\chi_A(A) = 0$ равносильно равенству $\chi_A(A') = 0$, а так как A и A' – матрицы одного и того же оператора в разных базисах, то $\chi_A = \chi_{A'}$. По определению $A' = \mathrm{diag}(\lambda_1 E_{k_1} + J_{k_1}, \dots, \lambda_m E_{k_m} + J_{k_m})$ для некоторых $m, k_1, \dots, k_m \in \mathbb{N}$ и $\lambda_1, \dots, \lambda_m \in \overline{F}$. Тогда $\chi_{A'}(t) = \prod_{i=1}^m (\lambda_i - t)^{k_i}$. Если подставить i-й диагональный блок матрицы A' в i-й сомножитель многочлена $\chi_{A'}$, то в соответствие с леммой 2.3 получится 0 (напомним, что при подстановке матрицы в многочлен свободный член многочлена умножается на единичную матрицу). Следовательно, значение $\chi_{A'}$ на любом диагональном блоке матрицы A' равно 0. Теперь требуемое равенство следует из леммы 2.5 □

Полезно знать и непосредственное доказательство теоремы Гамильтона–Кэли, без использования жордановой формы.

ВТОРОЕ ДОКАЗАТЕЛЬСТВО ТЕОРЕМЫ ГАМИЛЬТОНА–КЭЛИ. Заметим, что матричное кольцо $\mathrm{M}_n(F[t])$ над кольцом многочленов изоморфно кольцу $\mathrm{M}_n(F)[t]$ многочленов с матричными коэффициентами.[2] Рассмотрим матрицу $A - tE \in \mathrm{M}_n(F[t])$. По теореме 4.5 главы 5

$$(A - tE)(A - tE)^{\mathrm{ad}} = (A - tE)^{\mathrm{ad}}(A - tE) = \det(A - tE)E = \chi_A(t)E. \tag{28}$$

Переписывая это равенство в кольце $\mathrm{M}_n(F)[t]$ получаем

$$(A - tE)(B_0 + B_1 t + \cdots + B_{n-1}t^{n-1}) = \chi_A(t)E,$$

где $B_i \in \mathrm{M}_n(F)$ – коэффициенты в разложении матрицы $(A - tE)^{\mathrm{ad}}$ по степеням t.

Как видно, для доказательства теоремы достаточно подставить в это равенство A вместо t. Однако, такая подстановка законна только если A лежит в центре кольца коэффициентов. Действительно, формально, подстановка элемента a алгебры R в многочлен – это вычисление значения гомоморфизма $\varphi_a : R[t] \to R$, посылающего t в a и оставляющего на месте все элементы R.

Но t коммутирует со всеми элементами кольца R, а гомоморфизм сохраняет это свойство. Обратно, если a коммутирует со всеми элементами из R, то простая проверка показывает, что формула $\varphi_a(r_0 + \cdots + r_m t^m) = r_0 + \cdots + r_m a^m$ задает требуемый гомоморфизм (это небольшое обобщение универсального свойства кольца многочленов 16.2 главы 4).

Определим R, как минимальную F-подалгебра в $\mathrm{M}_n(F)$, содержащую матрицы A и $B_1, \ldots, B_{n-1}$. Так как $A - tE$ коммутирует с $(A - tE)^{\mathrm{ad}}$, то и A коммутирует с этой матрицей. Следовательно, $A(B_0 + \cdots + B_{n-1}t^{n-1}) = (B_0 + \cdots + B_{n-1}t^{n-1})A$, откуда сразу следует, что $AB_i = B_iA$ при всех i. Таким образом, A лежит в центре алгебры R и, значит, ее можно подставить вместо t в равенство (28). В левой части этого равенства, очевидно, получается 0, следовательно, и правая часть $\chi_A(A)E = 0$. □

Пусть R – ассоциативная (возможно некоммутативная) алгебра с 1 над полем F. Напомним, что для любого $r \in R$ существует единственный гомоморфизм F-алгебр

$$\varepsilon_r : F[t] \to R, \qquad p \mapsto p(r),$$

отображающий независимую переменную t в r. Ядро этого гомоморфизма – идеал кольца $F[t]$. Так как $F[t]$ – область главных идеалов, то $\operatorname{Ker}\varepsilon_r$ – главный идеал. Образующая этого идеала называется минимальным многочленом элемента r и обозначается φ_r. Этот многочлен определен однозначно с точностью до умножения на обратимый элемент кольца $F[t]$, т. е. ненулевую константу.

По теореме о гомоморфизме наименьшая подалгебра с 1 в R, содержащая r, изоморфна $F[t]/(\varphi_r)$. Она обозначается через $F[r]$. Таким образом, каждому многочлену от r (элементу $F[r]$) однозначно сопоставляется смежный класс этого многочлена в $F[t]/(\varphi_r)$. Как мы знаем, этот смежный класс имеет единственного представителя наименьшей степени, а именно, остаток от деления на φ_r.

Пусть теперь $R = M_n(F)$, а $A \in R$.

Лемма 2.8. *Характеристический многочлен матрицы (оператора) A делится на минимальный. Любое собственное число является корнем минимального многочлена.*

ДОКАЗАТЕЛЬСТВО. Первое утверждение сразу следует из теоремы Гамильтона–Кэли. Второе несложно вывести из теоремы о жордановой форме, но мы дадим непосредственное доказательство. Действительно, если $Ax = \lambda x$, то $A^n x = \lambda^n x$ и, следовательно, $p(A)x = p(\lambda)x$ для любого многочлена $p \in F[t]$. Следовательно, $0 = \varphi_A(A)x = \varphi_A(\lambda)x$, откуда (т. к. $x \neq 0$) следует, что $\varphi_A(\lambda) = 0$. □

[2] Мы определяли только кольцо многочленов над коммутативным кольцом, но определение над некоммутативным кольцом ничем не отличается. Важно только понимать, что независимая переменная t коммутирует со всеми элементами кольца.

Замечание 2.9. С использованием жордановой формы нетрудно доказать, что кратность собственного числа в минимальном многочлене матрицы равна размеру наибольшей жордановой клетки с этим собственным числом.

3. Другое доказательство жордановой формы

Основным доказательством теоремы о жордановой форме в нашем курсе будет доказательство, использующее строение конечнопорожденных модулей над кольцом многочленов. Это будет сделано в параграфе 9. А сейчас мы приведем набросок другого доказательства. Первым шагом является теорема о ядре произведения многочленов от оператора. Напомним, что умножение в алгебре операторов на векторном пространстве – это композиция.

Теорема 3.1. *Пусть $L : V \to V$ – линейный оператор на векторном пространстве V над полем F (V не предполагается конечномерным), а $p, q \in F[t]$ – взаимно простые многочлены. Тогда*

$$\operatorname{Ker}\bigl(p(L)q(L)\bigr) = \operatorname{Ker} p(L) \oplus \operatorname{Ker} q(L).$$

Доказательство. Так как p и q взаимно просты, существуют многочлены $f, h \in F[t]$ такие, что $pf + qh = 1$. Подставляя в это равенство оператор L, получим $f(L)p(L) + h(L)q(L) = \mathrm{id}$. Действуя на вектор $x \in V$, получим $f(L)p(L)(x) + h(L)q(L)(x) = x$. Если $x \in \operatorname{Ker}\bigl(p(L)q(L)\bigr)$, то первое слагаемое лежит в ядре оператора $q(L)$, так как $q(L)f(L)p(L)(x) = f(L)\bigl(p(L)q(L)\bigr)(x) = f(L)(0) = 0$. Аналогично второе слагаемое лежит в ядре оператора $p(L)$. Таким образом, $\operatorname{Ker}\bigl(p(L)q(L)\bigr) = \operatorname{Ker} p(L) + \operatorname{Ker} q(L)$. Если же $x \in \operatorname{Ker} p(L) \cap \operatorname{Ker} q(L)$, то $x = f(L)p(L)(x) + h(L)q(L)(x) = 0$, следовательно, сумма прямая. □

Пусть теперь поле алгебраически замкнуто. Минимальный многочлен оператора L раскладывается на линейные множители

$$\varphi_L(t) = \prod_{i=1}^{m} (t - \lambda_i)^{k_i},$$

где $\lambda_1, \ldots, \lambda_m$ – все различные собственные числа оператора L.

Используя последнюю теорему, разложим пространство в прямую сумму подпространств

$$V = \operatorname{Ker} \varphi_L(L) = \bigoplus_{i=1}^{m} \operatorname{Ker}(L - \lambda_i \cdot \mathrm{id})^{k_i}.$$

Подпространство $\operatorname{Ker}(L - \lambda_i \cdot \mathrm{id})^{k_i}$ называется корневым подпространством оператора L, соответствующим собственному числу λ_i. Заметим, что ничего не изменится, если мы заменим k_i на большие числа, например на $n = \dim V$. Таким образом, корневые подпространства можно определить как $\operatorname{Ker}(L - \lambda_i \cdot \mathrm{id})^n$, не упоминая минимальный многочлен.

Обозначим через A сужение оператора $L - \lambda_i \cdot \mathrm{id}$ на корневое подпространство $U = \operatorname{Ker}(L - \lambda_i \cdot \mathrm{id})^{k_i}$. Для окончания доказательства теоремы о жордановой форме достаточно научиться выбирать базис u пространства U такой, что A_u состоит из жордановых блоков. Так как оператор A нильпотентен, то его собственные числа равны 0. Таким образом, необходимо выбрать базис, удовлетворяющий условиям $A(u_j) = u_{j-1}$ или $A(u_j) = 0$ для всех $j = 1, \ldots, k_j$. Набор ненулевых векторов $v_1, \ldots, v_k$ таких, что $A(v_1) = 0$, а $A(v_j) = v_{j-1}$ при $j = 2, \ldots, k$ называется жордановой цепочкой. Следующее утверждение заканчивает доказательство.

Теорема 3.2. *Для любого нильпотентного оператора над произвольным полем существует базис, состоящий из жордановых цепочек.*

Доказательство этой теоремы не очень сложно, но требует аккуратной организации индукции по размерности пространства.

4. Разложение Жордана

Элемент r кольца R называется нильпотентным, если $r^m = 0$ для некоторого натурального m. Наименьшее такое m называется степенью нильпотентности элемента r.

Лемма 4.1. *Сумма коммутирующих между собой нильпотентных элементов кольца нильпотентна.*

Доказательство. Если $y^k = z^r = 0$, то в сумме

$$(y+z)^{k+r-1} = \sum_{i=0}^{k+r-1} C^i_{k+r-1} y^i z^{k+r-1-i},$$

каждое слагаемое равно нулю, потому что либо $i \geqslant k$, либо $k + r - 1 - i \geqslant r$. □

Лемма 4.2. *Если $r, s \in R$ – таковы, что r нильпотентен, s обратим, и $rs = sr$, то $s - r$ обратим.*

Доказательство. $(s-r)s^{-1} = 1 - rs^{-1}$. Элемент $\tilde{r} = rs^{-1}$ нильпотентен, так как r нильпотентен и коммутирует с s^{-1}. С другой строны, $(1-\tilde{r})(1+\tilde{r}+\cdots+\tilde{r}^{k-1}) = 1-\tilde{r}^k = 1$ для достаточно большого k. □

Оператор называется полупростым, если его матрица диагонализуема над алгебраическим замыканием исходного поля. Более подробно полупростота оператора над незамкнутым полем будет обсуждаться в следующих главах.

Поле F называется совершенным, если любой неприводимый многочлен имеет ненулевую производную. Это условие равносильно тому, что многочлен, неприводимый над F, не имеет кратных корней над замыканием F.

Теорема 4.3 (аддитивное разложение Жордана)**.** *Предположим, что F – совершенное поле. Тогда для любой матрицы $A \in \mathrm{M}_n(F)$ существуют единственные матрицы $A^{(s)}, A^{(n)} \in \mathrm{M}_n(F)$ такие, что:*

(1) *$A^{(s)}$ – полупростая, $A^{(n)}$ – нильпотентная;*
(2) *$A = A^{(s)} + A^{(n)}$;*
(3) *$A^{(s)}A^{(n)} = A^{(n)}A^{(s)}$.*

При этом матрицы $A^{(s)}$ и $A^{(n)}$ выражаются, как многочлены без свободного члена от матрицы A.

Доказательство. Мы докажем эту теорему только при условии, что F алгебраически замкнуто. По теореме 2.2 существует обратимая матрица C такая, что

$$C^{-1}AC = \operatorname{diag}(\lambda_1 E, \dots, \lambda_m E) + \operatorname{diag}(N_1, \dots, N_m),$$

где $\lambda_1, \dots, \lambda_m$ – все различные собственные числа матрицы A, а N_i – матрица с некоторым количеством единиц непосредственно над главной диагональю и нулями в остальных местах. Обозначим первое слагаемое (диагональную матрицу) через S, а второе – через N. Ясно, что $\lambda_i E$ коммутирует с N_i, поэтому $SN = NS$. Очевидно также, что S полупроста, а N нильпотентна. Таким образом,

$$A = CSC^{-1} + CNC^{-1}$$

является аддитивным разложением Жордана.

На **втором шаге** докажем, что из условий (1) – (3) следует, что $A^{(s)}$ и $A^{(n)}$ являются многочленами от A. Пусть $\lambda_1, \dots, \lambda_m$ – собственные числа матрицы $A^{(s)}$, $V_1, \dots, V_m$ – соответствующие собственные подпространства, а k – степень нильпотентности матрицы $A^{(n)}$. Так как $A^{(s)}$ диагонализуема над F, то V является прямой суммой собственных подпространств. Если $x \in V_i$, то $(A - \lambda_i E)x = A^{(n)}x$, следовательно,

$$(A - \lambda_i E)^k x = (A - \lambda_i E)^{k-1} A^{(n)} x = A^{(n)}(A - \lambda_i E)^{k-1} x = \dots = (A^{(n)})^k x = 0.$$

При помощи китайской теоремы об остатках найдем многочлен $p \in F[t]$ удовлетворяющий сравнениям $p \equiv \lambda_i \mod (t - \lambda_i)^k$ при всех $i = 1, \dots, m$, и $p \equiv 0 \mod t$ (если одно из λ_i равно нулю, то последнее условие лишнее). Так как $(A - \lambda_i E)^k = 0$ на V_i, то на этом подпространстве $p(A) = \lambda_i E = A^{(s)}$. Следовательно, и на всем пространстве $p(A) = A^{(s)}$.

Третий шаг – доказательство единственности разложения. Достаточно доказать это для матрицы $B = C^{-1}AC = S + N$. Пусть $B = S' + N'$, где S' полупростая, N' нильпотентная, и $S'N' = N'S'$. Из последнего равенства следует, что S' и N' коммутируют с B, а, значит, и с любым многочленом от B, в частности, с N и S. следовательно, все 4 матрицы N, S, N' и S' коммутируют друг с другом.

Так как $S = \operatorname{diag}(\lambda_1 E, \dots, \lambda_m E)$, где все λ_i различны, а $SS' = S'S$, то простое матричное вычисление показывает, что $S' = \operatorname{diag}(S^{(1)}, \dots, S^{(m)})$, где размеры диагональных блоков равны размерам диагональных блоков, на которые разбита матрица S. По лемме 4.1, матрица $S' - S = N - N'$ нильпотентна.

Пусть $J^{(i)} = (C^{(i)})^{-1} S^{(i)} C^{(i)}$ – жорданова форма матрицы $S^{(i)}$. Положим

$$C' = \operatorname{diag}(C^{(1)}, \dots, C^{(m)}) \text{ и } J' = \operatorname{diag}(J^{(1)}, \dots, J^{(m)}).$$

Тогда $(C')^{-1} S' C' = J'$. Так как S' диагонализуема, то этим свойством обладает и матрица J'. Но J' – блочно диагональная матрица с жордановыми блоками по диагонали. Если не все эти блоки имеют размер 1 на 1, то алгебраическая кратность какого-то собственного числа не совпадает с геометрической, что противоречит диагонализуемости. Таким образом, матрица J' диагональна. Заметим, что C' коммутирует с S, потому что S коммутирует с любой блочно диагональной матрицей с такими размерами блоков. Следовательно, матрица $S - J' = (C')^{-1}(S - S')C$ диагональна и нильпотентна. Легко видеть, что она нулевая. Таким образом, $S = S'$, откуда следует единственность. □

Элемент r кольца с 1 называется унипотентным, если $r - 1$ нильпотентен.

Следствие 4.4 (мультипликативное разложение Жордана). *Для любой матрицы $A \in \mathrm{GL}_n(F)$ над совершенным полем F существуют матрицы $S, U \in \mathrm{M}_n(F)$ такие, что $A = SU$, S – полупростая, U – унипотентная, и $SU = US$.*

При этом матрицы S и U выражаются, как многочлены от матрицы A.

Доказательство. Пусть $A = S + N$ – аддитивное разложение Жордана матрицы A. Так как A – обратимая матрица, то и $S = A - N$ обратима. Таким образом, $A = SU$, где $U = E + S^{-1}N$ и есть мультипликативное разложение Жордана.

Осталось доказать, что U – многочлен от A. Пусть $S = p(A)$, где $p \in F[t]$ – многочлен, существование которого утверждается в предыдущей теореме, а x – собственный вектор матрицы A, соответствующий собственному числу $\lambda \in F$. Тогда $Sx = p(A)x = p(\lambda)x$. Так как S обратима, а $x \neq 0$, то $p(\lambda) \neq 0$. Так как только собственные числа являются корнями минимального многочлена φ_A матрицы A, то p взаимно прост с φ_A. Напишем линейное представление НОД: $1 = pf + \varphi_A g$ и подставим туда A. Получим $E = p(A)f(A) + \varphi_A(A)g(A) = Sf(A)$. Таким образом, $S^{-1} = f(A)$, и $U = AS^{-1} = Af(A)$. □

5. Функции от матриц

Пусть $f : \mathbb{C} \to \mathbb{C}$ – функция комплексной переменной, раскладывающаяся в степенной ряд с бесконечным радиусом сходимости:[3]

$$f(z) = \sum_{k=0}^{\infty} \alpha_k(\mu)(z-\mu)^k.$$

Для матрицы $A \in \mathrm{M}_n(\mathbb{C})$ положим по определению,

$$f(A) = \lim_{m\to\infty} \sum_{k=0}^{m} \alpha_k(\mu)(A-\mu E)^k,$$

где предел берется поэлементно (впрочем, можно брать предел по любой норме на векторном пространстве $\mathrm{M}_n(\mathbb{C})$, так как в конечномерном пространстве предел не зависит от выбора нормы). Можно доказать, что этот предел всегда существует и не зависит от точки μ.

Умножение матриц задается многочленами и, поэтому, непрерывно, как функция $\mathbb{C}^{2n^2} \to \mathbb{C}^{n^2}$. Для $A \in \mathrm{M}_n(\mathbb{C})$ и $C \in \mathrm{GL}_n(\mathbb{C})$ имеем:

$$f(C^{-1}AC) = \lim_{m\to\infty} \sum_{k=0}^{m} \alpha_k(C^{-1}AC - \mu E)^k \overset{2.4}{=}$$

$$\lim_{m\to\infty} C^{-1}\Big(\sum_{k=0}^{m} \alpha_k(A-\mu E)^k\Big)C \overset{\text{непр.}}{=} C^{-1}f(A)C.$$

Равенство означает, что либо ряды в обеих частях расходятся, либо эти выражения равны. Предположим, что $C^{-1}AC = \operatorname{diag}(\lambda_1 E + J_{k_1}, \ldots, \lambda_l E + J_{k_l})$ – жорданова форма матрицы A. По лемме 2.5 для вычисления многочлена от блочно диагональной матрицы достаточно вычислить его от каждого диагонального блока. Аналогичное утверждение очевидно верно и с заменой слова "многочлен" на слово "предел" или "ряд".

Пусть $\lambda \in \mathbb{C}$, а $J = J_s$ – жорданов блок размера $s \times s$ с собственным числом 0. Тогда $f(\lambda E + J) = \sum_{k=0}^{s-1} \alpha_k(\lambda) J^k$.

Теорема 5.1. *Пусть $A \in \mathrm{M}_n(\mathbb{C})$ имеет собственные числа $\lambda_1, \ldots, \lambda_\ell \in \mathbb{C}$, а $C = C_{e\to u}$ – матрица перехода от стандартного базиса $\mathbb{C}^n$ к жорданову базису матрицы A. Тогда*

$$f(A) = C \operatorname{diag}\big(f(\lambda_1 E + J_{s_1}), \ldots, f(\lambda_l E + J_{s_\ell})\big)C^{-1},$$

а значение f на жордановом блоке вычисляется по формуле

$$f(\lambda E + J_s) = \sum_{k=0}^{s-1} \alpha_k(\lambda) J_s^k.$$

Вычисление экспоненты от матрицы применяется при решении системы дифференциальных уравнений $x' = Ax$, где $x = (x_1, \ldots, x_n)^\intercal$, x_k – дифференцируемые функции $\mathbb{C} \to \mathbb{C}$, а $A \in \mathrm{M}_n(\mathbb{C})$. Общее решение этой системы имеет вид $x = e^{At}c$, где c пробегает множество столбцов $\mathbb{C}^n$.

[3] Можно рассматривать произвольные функции, аналитические в данном круге, но мне не хочется вдаваться в подробности мат. анализа.

6. Дифференциальные и рекуррентные уравнения

Однородное линейное дифференциальное уравнение с постоянными коэффициентами

$$y^{(n)} + a_{n-1}y^{(n-1)} + \cdots + a_0 y = 0. \tag{29}$$

можно переписать в виде системы линейных уравнений $x' = Ax$, где $x = (y, y', \ldots, y^{(n-1)})^\intercal$, а

$$A = \begin{pmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \ldots & \ldots & 1 \\ -a_0 & -a_1 & -a_2 & \ldots & -a_{n-1} \end{pmatrix} \tag{30}$$

фробениусова клетка. Таким образом, задача о решении уравнения (29) сводится к решению системы линейных дифференциальных уравнений, т. е. к нахождению экспоненты от матрицы.

Однако есть и более прямой путь решения уравнения (29). Запишем его в виде $p(D)(y) = 0$, где D – оператор дифференцирования на множестве дифференцируемых функций $\mathbb{C} \to \mathbb{C}$, а $p = t^n + a_{n-1}t^{n-1} + \cdots + a_0$. Другими словами, нам надо найти ядро дифференциального оператора $p(D)$.

По теореме 3.1 для решения этой задачи достаточно рассмотреть случай $p(t) = (t-\lambda)^k$. При $\lambda = 0$ мы умеем решать такую задачу: решение дифференциального уравнения $y^{(k)} = 0$ – множество многочленов степени $\leqslant k-1$. Общий случай следует из теоремы сдвига.

Теорема 6.1. *Пусть D – оператор дифференцирования, определенный выше, а p – многочлен. Тогда*

$$\operatorname{Ker} p(D - \lambda \cdot \mathrm{id}) = e^{\lambda t} \operatorname{Ker} p(D).$$

Таким образом, общее решение уравнения (29) выглядит следующим образом.

Теорема 6.2. *Пусть D – оператор дифференцирования, а $p(t) = \prod_{i=1}^{m} (t-\lambda_i)^{k_i}$. Тогда*

$$\operatorname{Ker} p(D) = \{\sum_{i=1}^{m} e^{\lambda_i t} f_i(t) \mid f_i \in \mathbb{C}[t],\ \deg f_i < k_i\}.$$

Пусть теперь F^∞ обозначает векторное пространство бесконечных последовательностей $x : \mathbb{N}_0 \to F$, где F алгебраически замкнутое поле. Рекуррентное уравнение

$$x_{m+n} + a_{n-1}x_{m+n-1} + \cdots + a_0 x_m = 0 \text{ при любом } m \in \mathbb{N}_0 \tag{31}$$

сводится к нахождению ядра оператора $p(S)$, где S – оператор сдвига номеров, т. е. $S(x)_k = x_{k+1}$ при всех $k \in \mathbb{N}_0$. Правда для рекуррентных уравнений теорему сдвига написать не удается. Поэтому ядро оператора $(S - \lambda \cdot \mathrm{id})^k$ приходится искать индукцией по k.

Альтернативно, уравнение (31) можно переписать в виде $x^{(m+1)} = Ax^{(m)}$, где A – фробениусова клетка (30). В этих обозначениях $x^{(m)} = A^m x^{(0)}$, где $x^{(0)}$ произвольный столбец. Для возведения матрицы в степень можно воспользоваться жордановой формой.

Так или иначе, общее решение уравнения (31) выглядит следующим образом.

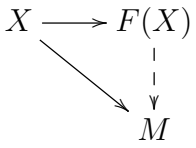
Теорема 6.3. *Пусть S – оператор сдвига, а $p(t) = \prod_{i=1}^{m} (t-\lambda_i)^{k_i}$. Тогда*

$$\operatorname{Ker} p(S) = \{\sum_{i=1}^{m} \lambda_i^n f_i(n) \mid f_i \in F[t],\ \deg f_i < k_i\}.$$

7. Свободные модули

Пусть R – коммутативное кольцо. Напомним, что модуль называется свободным, если он имеет базис. Свободный модуль $F(X)$ с базисом X обладает следующим универсальным свойством.

Предложение 7.1. *Для любого R-модуля M и функции $X \to M$ существует единственный гомоморфизм модулей $F(X) \to M$ такой, что диаграмма*



коммутативна.

Доказательство. Из условия коммутативности диаграммы мы знаем образы базисных элементов. Так как любой элемент $F(X)$ является линейной комбинацией базисных, а гомоморфизм модулей переводит линейные комбинации в линейные комбинации, то образы всех элементов $F(X)$ определены однозначно. Проверка того, что таким образом мы действительно получим гомоморфизм модулей тривиальна. □

Следующее утверждение – конструкция свободного модуля.

Предложение 7.2. *Пусть $F(X)$ – множество финитных функций $X \to R$. Они образуют модуль над R относительно поточечных операций. Зададим функцию $i : X \to F(X)$ формулой $i(x) = \mathbb{1}_x$ – характеристическая функция множества $\{x\}$, т. е. $\mathbb{1}_x(y) = \begin{cases} 1, & y = x \\ 0, & y \neq x \end{cases}$. Тогда $F(X)$ вместе с отображением i – свободный модуль с базисом X.*

Если X конечно, то $F(X) \cong R^{|X|} \cong \underbrace{R \oplus \cdots \oplus R}_{|X| \text{ раз}}$.

Доказательство. Фактически множество столбцов высоты $n = |X|$ – это и есть множество функций $X \to R$, или множество кортежей длины n, которое совпадает с прямой суммой. Легко проверить, что после отождествления окажется, что операции определены одинаково. □

Определим аналог размерности векторных пространств для конечнопорожденных свободных модулей. Если u и v – конечные базисы свободного модуля, то, также как и для векторных пространств, существуют взаимно обратные матрицы перехода $C_{u \to v}$ и $C_{v \to u}$. Неквадратная матрица над коммутативным кольцом не может быть двусторонне обратима. Действительно, если бы она была двусторонне обратима, то ее образ над любым факторкольцом обладал бы тем же свойством. С другой стороны, в любом коммутативном кольце существует максимальный идеал, фактор по которому – поле, а для поля результат известен из линейной алгебры. Поэтому любой базис свободного модуля имеет одинаковое число элементов, которое называется рангом этого модуля.[4]

8. Подмодули свободного модуля

Следующие два параграфа лежат на стыке теории коммутативных колец и линейной алгебры. Мы докажем теорему классификации конечнопорожденных модулей над кольцами главных

[4] Для некоммутативных колец это утверждение неверно, т. е. существует некоммутативное кольцо Λ для которого модули Λ^n и Λ^m изоморфны при $m \neq n$. Простейший пример такого кольца – кольцо эндоморфизмов бесконечномерного векторного пространства. Кольца, для которых ранг свободного модуля определен однозначно, называются ИБЧ или IBN-кольцами (Invariant Basis Number). Нетрудно доказать, что из не-ИБЧ-кольца не существует ни одного гомоморфизма ни в одно ИБЧ-кольцо, т. е. в каком-то смысле класс не-ИБЧ-колец – это черная дыра: попав туда, выбраться уже невозможно.

идеалов, из которой затем выведем теорему о Жордановой форме оператора. В качестве бесплатного приложения мы получим теорему о строении конечнопорожденных абелевых групп. Пусть R – коммутативное кольцо, а M – конечнопорожденный R-модуль, т. е. в M существует конечная система образующих $X = \{x_1, \dots, x_n\}$. По универсальному свойству свободного модуля существует единственный гомоморфизм свободного модуля $F(X)$ в M, тождественный на X. Так как X – система образующих, то этот гомоморфизм сюръективен. По теореме о гомоморфизме $M \cong F(X)/N$ для некоторого подмодуля N в $F(X)$. Учитывая, что $F(X) \cong R^{|X|}$, для описания произвольного конечнопорожденного R-модуля достаточно описать подмодули в R^n для произвольного натурального n. Мы знаем, что подмодуль конечнопорожденного модуля над нетеровым кольцом конечнопорожден. Для описания подмодуля $N \leqslant R^n$ возьмем матрицу, состоящую из порождающих N и посмотрим, к какому виду ее можно привести заменой базиса R^n и заменой системы образующих N.

Теорема 8.1 (Нормальная форма Смита)**.** *Для любой матрицы $A \in \mathrm{M}_{n,k}(R)$ над областью главных идеалов R существуют матрицы $B \in \mathrm{GL}_n(R)$ и $C \in \mathrm{GL}_k(R)$ такие, что все недиагональные элементы матрицы BAC равны нулю, а каждый диагональный элемент делится на предыдущий.*

Специфика области главных идеалов, необходимая для доказательства этой теоремы, выражена в следующей лемме. Фактически, эта лемма является частным случаем теоремы при $k = 1$.

Лемма 8.2. *Пусть R – область главных идеалов. Для любого столбца $x \in R^n$ существует матрица $B \in \mathrm{GL}_n(R)$ такая, что $Bx = e_{*1}\alpha$. При этом α является наибольшим общим делителем элементов столбца x.*

Доказательство. Проведем доказательство индукцией по n. Для $n = 1$ доказывать нечего. При $n > 1$ положим $d = \gcd(x_{n-1}, x_n)$, $a = \frac{x_{n-1}}{d}$ и $b = \frac{x_n}{d}$. По теореме о линейном представлении НОД существуют $r, s \in R$ такие, что $x_{n-1}r + x_n s = d$, откуда $ar + bs = 1$. Тогда

$$\begin{pmatrix} E & 0 & 0 \\ 0 & r & s \\ 0 & -b & a \end{pmatrix} \begin{pmatrix} x' \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} x' \\ d \\ 0 \end{pmatrix},$$

где $x' = (x_1, \dots, x_{n-2})^\intercal$. Заметим, что определитель первого сомножителя равен 1, следовательно, эта матрица обратима. По индукционному предположению существует обратимая матрица A' такая, что $A' \left(\begin{smallmatrix} x' \\ d \end{smallmatrix}\right) = \left(\begin{smallmatrix} \alpha \\ 0 \end{smallmatrix}\right)$, где $\alpha = \gcd(x_1, \dots, x_{n-2}, d) = \gcd(x_1, \dots, x_n)$. □

Упражнение 8.3. Докажите, что если R – нетерова область целостности, то свойство из предыдущей леммы равносильно тому, что любой идеал в R является главным.

Следующий шаг – изучение матрицы с двумя столбцами.

Лемма 8.4. *Пусть R – область главных идеалов, $A \in \mathrm{M}_{n2}(R)$ – матрица с первым столбцом $e_{*1}\alpha$. Если не все элементы матрицы A делятся на α, то существуют матрицы $B \in \mathrm{GL}_n(R)$ и $C \in \mathrm{GL}_2(R)$ такие, что первый столбец матрицы BAC равен $e_{*1}\beta$, причем $\alpha R \subsetneqq \beta R$.*

Доказательство. По условию a_{i2} не делится на α для некоторого $i = 1, \dots, n$. Если a_{12} делится на α, то прибавим i-ю строку к первой. Таким образом, можно считать, что изначально a_{12} не делится на α.

По предыдущей лемме найдем матрицу $C \in \mathrm{GL}_2(R)$ такую, что $C\left(\begin{smallmatrix} \alpha \\ a_{12} \end{smallmatrix}\right) = \left(\begin{smallmatrix} \delta \\ 0 \end{smallmatrix}\right)$, где $\delta = \gcd(\alpha, a_{12})$. Пусть $D = AC^\intercal$. Тогда $d_{11} = \delta$. Снова по предыдущей лемме существует матрица $B \in \mathrm{M}_n(R)$ такая, что первый столбец матрицы $BAC^\intercal$ равен $e_{*1}\beta$, где $\beta = \gcd(\delta, d_{21}, \dots, d_{n1})$. Осталось заметить, что $\alpha R \subsetneqq \delta R \subseteq \beta R$. □

Доказательство теоремы о нормальной форме Смита. Пусть S – множество тех $\gamma \in R$, для которых существют матрицы $B \in \mathrm{GL}_n(R)$ и $C \in \mathrm{GL}_k(R)$ такие, что первый столбец матрицы BAC равен $e_{*1}\gamma$. По лемме 8.2 это множество не пусто. Возьмем максимальный из идеалов αR

по всем $\alpha \in S$ (так как R нетерово, то в любом множестве идеалов существует максимальный). Пусть $D = BAC$ такова, что $d_{*1} = e_{*1}\alpha$. По лемме 8.4 все d_{ij} делятся на α. Вычитая первый столбец из всех остальных, получим матрицу $BACT = \operatorname{diag}(\alpha, \alpha A')$. Доказательство заканчивает индукция по размеру матрицы A. □

Следствие 8.5 (классификации подмодулей конечнопорожденного свободного модуля)**.** *Пусть R – кольцо главных идеалов, а N – подмодуль свободного модуля R^n. Существует базис $u = (u_1, \dots, u_n)$ модуля R^n и элементы $\alpha_1, \dots, \alpha_n \in R$ такие, что*

(1) *элементы $u_1\alpha_1, \dots, u_n\alpha_n$ лежат в N и порождают его;*
(2) *α_{i+1} делится на α_i при всех $i = 1, \dots, n-1$ (мы считаем, что 0 делится на 0).*

Доказательство. Пусть столбцы $x_1, \dots, x_k$ порождают подмодуль $N \subseteq R^n$. Добавив нулевые элементы, можно считать, что $k \geqslant n$. Рассмотрим матрицу $A = (x_1, \dots, x_k)$ и возьмем B, C из теоремы 8.1. Пусть u_i – i-й столбец матрицы B^{-1}. Тогда $u = (u_1, \dots, u_n)$ – базис модуля R^n, а $AC = B^{-1}(BAC) = (u_1\alpha_1, \dots, u_n\alpha_n, 0, \dots, 0)$. Ясно, что столбцы матрицы AC порождают N, откуда следует результат. □

Пусть m – наибольшее целое для которого $\alpha_m \neq 0$. Тогда $(u_1\alpha_1, \dots, u_m\alpha_m)$ линейно независимы, так как уже $(u_1, \dots, u_m)$ линейно независимы, следовательно, являются базисом N.

Следствие 8.6. *Любой подмодуль конечнопорожденного свободного модуля над областью главных идеалов является свободным, причем ранг подмодуля не превосходит ранга модуля.*

9. Конечнопорожденные модули над кольцами главных идеалов

Модуль M над кольцом R называется циклическим, если он порожден одним элементом. Любой циклический модуль изоморфен R/I для некоторого идеала I кольца R. Действительно, если $M = \langle m \rangle$, то отображение $\pi : R \to M$, заданное формулой $\pi(r) = mr$ является эпиморфизмом модулей, следовательно, $M \cong R/\operatorname{Ker}\pi$. Ненулевой модуль называется неприводимым, если у него нет нетривиальных подмодулей. Модуль называется неразложимым, если он не раскладывается в прямую сумму ненулевых подмодулей.

Лемма 9.1. *Циклический модуль R/I неприводим тогда и только тогда, когда I – максимальный идеал. Циклический модуль R/I неразложим тогда и только тогда, когда I не раскладывается в произведение собственных взаимно простых идеалов.*

Доказательство. Первое утверждение следует из третьей теоремы о гомоморфизме для модулей.

Если $I = I_1I_2$, а $I_1 + I_2 = R$, то по китайской теореме об остатках $R/I \cong R/I_1 \oplus R/I_2$ как кольца. Ясно, что это разложение совместимо со структурой R-модулей.

Обратно, пусть $R/I \cong M_1 \oplus M_2$. Каноническая проекция $R \to R/I$ в композицией с проекцией R/I на M_k дает эпиморфизм $R \to M_k$, следовательно, $M_k \cong R/I_k$ (здесь $k = 1, 2$). При этом $I_k \supseteq I$, а $\varphi(r + I) = (r + I_1, r + I_2)$ – изоморфизм $R/I \cong R/I_1 \oplus R/I_2$. Тогда $\operatorname{Ker}\varphi = \{r + I \mid r \in I_1 \cap I_2\} = \{0_{R/I}\}$, откуда $I_1 \cap I_2 = I$. Так как φ сюръективно, то существуют $a, b \in R$ такие, что $\varphi(a + I) = (I_1, 1 + I_2)$ и $\varphi(b + I) = (1 + I_1, I_2)$. Другими словами, $a, b - 1 \in I_1$, а $a - 1, b \in I_2$. Тогда $c = a + b - 1 \in I_1 \cap I_2 = I$, откуда $(a - c) + b = 1$, причем $a - c \in I_1$, а $b \in I_2$. Так как идеал $I_1 + I_2$ содержит 1, то он равен R, т. е. I_1 и I_2 взаимно просты. Наконец, по лемме 8.2 главы 4 $I_1I_2 = I_1 \cap I_2 = I$. □

Следствие 9.2. *Пусть R – область главных идеалов. Циклический R-модуль R/qR неразложим тогда и только тогда, когда q – степень неприводимого элемента или $q = 0$.*

Теорема 9.3 (Классификация конечнопорожденных модулей)**.** *Любой конечнопорожденный модуль над областью главных идеалов изоморфен прямой сумме неразложимых циклических модулей. При этом набор таких модулей определен однозначно.*

Доказательство. В этом параграфе мы докажем только существование такого разложения и выведем несколько следствий. Единственность будет доказана в следующем разделе.

Пусть M – конечнопорожденный модуль над областью главных идеалов R. Как уже говорилось, существует эпиморфизм $R^k \to M$. Обозначим его ядро через N. По теореме 8.5 существует базис $u_1, \dots, u_k$ модуля R^k и элементы $\alpha_1, \dots, \alpha_k \in R$ такие, что $N = \langle u_1\alpha_1, \dots, u_n\alpha_n \rangle$. Зададим гомоморфизм $\rho : R^k \to \bigoplus_{i=1}^{k} R/\alpha_i R$ формулой

$$\rho(u_1\beta_1 + \dots + u_k\beta_k) = (\beta_1 \mod \alpha_1, \dots, \beta_k \mod \alpha_k)$$

(так как u – базис, то эта формула однозначно задает отображение). Легко видеть, что ρ – эпиморфизм, а его ядро равно N. По теореме о единственности эпиморфизма с данным ядром $M \cong \bigoplus_{i=1}^{k} R/\alpha_i R$, т. е. M является прямой суммой циклических модулей. С другой стороны, по лемме 9.1 любой циклический модуль изоморфен прямой сумме неразложимых. □

Следствие 9.4 (Классификация конечнопорожденных абелевых групп)**.** *Любая конечнопорожденная абелева группа изоморфна прямой сумме циклических групп бесконечного порядка и порядка p^n, где p – простое число. При этом такое разложение единственно с точностью до перестановки прямых слагаемых.*

В частности, если порядок абелевой группы делится на простое число p, то среди прямых слагаемых обязательно должна быть циклическая группа порядка p^n, следовательно, в ней есть элемент порядка p – утверждение, которое будет использовано при доказательстве первой теоремы Силова.

10. Фробениусова и жорданова форма

Из теоремы о строении конечнопорожденных модулей, примененной к кольцу многочленов, получаем доказательство существования нормальных форм матрицы оператора. Клеткой Фробениуса называется матрица $A \in \mathrm{M}_n(F)$ с элементами $a_{i+1,i} = 1$ при $i = 1, \dots, n-1$, произвольными a_{in}, $i = 1, \dots, n$, и остальными нулями.

Теорема 10.1 (нормальная форма Фробениуса)**.** *Для любого линейного оператора $L : V \to V$ существует базис и пространства V такой, что матрица L_u клеточно диагональная с клетками Фробениуса по диагонали.*

В этой теореме единственности нет, потому что клетки Фробениуса соответствуют циклическим модулям, которые не обязательно неразложимы. Заметим, что характеристический многочлен фробениусовой клетки A равен $\pm(t^n - a_{nn}t^{n-1} - \dots - a_{2n}t - a_{1n})$.

Упражнение 10.2. Докажите, что фробениусова клетка не приводится над F к блочно диагональному виду тогда и только тогда, когда ее характеристический многочлен является степенью неприводимого над F многочлена.

Доказательство. Используя гомоморфизм F-алгебр $\varepsilon_L : F[t] \to \mathrm{End}(V)$, посылающий t в L, рассмотрим V как $F[t]$-модуль. Другими словами, для $p \in F[t]$ положим $pv = p(L)(v)$. Так как $F[t]$ – кольцо главных идеалов, то по теореме 9.3 V изоморфно прямой сумме циклических модулей (допуская вольность речи будем считать, что V равно прямой сумме циклических модулей). Заметим, что $F[t]$ не может встречаться в прямой сумме, потому что V конечномерно над F, а $F[t]$ – нет.

Каждый из циклических модулей является L-инвариантным подпространством, поэтому в F-базисе V, состоящем из F-базисов циклических модулей, матрица оператора L будет клеточно диагональной. В каждом циклическом модуле $F[t]/pF[t]$ выберем базис из смежных классов $\overline{t^k} = t^k + pF[t]$ $(k = 0, \dots, \deg p - 1)$. Тогда $L(\overline{t^k}) = t \cdot \overline{t^k} = \overline{t^{k+1}}$. Если $k < \deg p - 1$, то оператор L отображает базисный вектор в следующий. Следовательно, в соответствующих столбцах матрицы

оператора L появляются единицы под главной диагональю и остальные нули. Таким образом, в диагональных блоках стоят клетки Фробениуса. □

Доказательство существования жордановой формы. Пусть F – алгебраически замкнутое поле, V – векторное пространство над F, а $L : V \to V$ – линейное отображение. Также, как в доказательстве предыдущей теоремы, рассмотрим V как $F[t]$-модуль. Так как $F[t]$ – область главных идеалов, то по теореме 9.3 V изоморфно прямой сумме неразложимых циклических модулей. Заметим, что $F[t]$ не может встречаться в прямой сумме, потому что V конечномерно над F, а $F[t]$ – нет.

Так как F замкнуто, то неприводимыми элементами кольца $F[t]$ являются только многочлены $t - \lambda$. По следствию 9.2 модуль V можно отождествить с прямой суммой модулей вида $M = F[t]/(t-\lambda)^k F[t]$. Элементы этого модуля взаимно однозначно соответствуют многочленам степени, меньшей k.

Поэтому любой элемент M может быть единственным образом представлен в виде линейной комбинации $\sum_{i=0}^{k-1} \alpha_i \overline{(t-\lambda)^i}$, где $\alpha_i \in F$, а черта обозначает каноническую проекцию $F[t] \to M$. Таким образом, элементы $v_i := \overline{(t-\lambda)^{k-i}}$, $i = 1, \dots, k$, образуют базис M, как векторного пространства над F (мы меняем естественный порядок базисных векторов, чтобы получилась в точности жорданова клетка, а не транспонированная к ней). Ясно, что $(t-\lambda)v_i = v_{i-1}$ при $i > 1$, а $(t-\lambda)v_1 = 0$. По определению действия t на M получаем:

$$L(v_i) = v_{i-1} + \lambda v_i \text{ при } i > 1, \text{ и } L(v_1) = \lambda v_1.$$

Другими словами, матрица сужения оператора L на M в выбранном базисе – это жорданова клетка. Таким образом, пространство V разбилось в прямую сумму подпространств, в каждом из которых матрица сужения L в подходящем базисе является жордановой клеткой. Объединение базисов прямых слагаемых является базисом пространства. Ясно, что в объединении подходящих базисов матрица оператора L будет клеточно-диагональной с жордановыми блоками по диагонали.

Утверждение о единственности следует из единственности разложения конечнопорожденного модуля в прямую сумму неразложимых циклических модулей. □

11. Единственность разложения

Пусть R – область целостности, а M – R-модуль. Множество элементов из M, которые при умножении на какой-нибудь ненулевой элемент кольца обращаются в 0, называется подмодулем кручения. Это действительно подмодуль, потому что если $m_1 r_1 = 0$ и $m_2 r_2 = 0$, то $(m_1+m_2)r_1 r_2 = 0$ (здесь $m_1, m_2 \in M$, а $r_1, r_2 \in R \setminus \{0\}$). Говорят, что M – модуль без кручения, если его подмодуль кручения нулевой.

Фактормодуль по подмодулю кручения является модулем без кручения. Действительно, Обозначим через $T = T(M)$ подмодуль кручения. Тогда условие $x + T \in T(M/T)$ равносильно включению $xr \in T$ для некоторого ненулевого $r \in R$. Но тогда по определению подмодуля кручения $xrr' = 0$ для какого-то $r' \in R \setminus \{0\}$, при этом $rr' \neq 0$, так как R – область целостности. Следовательно, $x \in T$, т. е. $x + T = 0_{M/T}$.

Пусть теперь R – область главных идеалов, а $p \in R$ – неприводимый элемент. Тогда R/pR – поле. Для каждого натурального числа k рассмотрим фактормодуль $p^{k-1}M/p^k M$. Так как умножение на p аннулирует этот модуль, то можно естественным образом определить на этом модуле структуру R/pR-векторного пространства: $(r + pR)(p^{k-1}m + p^k M) = p^{k-1}rm + p^k M$. Оказывается числа $\ell_{p,k} = \ell_{p,k}(M) := \dim_{R/pR} p^{k-1}M/p^k M$ по всем неприводимым $p \in R$ и всем $k \in \mathbb{N}$ полностью определяют строение подмодуля кручения $T(M)$.

Лемма 11.1. *Пусть q, p – неассоциированные неприводимые элементы в R, а $M = R/p^n R$. Тогда $\ell_{q,k} = 0$, $\ell_{p,k} = 1$ при $n \geqslant k$, и $\ell_{p,k} = 0$ в противном случае.*

$\ell_{p,k}(M_1 \oplus M_2) = \ell_{p,k}(M_1) + \ell_{p,k}(M_2)$.

Если M – прямая сумма неразложимых циклических модулей не изоморфных R, то $\ell_{p,k}(M)$ равно количеству прямых слагаемых в разложении M, изоморфных R/p^hR с $h \geqslant k$.

Доказательство. Так как q обратим в кольце R/p^nR, то умножение на q является автоморфизмом модуля M, откуда $q^kM = q^{k-1}M$. Фактормодуль циклического модуля циклический, поэтому если он ненулевой, то он порожден 1 элементом. При $n \geqslant k$ модуль $p^{k-1}(R/p^nR)$ порожден элементом p^{k-1}, отсюда первое утверждение.

Ясно, что $p^k(M_1 \oplus M_2) = p^kM_1 \oplus p^kM_2$. Следовательно,

$$p^{k-1}(M_1 \oplus M_2)/p^k(M_1 \oplus M_2) \cong p^{k-1}M_1/p^kM_1 \oplus p^{k-1}M_2/p^kM_2,$$

откуда вытекает второе утверждение леммы. Третье утверждение сразу вытекает из двух первых. □

Замечание 11.2. В предыдущей лемме можно заменить идеалы вида p^nR на $\mathfrak{m}^n$, где $\mathfrak{m}$ – максимальный идеал. Тогда можно не накладывать никаких ограничений на коммутативное кольцо R.

Доказательство единственности разложения из теоремы 9.3. Пусть R – область главных идеалов, а

$$M \cong R^k \oplus M_1 \oplus \cdots \oplus M_s, \text{ где } M_i = \bigoplus_{j=1}^{t_i} R/p_i^{h_{ij}}R \text{ при всех } i = 1, \ldots, s,$$

а $p_1, \ldots, p_s$ – попарно неасоциированные неприводимые элементы R. Мы хотим доказать, что набор $p_1, \ldots, p_s$, а также числа k, s, t_i и h_{ij} зависят только от самого модуля M, а не от разложения в прямую сумму.

Ясно, что $T(M) = M_1 \oplus \cdots \oplus M_s$ является подмодулем кручения в M, а $R^k \cong M/T(M)$. Так как ранг свободного модуля определен однозначно, то k определено однозначно. Числа $\ell_{q,h} = \ell_{q,h}\big(T(M)\big)$ также не зависят от разложения.

Количество прямых слагаемых, изоморфных R/p^hR, в разложении модуля M в прямую сумму неприводимых циклических, равно $\ell_{p,h} - \ell_{p,h+1}$. Действительно, по предыдущей лемме $\ell_{p,h+1}$ равно количеству слагаемых R/p^nR, где $n \geqslant h+1$, а $\ell_{p,h}$ – количеству слагаемых R/p^nR, где $n \geqslant h$. Таким образом, количество слагаемых изоморфных R/p^hR также определяется модулем M однозначно, что завершает доказательство. □

Глава 7

Билинейные и квадратичные формы

Везде в этой главе мы считаем, что характеристика основного поля не равна 2. Изучение квадратичных форм в характеристике 2 – это существенно более сложная задача, даже на уровне определений.

1. Формы и их матрицы

Пусть V – векторное пространство над полем F характеристики не 2.

Напомним, что функция $B : V \times V \to F$ называется билинейной формой, если для любых $x, y, z \in V$ и $\alpha, \beta \in F$ имеют место равенства:

$$B(x\alpha + y\beta, z) = B(x, z)\alpha + B(y, z)\beta \text{ и } B(z, x\alpha + y\beta) = B(z, x)\alpha + B(z, y)\beta.$$

Билинейная форма B называется симметричной (антисимметричной), если $B(x, y) = B(y, x)$ (соотв. $B(x, y) = -B(y, x)$) для любых $x, y \in V$.

Обозначим через $\mathrm{Bil}(V)$ – множество всех билинейных форм на V, а через $\mathrm{Bil}^{(s)}(V)$ и $\mathrm{Bil}^{(a)}(V)$ – множества симметричных и антисимметричных билинейных форм. Ясно, что все 3 множества являются подпространствами пространства всех функций $V \times V \to F$ (с поточечными операциями).

Лемма 1.1. $\mathrm{Bil}(V) = \mathrm{Bil}^{(s)}(V) \oplus \mathrm{Bil}^{(a)}(V)$.

Доказательство. Положим $B^{(s)}(x, y) = \frac{1}{2}\big(B(x, y) + B(y, x)\big)$ и $B^{(a)}(x, y) = \frac{1}{2}\big(B(x, y) - B(y, x)\big)$. Ясно, что $B^{(s)}$ является симметричной билинейной формой, $B^{(a)}$ антисимметричной, а $B = B^{(s)} + B^{(a)}$. Тот факт, что $\mathrm{Bil}^{(s)}(V) \cap \mathrm{Bil}^{(a)}(V) = \{0\}$ очевиден (используя условие, что $2 \neq 0$ в поле F). □

Форма $B^{(s)}$ из последнего доказательства называется симметризацией формы B.

Определение 1.2. Функция $Q : V \to F$ называется квадратичной формой, если существует билинейная форма $B : V \times V \to F$ такая, что $Q(x) = B(x, x)$.

Теорема 1.3 (поляризация квадратичной формы)**.** *Пусть Q – квадратичная форма на V, соответствующая билинейной форме B. Положим*

$$B^{(s)}(x, y) = \frac{1}{2}\big(Q(x + y) - Q(x) - Q(y)\big).$$

Тогда $B^{(s)}$ является симметризацией формы B, и $Q(x) = B^{(s)}(x, x)$.

Таким образом, мы имеем биекцию между множеством квадратичных и симметричных билинейных форм, которая на самом деле является изоморфизмом векторных пространств.

В некоторых ситуациях приходится рассматривать чуть более общий случай. Пусть $\bar{\cdot} : F \to F$ – инволюция, т. е. автоморфизм поля порядка 2. Иными словами, для любых $\alpha, \beta \in F$ выполнены равенства

(1) $\overline{\alpha + \beta} = \overline{\alpha} + \overline{\beta}$;
(2) $\overline{\alpha\beta} = \overline{\alpha} \cdot \overline{\beta}$ и
(3) $\overline{\overline{\alpha}} = \alpha$.

Определение 1.4. Функция $B : V \times V \to F$ называется полуторалинейной формой, если для любых $x, y, z \in V$ и $\alpha, \beta \in F$ имеют место равенства:

$$B(x\alpha + y\beta, z) = B(x, z)\overline{\alpha} + B(y, z)\overline{\beta} \text{ и } B(z, x\alpha + y\beta) = B(z, x)\alpha + B(z, y)\beta.$$

Полуторалинейная форма B называется эрмитовой, если $B(x, y) = \overline{B(y, x)}$ для любых $x, y \in V$.

Симметричная билинейная форма является частным случаем эрмитовой формы, если инволюция тривиальна. Поэтому в дальнейшем, везде где можно, мы будем рассматривать эрмитовы формы вместо симметричных билинейных.

Для матрицы $A \in \mathrm{M}_{m,n}(F)$ обозначим $A^* = \overline{A}^\intercal$. Другими словами, $a^*_{ij} = \overline{a_{ji}}$.

Предложение 1.5. *Пусть B полуторалинейная форма на V, а $v = (v_1, \ldots, v_n)$ – базис V. Тогда существует матрица $A \in \mathrm{M}_n(F)$ такая, что $B(x, y) = x_v^* A y_v$ для любых $x, y \in V$. При этом $a_{ij} = B(v_i, v_j)$.*

Доказательство. Пусть $x_v = (x_1, \ldots, x_n)^\intercal$ и $y_v = (y_1, \ldots, y_n)^\intercal$. Тогда

$$B(x, y) = B(\sum_{i=1}^{n} v_i x_i, \sum_{j=1}^{n} v_j y_j) = \sum_{i,j=1}^{n} \overline{x}_i B(v_i, v_j) y_j = x_v^* A y_v.$$

□

Определение 1.6. Матрица A, построенная в предыдущей лемме, называется матрицей формы B в базисе v и обозначается через B_v. Матрицей квадратичной формы называется матрица ассоциированной с ней симметричной билинейной формы.

Определение 1.7. Полуторалинейная форма называется вырожденной, если существует $y \in V \setminus \{0\}$ такой, что $B(y, x) = 0$ при всех $x \in V$.

Если форма вырождена, то $y_v^* B_v x_v = 0$ для любого столбца $x_v \in F^n$, откуда система линейных уравнений $z B_v = 0$ имеет ненулевое решение. Таким образом, вырожденность формы равносильна вырожденности матрицы этой формы (в любом базисе). Следовательно, вырожденность "слева" эквивалентна вырожденности "справа":

$$\exists\, y \in V \setminus \{0\}\ \forall x \in V :\ B(x, y) = 0.$$

Ясно, что матрица квадратичной формы симметрична, т. е. $B_v^\intercal = B_v$, а матрица эрмитовой формы эрмитова, т. е. $B_v^* = B_v$ (другое название такой матрицы – самосопряженная).

Замечание 1.8. Пусть F – поле с инволюцией, а $K = \{\alpha \in F \mid \overline{\alpha} = \alpha\}$. Нетрудно проверить, что K – подполе в F. Диагональные элементы и определитель любой эрмитовой матрицы A лежит в K. Действительно, $A = A^* \implies \det A = \det A^* = \det \overline{A} = \overline{\det A}$, а утверждение про диагональные элементы очевидно.

Предложение 1.9. *Пусть B полуторалинейная форма на V, а u и v – базисы пространства V. Тогда $B_v = C^*_{u \to v} B_u C_{u \to v}$.*

2. Диагонализация эрмитовой формы

В этом параграфе мы начинаем классификацию эрмитовых (квадратичных) форм. Сначала для этого надо определить, какие формы мы считаем одинаковыми. Вместо квадратичной формы можно всегда рассматривать соответствующую симметричную билинейную форму, являющуюся эрмитовой формой с тривиальной инволюцией поля. Поэтому в дальнейшем мы в основном формулируем все для эрмитовых форм.

Определение 2.1. Если B – эрмитова форма на векторном пространстве V, то пара (V, B) называется эрмитовым пространством. Эрмитово пространство называется невырожденным, если форма B невырождена.

Изометрией эрмитовых пространств $(V, B) \to (V', B')$ называется биективное линейное отображение $L : V \to V'$, обладающее свойством $B'\big(L(x), L(y)\big) = B\big(x, y\big)$ для всех $x, y \in V$.

Предложение 2.2. *Пусть B эрмитова форма на V. Положим*

$$V^\perp = \{x \in V \mid B(x, y) = 0\, \forall y \in V\}.$$

Если $V = V^\perp \oplus U$ (заметим, что такое U всегда существует), то сужение формы B на $U \times U$ невырождено. При этом, если $V = V^\perp \oplus U'$, то U и U' изометричны.

Доказательство. Пусть $x \in U$ такой, что $B(x, y) = 0$ для любого $y \in U$. Любой вектор $z \in V$ представляется в виде суммы $z = y + w$ для некоторых $y \in U$ и $w \in V^\perp$. Тогда $B(x, z) = B(x, y) + B(x, w) = 0$. Таким образом, $x \in U \cap V^\perp = \{0\}$, следовательно, сужение B на $U \times U$ невырождено.

Пусть $V^\perp \oplus U = V^\perp \oplus U' = V$. Тогда для любого $x' \in U'$ существует единственный $x \in U$ такой, что $x' \in x + V^\perp$ (такой x называется проекцией x' на U параллельно $V^\perp$). Нетрудно проверить, что такое проектирование (т. е. сопоставление $x' \mapsto x$) является линейным отображением, а обратным к нему является проектирование U на U' параллельно $V^\perp$. Так что это изоморфизм векторных пространств и

$$B(x', y') = B(x + v, y + w) = B(x, y) + B(v, y) + B(x, w) + B(v, w) = B(x, y),$$

откуда следует, что этот изоморфизм – изометрия. □

Задача теории эрмитовых форм – классификация всех конечномерных эрмитовых пространств над данным полем с точностью до изометрии. Ясно, что пространства разной размерности не могут быть изометричными. Очевидно также, что изометрия сохраняет подпространство $V^\perp$ из предложения 2.2. Таким образом, достаточно классифицировать невырожденные эрмитовы пространства любой фиксированной размерности.

Лемма 2.3. *Эрмитовы пространства (V, B) и (V', B') изометричны тогда и только тогда, когда существуют такие базисы v и v' пространств V и V' соответственно, что $B_v = B'_{v'}$.*

Доказательство. Пусть $L : V \to V'$ – изометрия эрмитовых пространств, а $v = (v_1, \dots, v_n)$ – базис пространства V. Тогда $v' = \big(L(v_1), \dots, L(v_n)\big)$ является базисом V'. По определению изометрии $B'\big(L(x), L(y)\big) = B\big(x, y\big)$ для любых $x, y \in V$, откуда $B_v = B'_{v'}$.

Обратно, если $B_v = B'_{v'}$, то отображение $L(x) = v' x_v$ является изометрией эрмитовых пространств. Действительно, $L(x)_{v'} = x_v$, поэтому

$$B'\big(L(x), L(y)\big) = L(x)^*_{v'} B'_{v'} L(y)_{v'} = x_v^* B_v y_v.$$

□

Вектор $x \in V$ называется изотропным (относительно эрмитовой формы B), если $B(x, x) = 0$.

Лемма 2.4. *Для любой ненулевой эрмитовой формы существует неизотропный вектор.*

Доказательство. Так как форма ненулевая, существует пара векторов $y, z \in V$ таких, что $B(y, z) \neq 0$. Если $B(y, y)$ или $B(z, z)$ не равно нулю, то все доказано. Иначе

$$B(y + \lambda z, y + \lambda z) = B(y, y) + B(z, z)\lambda\overline{\lambda} + B(y, \lambda z) + B(\lambda z, y) = B(y, z)\lambda + \overline{\lambda} \cdot \overline{B(y, z)}$$

При $\lambda = \overline{B(y, z)}$ получим $B(y + \lambda z, y + \lambda z) = 2B(y, z)\overline{B(y, z)} \neq 0$. □

Теорема 2.5 (диагонализация матрицы эрмитовой формы)**.** *Для любой эрмитовой формы существует базис, в котором ее матрица диагональна.*

Доказательство. Пусть B – эрмитова форма на V. Мы должны доказать, что существует базис v пространства V ортогональный относительно формы B. Проведем доказательство индукцией по $n = \dim V$. Если $n = 1$ или форма нулевая, то доказывать нечего. Пусть $n > 1$, а

$B \neq 0$. По предыдущей лемме существует $u_1 \in V$ такой, что $B(u_1, u_1) \neq 0$. Дополним u_1 до базиса $u = (u_1, \ldots, u_n)$ пространства V. Положим $w_k = u_k - \frac{B(u_k,u_1)}{B(u_1,u_1)}u_1$ при $2 \leqslant k \leqslant n$. Тогда

$$B(w_k, u_1) = B\Big(u_k - \frac{B(u_k, u_1)}{B(u_1, u_1)}u_1, u_1\Big) = B(u_k, u_1) - \frac{B(u_k, u_1)}{B(u_1, u_1)}B(u_1, u_1) = 0.$$

Таким образом, u_1 B-ортогонален всем векторам w_k при $2 \leqslant k \leqslant n$, а, следовательно, и всему подпространству $W = \langle w_2, \ldots, w_n\rangle$. Заметим, что $(u_1, w_2, \ldots, w_n)$ получено из u преобразованиями Гаусса, следовательно, является базисом. Поэтому $\dim W = n - 1$. По индукционному предположению существует B-ортогональный базис $(v_2, \ldots, v_n)$ подпространства W. Положив $v_1 = u_1$, получим ортогональный базис $(v_1, \ldots, v_n)$ пространства V. □

Покажем, как алгоритм диагонализации, приведенный в доказательстве теоремы, работает на практике. Путь u – произвольный базис пространства V, и $A = B_u$. При переходе к другому базису матрица A меняется на C^*AC. Если C – трансвекция, то умножение справа на C производит преобразование Гаусса со столбцами, а умножение слева на C^* – такое же преобразование со строками, но с сопряженными коэффициентами. Если $a_{11} = B(u_1, u_1) \neq 0$, то

$$\begin{pmatrix} a_{11} & c \\ c^* & \star \end{pmatrix} \approx \begin{pmatrix} 1 & 0 \\ -c^*/a_{11} & E \end{pmatrix} \begin{pmatrix} a_{11} & c \\ c^* & \star \end{pmatrix} \begin{pmatrix} 1 & -c/a_{11} \\ 0 & E \end{pmatrix} = \begin{pmatrix} a_{11} & 0 \\ 0 & \star \end{pmatrix},$$

где $c = (a_{12}, \ldots, a_{1n})$, а знак "$\approx$" стоит между матрицами одной и той же формы в разных базисах. После этого мы работаем с получившейся матрицей $(n-1) \times (n-1)$, обозначенной звездочкой.

Если $a_{11} = 0$, но какой-нибудь другой диагональный элемент матрицы A не равен нулю, то проделываем аналогичные преобразования используя этот элемент вместо a_{11}. Если же все диагональные элементы равны нулю, но $a_{ij} = \alpha \neq 0$ (такой элемент всегда найдется в ненулевой матрице), то по лемме 2.4 вектор $u_i + u_j\overline{\alpha}$ неизотропен относительно B. Значит достаточно к i-му столбцу прибавить j-й с коэффициентом $\overline{\alpha}$, к i-й строке прибавить j-ю с коэффициентом α, чтобы заработать ненулевой диагональный элемент. Проиллюстрируем сказанное на примере матрицы 2×2 с $a_{11} = a_{22} = 0$.

$$\begin{pmatrix} 0 & \alpha \\ \overline{\alpha} & 0 \end{pmatrix} \approx \begin{pmatrix} 1 & \alpha \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & \alpha \\ \overline{\alpha} & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ \overline{\alpha} & 1 \end{pmatrix} = \begin{pmatrix} 2\alpha\overline{\alpha} & \alpha \\ \overline{\alpha} & 0 \end{pmatrix} \approx \begin{pmatrix} 2\alpha\overline{\alpha} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix}.$$

Пусть Q – квадратичная форма. Если матрица Q_v диагональна, а $u_k = v_k\lambda_k$, то матрица Q_u также диагональна, причем ее диагональные элементы отличаются от диагональных элементов Q_v умножением на λ_k^2. Таким образом, для невырожденных форм, играют роль только классы вычетов диагональных элементов Q_v в $F^*/(F^*)^2$. В частности, если из любого элемента поля F можно извлечь квадратный корень, то классификация совсем простая.

Определение 2.6. Поле F называется квадратично замкнутым, если для любого $\alpha \in F$ уравнение $x^2 = \alpha$ имеет хотя бы одно решение. В частности, алгебраически замкнутое поле является квадратично замкнутым.

Следствие 2.7. *Любые два невырожденных квадратичных пространства одинаковой размерности над квадратично замкнутым полем изоморфны.*

3. Вещественные квадратичные формы

В этом параграфе V обозначает вещественное векторное пространство размерности m. В общем случае неверно, что любые две диагональные матрицы Q_u и Q_v квадратичной формы Q отличаются только перестановкой диагональных элементов и умножением их на квадраты. Однако, для поля вещественных чисел это так, и роль играют только знаки диагональных элементов.

Определение 3.1. Сигнатурой последовательности вещественных чисел называется пара чисел (p, n), где p – количество положительных среди этих чисел, а n – отрицательных.

Сигнатурой вещественной диагональной матрицы называется сигнатура последовательности ее диагональных элементов.

Лемма 3.2. *Пусть B – симметричная билинейная форма на V, а $v_1, \dots, v_k \in V$. Если $B(v_i, v_i) > 0$ при всех i, а $B(v_i, v_j) = 0$ при всех $j \neq i$, то форма B положительно определена на подпространстве $\langle v_1, \dots, v_k \rangle$.*

Теорема 3.3 (закон инерции квадратичных форм)**.** *Пусть Q – квадратичная форма на вещественном векторном пространстве V, а u, v – базисы V такие, что матрицы Q_u и Q_v диагональны. Тогда сигнатуры матриц Q_u и Q_v равны.*

Доказательство. По предложению 2.2 можно считать, что форма Q невырождена. Пусть (p_u, n_u) – сигнатура матрицы Q_u, а (p_v, n_v) – сигнатура матрицы Q_v. Перенумеровав при необходимости базисные элементы, можно считать, что положительные диагональные элементы матриц Q_u и Q_v стоят выше (и левее) отрицательных. Предположим, что $p_u > p_v$. По предыдущей лемме форма Q положительно определена на подпространстве $U = \langle u_1, \dots, u_{p_u} \rangle$. Аналогично, Q отрицательно определена на $W = \langle v_{p_v+1}, \dots, v_m \rangle$. Но по теореме о размерности суммы и пересечения

$$\dim U \cap W = p_u + (m - p_v) - \dim(U + W) \geqslant p_u + (m - p_v) - m = p_u - p_v > 0.$$

Следовательно, $U \cap W \neq \{0\}$, но форма Q одновременно положительно и отрицательно определена на этом подпространстве. Противоречие показывает, что неравенство $p_u > p_v$ невозможно. По аналогичным причинам невозможно и обратное неравенство, следовательно, $p_u = p_v$. □

В соответствии с законом инерции можно дать следующее определение. Сигнатурой квадратичной формы называется сигнатура ее матрицы в таком базисе, в котором эта матрица диагональна. Сигнатуру квадратичной формы Q будем обозначать через $\operatorname{Sign} Q$.

Следствие 3.4 (классификация вещественных квадратичных форм)**.** *Два вещественных квадратичных пространства (V, Q) и (V', Q') изометричны тогда и только тогда, когда $\dim V = \dim V'$, а $\operatorname{Sign} Q = \operatorname{Sign} Q'$.*

Пусть A – квадратная матрица. Определитель главной подматрицы $\begin{vmatrix} a_{11} & \dots & a_{1k} \\ \dots & \dots & \dots \\ a_{k1} & \dots & a_{kk} \end{vmatrix}$ называется главным минором k-ого порядка матрицы A. Обозначим его для краткости Δ_k. Напомним, что матрица называется унитреугольной, если она треугольна с 1 на главной диагонали.

Лемма 3.5. *Главные миноры матрицы не меняются при умножении на нижнюю унитреугольную матрицу слева и на верхнюю унитреугольную матрицу справа.*

Доказательство. При умножении матрицы слева на нижнюю унитреугольную главные подматрицы также умножаются на унитреугольные (это просто блочное умножение матриц) и, значит, их определители не меняются. □

По определению положим $\Delta_0 = 1$.

Теорема 3.6 (критерий Сильвестра)**.** *Пусть Q – квадратичная форма на V, а A – матрица формы Q. Предположим, что главные миноры $\Delta_1, \dots, \Delta_m$ матрицы A не равны нулю. Тогда сигнатура формы Q равна сигнатуре последовательности $(\frac{\Delta_1}{\Delta_0}, \dots, \frac{\Delta_m}{\Delta_{m-1}})$.*

В частности, форма положительно определена тогда и только тогда, когда все ее главные миноры больше 0.

Доказательство. По лемме 12.6 главы 2 матрица A лежит в главной клетке Гаусса, т. е. $A = BDC$, где B – нижняя унитреугольная матрица, C – верхняя унитреугольная, а D – диагональная. Так как A симметрична, то $BDC = (BDC)^\intercal = C^\intercal D B^\intercal$, откуда $B^{-1} C^\intercal D = DC(B^\intercal)^{-1}$. Но последняя матрица одновременно нижняя и верхняя треугольная, следовательно $B^{-1} C^\intercal = E$, т. е. $B = C^\intercal$. Это означает, что D – матрица той же квадратичной формы Q в другом базисе, а ее

сигнатура совпадает с сигнатурой формы Q. По лемме 3.5 главные миноры матриц A и D равны, т. е. $\Delta_k = d_{11} \dots d_{kk}$, откуда $d_{kk} = \Delta_k/\Delta_{k-1}$. Таким образом, последовательность из формулировке равна последовательности диагональных элементов матрицы D.

Если все главные миноры положительны, то в любой диагонализации формы Q все диагональные элементы положительны. По лемме 3.2 из этого следует, что форма положительно определена. Обратно, если Q положительно определена, то все диагональные элементы матрицы этой формы в любом базисе положительны. □

Замечание 3.7. Рассмотрим эрмитову форму B на комплексном векторном пространстве (инволюция – комплексное сопряжение). В соответствии с замечанием 1.8 диагональные элементы и главные миноры любой матрицы этой формы лежат в $\mathbb{R}$, поэтому для этой формы имеют смысл все утверждения настоящего параграфа. На самом деле все эти утверждения верны, а доказательства повторяют доказательства для вещественного случая с необходимыми уточнениями.

4. Пространства со скалярным произведением

В этом параграфе мы рассмотрим случай вещественных (комплексных) векторных пространств с положительно определенной симметричной билинейной (соотв. эрмитовой) формой, при этом инволюция на поле комплексных чисел – это обычное комплексное сопряжение. Можно рассматривать эти 2 случая одновременно, считая, что V – векторное пространство над полем $F = \mathbb{R}$ или $\mathbb{C}$, а в вещественном случае инволюция тривиальна. Для эрмитовой формы B и $x \in V$ имеем $B(x,x) = \overline{B(x,x)}$, откуда $B(x,x) \in \mathbb{R}$.

Определение 4.1. Эрмитова форма называется положительно определенной, если $B(x,x) > 0$ для любого вектора $x \neq 0$. Положительно определенная эрмитова форма называется (эрмитовым) скалярным произведением. В вещественным случае эрмитовость означает симметричность и билинейность, а скалярное произведение называется евклидовым. Конечномерное векторное пространство со скалярным произведением называется евклидовым или (классическим) эрмитовым пространством.

Квадратичная форма Q над произвольным полем называется анизотропной, если $Q(x) \neq 0$ при $x \neq 0$. Аналогично, мы будем говорить, что эрмитова форма B анизотропна, если $B(x,x) \neq 0$ при $x \neq 0$. В частности, скалярное произведение, является анизотропной формой. Позже мы узнаем, что для классификации квадратичных форм достаточно классифицировать анизотропные формы, так что они занимают особое место в теории квадратичных форм. По определению скалярное произведение является анизотропной формой, поэтому все утверждения, в которых не используется специфика поля вещественных или комплексных чисел [1]будут доказаны для произвольных анизотропных форм.

В дальнейшем скалярное произведение будет обозначаться просто (x,y) вместо $B(x,y)$, а матрица этой формы будет называться матрицей Грама и обозначаться через Γ или Γ_v, где v – базис. Соответствующая квадратичная форма будет обозначаться через $\|x\|^2 = (x,x)$. Другими словами, $\|x\| = \sqrt{(x,x)}$, что имеет смысл за счет положительной определенности.

Как следует из предложения 1.5, любое скалярное произведение на F^n имеет вид $(x,y) = x^*\Gamma y$, где $\Gamma \in \mathrm{M}_n(F)$ – матрица Грама в стандартном базисе. Эта формула всегда задает полуторалинейную форму, эрмитовость такой формы равносильна эрмитовости матрицы Γ, а положительная определенность, как мы узнаем чуть позже, – положительности собственных чисел Γ. В случае $\Gamma = E$ такое скалярное произведение называется стандартным.

Теорема 4.2. *Пусть V – векторное пространство (не обязательно конечномерное) со скалярным произведением. Тогда для любых $x, y \in V$ имеют место неравенства:*

(1) $|(x,y)| \leqslant \|x\|\|y\|$ *(неравенство Коши–Буняковского–Шварца — КБШ);*
(2) $\|x+y\| \leqslant \|x\| + \|y\|$ *(неравенство треугольника).*

[1]Эта специфика чаще всего проявляется в наличии неравенств.

Доказательство. (1). Если $y = 0$, то обе части равенства равны 0, поэтому можно считать, что $y \neq 0$. $0 \leqslant \|x + y\lambda\|^2 = \|x\|^2 + (x, y)\lambda + \bar{\lambda}(y, x) + \|y\|^2|\lambda|^2$. Положим $\lambda = -\frac{(y,x)}{(y,y)}$. Тогда последнее выражение равно $\|x\|^2 - \frac{(x,y)(y,x)+\overline{(y,x)}(y,x)}{(y,y)} + \frac{\|y\|^2|(y,x)|^2}{\|y\|^4} = \|x\|^2 - \frac{|(y,x)|^2}{\|y\|^2} \geqslant 0$. После домножения на знаменатель (который больше 0) получаем неравенство КБШ.

(2). Возводя неравенство треугольника в квадрат, получаем $(x+y, x+y) \leqslant \|x\|^2+\|y\|^2+2\|x\|\|y\|$. Раскрывая скобки в левой части и сокращая $\|x\|^2 + \|y\|^2$ имеем $(x, y) + (y, x) \leqslant 2\|x\|\|y\|$. Так как $(y, x) = \overline{(x, y)}$, левая часть равна $2\operatorname{Re}(x, y)$. Но $\operatorname{Re}(x, y) \leqslant |(x, y)| \leqslant \|x\| \cdot \|y\|$ по неравенству КБШ, откуда следует последнее неравенство. □

Следующие несколько утверждений верны для произвольной анизотропной эрмитовой формы B.

Лемма 4.3. *Набор ненулевых попарно B-ортогональных векторов линейно независим.*

Теорема 4.4 (ортогонализация Грама–Шмидта)**.** *Пусть B – анизотропная форма на V, $u_1, \dots, u_n \in V$. Положим*

$$v_1 = u_1$$
$$\dots\dots\dots\dots$$
$$v_n = u_n - \sum_{i=1}^{n-1} v_i \frac{B(v_i, u_n)}{B(v_i, v_i)}$$

(если $v_i = 0$, то соответствующее слагаемое отсутствует; строго говоря, надо было бы писать, что коэффициент при v_i в формуле для v_k равен 0, если $v_i = 0$, и $\frac{B(v_i,u_k)}{B(v_i,v_i)}$ в противном случае). Для любых $i, j, k \in \{1, \dots, n\}$, $i \neq j$ выполнены следующие утверждения.

(1) $B(v_i, v_j) = 0$.
(2) $\langle u_1, \dots, u_k\rangle = \langle v_1, \dots, v_k\rangle$.
(3) *Если $u_1, \dots, u_k$ линейно независимы, то и $v_1, \dots, v_k$ линейно независимы, в частности, $v_m \neq 0$ при всех $m = 1, \dots, k$.*
(4) *Если $u_k \in \langle u_1, \dots, u_{k-1}\rangle$, то $v_k = 0$.*
(5) *Если $(u_1, \dots, u_n)$ – система образующих V, то ненулевые из векторов $v_1, \dots, v_n$ образуют базис.*
(6) *Если $(u_1, \dots, u_n)$ – базис V, то $(v_1, \dots, v_n)$ – B-ортогональный базис V.*

Процесс ортогонализации – конструктивный способ построения ортогонального базиса пространства, существование которого мы уже доказали для произвольной эрмитовой формы в теореме 2.5. На самом деле доказательство теоремы 2.5 для анизотропной формы совпадает с процессом ортогонализации. В случае скалярного произведения можно сделать нормы всех базисных векторов равными 1. Такой базис будет называться ортонормированным.

В следующем утверждении в качестве B традиционно рассматривается стандартное скалярное произведение на F^m, хотя на самом деле это не имеет никакого значения. Для доказательства достаточно применить процесс ортогонализации к столбцам матрицы A.

Следствие 4.5 (QR-разложение)**.** *Пусть B – анизотропная эрмитова форма на F^m. Для любой матрицы $A \in \mathrm{M}_{m,n}(F)$ существует матрица $Q \in \mathrm{M}_{m,n}(F)$ с B-ортогональными столбцами и верхняя унитреугольная матрица $R \in \mathrm{M}_n(F)$ такие, что $A = QR$.*

Определение 4.6. Для подпространства $U \leqslant V$ определим ортогональное дополнение U формулой

$$U^{\perp_B} = U^{\perp} := \{x \in V \mid B(y, x) = 0\, \forall y \in U\}.$$

Нетрудно видеть, что сужение формы B на U невырождено тогда и только тогда, когда $U \cap U^{\perp} = \{0\}$. В конечномерном пространстве из подсчета размерностей сразу следует, что в этом случае ортогональное дополнение является и прямым дополнением.

Предложение 4.7. *Пусть B невырожденная эрмитова форма на конечномерном пространстве V, а $U \leqslant V$. Тогда* $\dim U^\perp = \dim V - \dim U$ *и* $(U^\perp)^\perp = U$.

Если сужение B на U невырождено, то $V = U \oplus U^\perp$. В частности, если B анизотропна, то равенства выполнены для любого подпространства U.

Доказательство. Пусть $u = (u_1, \dots, u_k)$ – базис U, а $L : V \to F^k$ – линейное отображение, заданное формулой $L(x) = \big(B(u_1, x), \dots, B(u_k, x)\big)^\intercal$. Ясно, что $U^\perp = \operatorname{Ker} L$. Если v – базис V, то можно записать наше отображение в координатной форме: $L(x) = CB_v x_v$, где $C = \big(\overline{(u_1)_v}, \dots, \overline{(u_k)_v}\big)^\intercal$. Так как набор u линейно независим, ранг матрицы C равен k. Так как B невырождена, матрица B_v обратима и умножение на нее не меняет ранга матрицы. Следовательно, размерность образа оператора L равна k, а по теореме о размерности ядра и образа $\dim \operatorname{Ker} L = \dim V - k$.

Очевидно, что $U \leqslant (U^\perp)^\perp$. С другой стороны, из доказанного следует, что их размерности равны, откуда вытекает второе равенство.

Если сужение B на U невырождено, то $U \cap U^\perp = \{0\}$. Но мы уже доказали, что $\dim U + \dim U^\perp = \dim V$, следовательно, $U \oplus U^\perp = V$. □

Вернемся к ситуации, когда форма анизотропна. Тогда другое доказательство того, что $V = U \oplus U^\perp$ дает процесс ортогонализации, примененный к последовательности $(u_1, \dots, u_k, x)$, где x произвольный элемент из V. Действительно, на последнем шаге мы получим вектор $y \in x + U$, ортогональный U, откуда $x \in y + U \subseteq U^\perp + U$.

Из последнего предложения следует, что каждый вектор $x \in V$ единственным образом представляется в виде $x = z + y$, где $z \in U$, а $y \in U^\perp$. В этом случае элемент z называется ортогональной проекцией x на U. Мы будем обозначать его $\operatorname{pr}_U x$. Проекция на вектор – это проекция на подпространство, порожденное этим вектором. Легко видеть, что

$$\operatorname{pr}_z x = \frac{(z, x)}{(z, z)} z.$$

В этих терминах формулу процесса ортогонализации можно произнести следующим образом: новый вектор v_k равен разности старого u_k и его проекций на все вектора, найденные на предыдущих шагах.

Обратите внимание, что если u является ортогональным базисом подпространства U, то $\operatorname{pr}_U x = \sum_{j=1}^k \operatorname{pr}_{u_j} x$, если же базис u не ортогонален, то такое равенство может выполняться только случайно.[2]

Предложение 4.8. *Пусть $u_1, \dots, u_k$ – ортогональный набор ненулевых векторов пространства V со скалярным произведением, а $x \in V$. Обозначим через U линейную оболочку векторов $u_1, \dots, u_k$. Тогда*

(1) $\operatorname{pr}_U x = \sum\limits_{j=1}^k \operatorname{pr}_{u_j} x = \sum\limits_{j=1}^k \frac{(u_j, x)}{(u_j, u_j)} u_j$;

(2) $\|x\|^2 \geqslant \|\operatorname{pr}_U x\|^2 = \sum\limits_{j=1}^k \frac{|(u_j, x)|^2}{(u_j, u_j)}$ *(неравенство Бесселя);*

(3) *равенство в последней формуле имеет место тогда и только тогда, когда $x \in U$ (равенство Парсеваля).*

Доказательство. Пусть $z = \operatorname{pr}_U x = \sum\limits_{j=1}^k u_j \alpha_j$. Тогда $x = z + y = \sum\limits_{j=1}^k u_j \alpha_j + y$ для некоторого $y \in U^\perp$. Домножая выражение для x скалярно (слева) на u_i, получаем формулу для α_i, откуда следует первый пункт предложения.

[2] *Упражнение.* Найдите условие на u равносильное существованию ненулевого x, для которого равенство верно.

Так как набор $u_1, \dots, u_k$ ортогонален, то скалярный квадрат правой части выражения для z равен сумме скалярных квадратов слагаемых. Следовательно, правая часть неравенства Бесселя равна $\|z\|^2$. Но $\|x\|^2 = \|z+y\|^2 = \|z\|^2 + \|y\|^2 \geqslant \|z\|^2$, причем равенство достигается, только если $y = 0$. □

Неравенство Бесселя является всего лишь выражением того факта, что длина гипотенузы не меньше длины катета. Следующее утверждение сравнивает длину гипотенузы того же прямоугольного треугольника с длиной другого катета.

Предложение 4.9. *Для любого $u \in U$ выполнено неравенство $\|x - u\| \geqslant \|x - \mathrm{pr}_U x\|$.*

Доказательство. Пусть $x = z + y$, где $z = \mathrm{pr}_U x$, а $y \in U^\perp$. Тогда для любого $u \in U$:

$$\|x-u\|^2 = \|y + (z-u)\|^2 = \|y\|^2 + \|z-u\|^2 \geqslant \|y\|^2 = \|x - \mathrm{pr}_U x\|^2.$$

□

До конца параграфа $V = F^m$ пространство со стандартным скалярным произведением.

Лемма 4.10. *Пусть $A \in \mathrm{M}_{m,n}(F)$ – матрица с линейно независимыми столбцами. Тогда матрица A^*A невырождена.*

Доказательство. Если $U \leqslant V$ – подпространство, порожденное столбцами матрицы A, то A^*A – матрица Грама сужения скалярного произведения на $U \times U$ в базисе из столбцов A. Так как наше скалярное произведение невырождено на любом подпространстве, то эта матрица невырождена. □

Теорема 4.11 (метод наименьших квадратов)**.** *Пусть $A \in \mathrm{M}_{m,n}(F)$ – матрица с линейно независимыми столбцами, а $b \in F^m$. Тогда $\|Ax - b\|$ минимальна, если $x \in F^n$ удовлетворяет уравнению $A^*Ax = A^*b$. (заметим, что по предыдущей лемме такой вектор x всегда существует и единственный).*

Доказательство. Для любого $y \in F^n$

$$(Ay, Ax - b) = y^*(A^*Ax - A^*b) = 0.$$

Таким образом, вектор $Ax - b$ ортогонален подпространству, порожденному столбцами матрицы A, т. е. Ax является ортогональной проекцией b на это подпространство. Теперь результат следует из предложения 4.9. □

5. Нормальные операторы

Определение 5.1. Пусть V – векторное пространство над произвольным полем F. Множество линейных отображений из V в F

$$V^* := \mathrm{Hom}_F(V, F)$$

с поточечными операциями называется пространством, двойственным к V (dual space).

Пусть U и V – векторные пространства, а $L : U \to V$ – линейное отображение. Тогда двойственное отображение $L^* : V^* \to U^*$ задается формулой $L^*(\varphi) = \varphi \circ L$. Двойственное отображение чаще называют сопряженным, в нашем курсе будет небольшая разница между ними.

Заметим, что $(L \circ M)^* = M^* \circ L^*$.

Если V конечномерно, то при помощи выбора базиса его можно отождествить с F^n. Тогда любое линейное отображение $V \to F$ – это умножение слева на строку длины n, следовательно, V^* отождествляется с ${}^nF$. Так как эти пространства имеют одинаковую размерность, то они изоморфны.[3] В случае бесконечномерных пространств это совсем не так, например базис пространства, двойственного к счетномерному, имеет мощность континуум.

[3] Это первый пример в нашем курсе неканонического изоморфизма, т. е. изоморфизма, который зависит от какого-то выбора.

Далее на протяжении этого параграфа пространство V конечномерно. Любая невырожденная полуторалинейная форма B задает полулинейную биекцию пространства на его двойственное. А именно, для $x \in V$ положим $\varphi_x(y) = B(x,y)$. Тогда $\varphi_x \in V^*$, а $x \mapsto \varphi_x$ – искомая биекция. Действительно, невырожденность формы B равносильна тривиальности ядра этого отображения; ясно, что отображение полулинейно, т. е. $\varphi_{x+z} = \varphi_x + \varphi_z$ и $\varphi_{\alpha x} = \overline{\alpha}\varphi_x$; а сюръективность следует из теоремы о размерности ядра и образа, которая верна и для полулинейных отображений. Большую часть утверждений настоящего параграфа можно сформулировать для эрмитовых пространств над произвольным полем. Мы, однако, оставим эти обобщения в качестве упражнения и будем рассматривать только евклидовы и классические эрмитовы пространства.

Если U и V эрмитовы пространства, то мы можем отождествить их со своими двойственными при помощи полулинейной биекции, определенной в предыдущем абзаце. Тогда L^* отождествляется с линейным отображением $L^\star : V \to U$, для которого следующая диаграмма коммутативна.

$$\begin{array}{ccc} V & \xrightarrow{L^\star} & U \\ {\scriptstyle x\mapsto\varphi_x}\downarrow & & \downarrow{\scriptstyle y\mapsto\varphi_y} \\ V^* & \xrightarrow{L^*} & U^* \end{array}$$

Обратите внимание, что пока мы различаем $L^* : V^* \to U^*$ и $L^\star : V \to U$ (у них разные символы “звездочка”).

Лемма 5.2. *Определенное выше отображение $L^\star$ удовлетворяет равенству*

$$\big(L^\star(x), y\big)_U = \big(x, L(y)\big)_V \tag{32}$$

и $L^{\star\star} = L$.

Доказательство. Пусть $x \in V$. Тогда коммутативность диаграммы говорит, что $L^*(\varphi_x) = \varphi_{L^\star(x)}$. Применяя обе части равенства к элементу $y \in U$ получим

$$L^*(\varphi_x)(y) = \varphi_x\big(L(y)\big) = \big(x, L(y)\big)_V = \varphi_{L^\star(x)}(y) = \big(L^\star(x), y\big)_U.$$

Далее,

$$\big(x, L(y)\big)_V = \big(L^\star(x), y\big)_U = \overline{\big(y, L^\star(x)\big)_U} = \overline{\big(L^{\star\star}(y), x\big)_V} = \big(x, L^{\star\star}(y)\big)_V.$$

Так как последнее равенство выполнено для любого $x \in V$, а скалярное произведение невырождено, то $L(y) = L^{\star\star}(y)$ для любого $y \in U$, что и означает $L^{\star\star} = L$. □

Последняя лемма говорит в частности, что отображение $L^\star$, удовлетворяющее равенству (32) существует. Единственность такого отображения очевидна. Таким образом, мы получаем 2 равносильных определения оператора $L^\star$, ни одно из которых не выглядит конструктивным, так как в каждом из них надо искать вектор, скалярное умножение на который реализует данный линейный функционал. В действительности же, для нахождения такого вектора надо просто решить систему линейных уравнений.

В дальнейшем, мы меняем обозначение $L^\star$ на общепринятое L^*. Какой из операторов имеется в виду $V \to U$ или $V^* \to U^*$, всегда будет ясно из контекста. Обратите внимание, что первая версия L^* употребляется, только если зафиксированы изоморфизмы между пространствами и сопряженными к ним, в частности, если U и V – эрмитовы (квадратичные) пространства или если там зафиксированы базисы. В этом и следующем параграфах L^* обозначает оператор $V \to U$.

Лемма 5.3. *Пусть u – ортонормированный базис эрмитова пространства V, а $L : V \to V$. Тогда $(L^*)_u = (L_u)^*$.*

Доказательство. $\big(y, L^*(x)\big) = \big(L(y), x\big)$. Так как базис u ортонормирован, то $(z,t) = z_u^* t_u$. Используя это и определение матрицы оператора получим $y_u^*(L^*)_u x_u = (L_u y_u)^* x_u = y_u^*(L_u)^* x_u$. Так как x_u и y_u – произвольные столбцы, получаем требуемое равенство матриц. □

Определение 5.4. Оператор $L : V \to V$ называется самосопряженным или эрмитовым, если $L^* = L$.

Изоморфизм $L : U \to V$ называется унитарным или изометрией, если $L^* = L^{-1}$.

Оператор $L : V \to V$ называется нормальным, если $L^*L = LL^*$.

Таким образом, нормальные операторы является одновременным обобщением таких важных типов, как самосопряженные и унитарные операторы (а на самом деле еще и косоэрмитовы, т. е. те, для которых $L^* = -L$). Следующее утверждение говорит о связи эрмитовых форм и самосопряженных операторов.

Лемма 5.5. *Пусть $L : V \to V$ – линейный оператор, а $B = B_L$ – полуторалинейная форма, заданная равенством $B(x, y) = \big(x, L(y)\big)$. Тогда*

- *$B_u = L_u$ в любом ортонормированном базисе u;*
- *L самосопряженный тогда и только тогда, когда B эрмитова.*

Теорема 5.6. *Для любого нормального оператора существует ортонормированный базис из собственных векторов.*

Собственные вектора нормального оператора, соответствующие различным собственным числам, ортогональны.

Собственные числа самосопряженного оператора вещественны.

Собственные числа унитарного оператора по модулю равны 1.

Доказательство первого утверждения опирается на 2 леммы.

Лемма 5.7. *Если $AB = BA$, то собственное подпространство оператора A инвариантно относительно B.*

Доказательство. Если $Ax = \lambda x$, то $A(Bx) = B(Ax) = B(\lambda x) = \lambda Bx$, т. е. Bx принадлежит тому же собственному подпространству, что и x. □

Лемма 5.8. *Если $U \leqslant V$ инвариантно относительно оператора $L : V \to V$, то $U^\perp$ инвариантно относительно L^*.*

Доказательство. Пусть $x \in U^\perp$, а y – произвольный вектор из U. Так как $L(y) \in U$, то $\big(L^*(x), y\big) = \big(x, L(y)\big) = 0$. Таким образом, $L^*(x) \in U^\perp$. □

Доказательство теоремы 5.6. Пусть L – нормальный оператор в эрмитовом пространстве V. Проведем доказательство индукцией по $\dim V$. Так как $\mathbb{C}$ алгебраически замкнуто, то существует хотя бы один корень характеристического многочлена, следовательно, хотя бы одно ненулевое собственное подпространство V_λ. Так как $LL^* = L^*L$, то по лемме 5.7 V_λ инвариантно относительно L^*. Теперь по лемме 5.8 $V_\lambda^\perp$ инвариантно относительно $L^{**} = L$. Размерность пространства $V_\lambda^\perp$ строго меньше $\dim V$, поэтому к нему можно применить индукционное предположение. Выберем ортогональный базис $(v_1, \dots, v_k)$ подпространства $V_\lambda^\perp$ из собственных векторов оператора L и ортогональный базис $(v_{k+1}, \dots, v_n)$ подпространства V_λ. Тогда $(v_1, \dots, v_n)$ – искомый базис пространства V.

Из доказанного следует, что собственные подпространства нормального оператора попарно ортогональны, откуда вытекает второе утверждение.

Пусть $L(x) = \lambda x$ для некоторого $x \neq 0$. Если $L = L^*$, то

$$\bar\lambda\big(x, x\big) = \big(L(x), x\big) = \big(x, L(x)\big) = \lambda\big(x, x\big),$$

откуда $\bar\lambda = \lambda$, т. е. $\lambda \in \mathbb{R}$.

Аналогично, если $L^{-1} = L^*$, то

$$\bar\lambda\lambda\big(x, x\big) = \big(L(x), L(x)\big) = \big(x, x\big).$$

Сокращая на (x, x), получаем $|\lambda|^2 = \bar\lambda\lambda = 1$. □

На самом деле условия, собранные в теореме 5.6 являются характеризациями соответствующих типов операторов. Следствие в одну сторону доказано в теореме, а обратное сразу следует из леммы 5.3.

Следствие 5.9. *Оператор L является нормальным тогда и только тогда, когда существует ортогональный базис из его собственных векторов. При этом он самосопряжен тогда и только тогда, когда его собственные числа вещественные, и является унитарным тогда и только тогда, когда его собственные числа по модулю равны* 1.

Следующее утверждение является частным случаем спектральной теоремы, которая говорит о строении самосопряженных операторов в гильбертовых пространствах. Она формулируется в терминах ортогональных проекторов.

Определение 5.10. Оператор $P : V \to V$ называется проектором, если $P^2 = P$. Если $\operatorname{Ker} P \perp \operatorname{Im} P$, то проектор P называется ортогональным.

Лемма 5.11. *Если P – проектор в произвольном векторном пространстве V, то $V = \operatorname{Ker} P \oplus \operatorname{Im} P$.*

Доказательство. Если P – проектор, то $P(x-P(x)) = 0$, т. е. $x-P(x) \in \operatorname{Ker} P$, следовательно, $x = x - P(x) + P(x) \in \operatorname{Ker} P + \operatorname{Im} P$. с другой стороны, $P(x) \in \operatorname{Ker} P \implies P(x) = P^2(x) = 0$, откуда $\operatorname{Im} P \cap \operatorname{Ker} P = \{0\}$. Отсюда следует, что для ортогонального проектора P имеет место равенство $\operatorname{Ker} P = (\operatorname{Im} P)^\perp$. □

Нетрудно видеть, что ортогональный проектор является самосопряженным оператором.

Следствие 5.12 (спектральная теорема)**.** *Оператор L в евклидовом (эрмитовом) пространстве V является самосопряженным тогда и только тогда, когда он представляется в виде линейной комбинации ортогональных проекторов на попарно ортогональные подпространства с вещественными коэффициентами.*

Эти ортогональные подпространства можно считать одномерными, порожденными собственными векторами $u_1, \dots, u_n$ оператора L. Тогда

$$L = \sum_{k=1}^{n} \lambda_k P_k, \text{ где } P_k(x) = \operatorname{pr}_{u_k} x = \frac{(u_k, x)}{(u_k, u_k)} u_k.$$

Доказательство. Очевидно, что линейная комбинация самосопряженных операторов с вещественными коэффициентами является самосопряженным оператором. Обратно, по теореме 5.6 существует ортогональный базис из собственных векторов $u_1, \dots, u_n$ оператора L. По формуле "координаты в ортогональном базисе" (теорема 4.8(1)) $x = \sum_{k=1}^{n} \frac{(u_k,x)}{(u_k,u_k)} u_k$, откуда

$$L(x) = \sum_{k=1}^{n} \frac{(u_k, x)}{(u_k, u_k)} L(u_k) = \sum_{k=1}^{n} \lambda_k \frac{(u_k, x)}{(u_k, u_k)} u_k.$$

При этом мы уже доказывали, что формула $P_k(x) = \operatorname{pr}_{u_k} x = \frac{(u_k,x)}{(u_k,u_k)} u_k$ задает ортогональный проектор на $\langle u_k \rangle$, что доказывает оба утверждения. □

Проекторы на собственные вектора, соответствующие одному и тому же собственному числу, можно собрать в один проектор на собственное подпространство. Так что можно выразить самосопряженный оператор через линейную комбинацию проекторов на его собственные подпространства.

Следствие 5.13. *Для любой эрмитовой формы в эрмитовом пространстве V существует ортонормированный базис, в котором матрица этой формы диагональна.*

Для любой квадратичной формы в евклидовом пространстве V существует ортонормированный базис, в котором матрица этой формы диагональна.

Доказательство. Пусть $L : V \to V$ – самосопряженный оператор, соответствующий данной форме. По теореме 5.6 существует ортогональный базис из собственных векторов этого оператора. Нормируя эти вектора (что не портит их ортогональности и "собственности") получаем ортонормированный базис из собственных векторов. В этом базисе из собственных векторов матрица оператора диагональна, а по лемме 5.5 она равна матрице исходной формы.

В евклидовом пространстве надо только заметить, что, так как собственные числа оператора L вещественны, то у него существуют собственные вектора в исходном пространстве, а дальше повторить доказательство теоремы 5.6 для евклидова случая. □

6. Матричные разложения

В этом параграфе снова F – это поле вещественных или комплексных чисел.

Определение 6.1. Самосопряженный оператор называется положительно (неотрицательно) определенным, если все его собственные числа положительны (неотрицательны). Это равносильно тому, что эрмитова форма, заданная этим оператором, положительно (неотрицательно) определена.

Эрмитова матрица $A \in \mathrm{M}_n(\mathbb{C})$ называется положительно (неотрицательно) определенной, если эрмитова форма $B(x, y) = x^*Ay$ положительно (неотрицательно) определена.

Для диагональной матрицы $D \in \mathrm{M}_n(\mathbb{R})$ с неотрицательными диагональными элементами положим $\sqrt{D} = \mathrm{diag}(\sqrt{d_{11}}, \ldots, \sqrt{d_{nn}})$ (берутся арифметические квадратные корни). Сейчас мы определим, точнее, докажем корректность определения квадратного корня из произвольной неотрицательно определенной эрмитовой матрицы. Это утверждение понадобится для доказательства единственности полярного разложения. Обобщение этого утверждения на другие полиномиальные матричные уравнения (вместо $X^2 = A$) оставляется читателю в качестве упражнения.

Лемма 6.2. *Пусть $A \in \mathrm{M}_n(\mathbb{C})$ – неотрицательно определенная эрмитова матрица. Тогда существует единственная неотрицательно определенная эрмитова матрица $H \in \mathrm{M}_n(\mathbb{C})$ такая, что $H^2 = A$.*

Доказательство. По теореме 5.6 существует матрица $C \in \mathrm{GL}_n(\mathbb{C})$ такая, что $A = C^{-1}DC$, где D диагональна с неотрицательными диагональными элементами. Тогда $H = C^{-1}\sqrt{D}C$ удовлетворяет условиям леммы.

Если H удовлетворяет условиям леммы, то $H = G^{-1}\tilde{D}G$ для некоторой матрицы $G \in \mathrm{GL}_n(\mathbb{C})$ и диагональной $\tilde{D}$ с неотрицательными элементами по диагонали. Тогда $H^2 = G^{-1}\tilde{D}^2G = C^{-1}DC$. Поэтому собственные числа матриц $\tilde{D}^2$ и D и их кратности совпадают. Меняя при необходимости порядок собственных векторов матриц A и H можно считать, что $\tilde{D}^2 = D = \mathrm{diag}(\lambda_1 E, \ldots, \lambda_k E)$, где $\lambda_j \neq \lambda_\ell$ при $j \neq \ell$. Из выражения для H^2 получаем $(CG^{-1})D = D(CG^{-1})$. Вычисление показывает, что CG^{-1} – блочно диагональная матрица с блоками тех же размеров, что и диагональные блоки матрицы D. Заметим, что $\tilde{D} = \mathrm{diag}(\sqrt{\lambda_1}E, \ldots, \sqrt{\lambda_k}E)$ также коммутирует с любой такой блочно диагональной матрицей, откуда $H = G^{-1}\tilde{D}G = C^{-1}\tilde{D}C$ – единственная матрица с неотрицательными собственными числами, удовлетворяющая равенству $H^2 = A$. □

Теорема 6.3 (разложение Холецкого)**.** *Любая положительно определенная эрмитова матрица $A \in \mathrm{M}_n(\mathbb{C})$ представляется в виде произведения $A = C^*C$, для некоторой верхнетреугольной матрицы C с положительными диагональными элементами. При этом матрица C определена единственным образом.*

Доказательство. Пусть $B(x, y) = x^*Ay$ – эрмитова форма на $\mathbb{C}^n$. Ее матрица в стандартном базисе $B_e = A$. По определению положительная определенность A равносильна положительной определенности B. По теореме 2.5 существует базис, в котором эта форма диагональна, а так как она положительно определена, то диагональные элементы вещественны и положительны. Домножая базисные вектора на соответствующие вещественные константы можно

добиться того, чтобы эта матрица стала единичной. Если u такой базис $\mathbb{C}^n$, что $B_u = E$, то $A = B_e = C^*_{u\to e} B_u C_{u\to e} = C^*_{u\to e} C_{u\to e}$. По QR-разложению $C_{u\to e} = QC$ для некоторой матрицы Q с ортонормированными столбцами и верхнетреугольной матрицы C с положительными элементами по диагонали. Тогда $A = (QC)^*QC = C^*Q^*QC = C^*C$.

На самом деле, это доказательство равносильно применению процесса ортогонализации Грама–Шмидта относительно скалярного произведения B к векторам стандартного базиса. Действительно, процесс ортогонализации (вместе с нормированием) говорит, что умножая стандартный базис на верхнетреугольную матрицу с положительными диагональными элементами, можно получить базис v, ортонормированный относительно B. Это означает, что $B_v = E$, а $C_{e\to v}$ верхнетреугольная, откуда $A = B_e = C^*_{e\to v} C_{e\to v}$.

Другое доказательство существования вытекает из разложения Гаусса и совпадает с началом доказательства критерия Сильвестра. А именно, так как форма $B(x,y) = x^*Ay$ положительно определена, то все главные миноры ее матрицы в любом базисе положительны. Следовательно, по лемме 12.6 главы 2 A лежит в главной клетке Гаусса, т. е. $A = GDH$ для некоторой верхней унитреугольной матрицы H, диагональной матрицы D и нижней унитреугольной матрицы G. Так как $A = A^*$, то $GDH = H^*D^*G^*$ откуда $(H^*)^{-1}GD = D^*G^*H^{-1}$. Последняя матрица является одновременно верхней и нижней треугольной, следовательно, $G = H^*$ и $D = D^*$. Из положительной определенности сразу следует, что диагональные элементы матрицы D положительны. Следовательно, $A = (H^*\sqrt{D})(\sqrt{D}H)$ – искомое разложение.

Единственность следует из единственности разложения Гаусса. Если $C^*C = \tilde{C}^*\tilde{C}$ для некоторых верхнетреугольных матриц $C, \tilde{C}$ с положительными диагональными элементами, то матрица $(\tilde{C}^*)^{-1}C^* = \tilde{C}C^{-1}$ верхнетреугольна и нижнетреугольна одновременно, следовательно, она диагональна. То есть $\tilde{C} = LC$ для некоторой диагональной матрицы L. Так как диагональные элементы C и $\tilde{C}$ положительны, то L обладает тем же свойством. Кроме того, $C^*C = \tilde{C}^*\tilde{C} = (LC)^*LC = C^*(L^*L)C$, откуда $L^*L = L^2 = E$. Таким образом, $L = E$ и $\tilde{C} = C$. □

Теорема 6.4 (сингулярное разложение). *Любая матрица $A \in \mathrm{M}_{m,n}(\mathbb{C})$ представляется в виде произведения $A = BDC$, где B и C – квадратные унитарные матрицы, а $D \in \mathrm{M}_{m,n}(\mathbb{C})$ диагональна (т. е. $d_{ij} = 0$ при всех $i \neq j$) с неотрицательными элементами по диагонали.*

На языке линейных отображений это же утверждение звучит следующим образом.

Для любого линейного отображения $L : U \to V$, где U и V – классические эрмитовы пространства, существуют ортонормированные базисы u и v пространств U и V соответственно, в которых матрица $D = L^u_v$ оператора L диагональна с неотрицательными элементами по диагонали.

При этом диагональные элементы матрицы D определены однозначно с точностью до перестановки и равны корням из собственных чисел матрицы A^*A (соотв. оператора L^*L) или 0.

Доказательство. Пусть u и v ортонормированные базисы пространств U и V соответственно. Матрица L^u_v диагональна с элементами α_i по диагонали тогда и только тогда, когда $L(u_i) = \alpha_i v_i$ или $L(u_i) = 0$ (последнее существенно, если $\dim U > \dim V$). Ортогональность базиса v влечет $\big(L(u_i), L(u_j)\big) = 0$ при всех $i \neq j$. Перепишем последнее равенство в виде $\big(L^*L(u_i), u_j\big) = 0$, т. е. вектор $L^*L(u_i)$ лежит в ортогональном дополнении подпространства, порожденного всеми u_j при $j \neq i$. Заметим, что это ортогональное дополнение одномерно и содержит u_i. Следовательно, $L^*L(u_i) = \lambda_i u_i$. Таким образом, если сингулярное разложение оператора существует, то u_i – собственные вектора оператора L^*L. Этот оператор самосопряженный и неотрицательно определенный, поэтому все λ_i – вещественные неотрицательные числа. Так как $\|v_i\| = 1$, а $\alpha_i > 0$, то $\alpha_i = \|L(u_i)\|$. Имеем

$$\alpha_i^2 = \big(L(u_i), L(u_i)\big) = \big(L^*L(u_i), u_i\big) = \lambda_i\big(u_i, u_i\big) = \lambda_i.$$

Мы доказали, что если сингулярное разложение существует, то матрица D определена единственным образом с точностью до перестановки диагональных элементов.

По теореме 5.6 существует ортонормированный базис из собственных векторов оператора L^*L. Расположим эти вектора так, чтобы $u_1, \dots, u_k$ соответствовали ненулевым собственным числам, а остальные – собственному числу 0. Тогда при $i > k$ имеем $0 = \big(L^*L(u_i), u_i\big) = \big(L(u_i), L(u_i)\big)$, откуда $L(u_i) = 0$. Вектора $v_j = \frac{L(u_j)}{\|L(u_j)\|}$ при $j \leqslant k$ образуют базис образа оператора L. Как было показано выше, $L(u_j) \perp L(u_h)$ при $j \neq h$, поэтому этот базис ортонормированный. Дополним его до ортонормированного базиса v пространства V (это можно сделать, дополнив до произвольного базиса, применив процесс ортогонализации и нормировав полученные вектора). Таким образом, в ортонормированных базисах u и v матрица оператора L диагональна, что завершает доказательство. □

Следствие 6.5 (полярное разложение)**.** *Любая матрица $A \in \mathrm{M}_n(\mathbb{C})$ представляется в виде произведения $A = GH$, где G – унитарная матрицы, а $H \in \mathrm{M}_n(\mathbb{C})$ – неотрицательно определенная эрмитова. При этом матрица H определена единственным образом, а если A обратима, то и G единственна.*

Доказательство. Пусть $A = BDC$ – сингулярное разложение матрицы A. Тогда $A = (BC)(C^{-1}DC)$ – ее полярное разложение. Действительно, произведение унитарных матриц унитарно, а $C^{-1}DC = C^*DC$ – эрмитова.

Если $A = GH$, где G унитарная, а H эрмитова, то $A^*A = H^*G^*GH = H^2$. Но по лемме 6.2 такая положительно определенная матрица H единственна. Если же A обратима, то обратима и H, поэтому $G = AH^{-1}$ единственна. □

Очевидно, что самосопряженность оператора равносильна самосопряженности его матрицы в ортонормированном базисе. Таким образом, можно говорить о полярном разложении оператора.

Следствие 6.6. *Пусть V – евклидово или эрмитово пространство. Линейный оператор $L : V \to V$ является нормальным тогда и только тогда, когда унитарный и самосопряженный операторы в его полярном разложении коммутируют.*

Доказательство. Если $L = GH$, где G – унитарный, а H – эрмитов, то $L^*L = H^*G^*GH = H^2$, а $LL^* = GHH^*G^* = GH^2G^{-1}$. Таким образом, $L^*L = LL^*$ тогда и только тогда, когда G коммутирует с H^2. В этом случае $(GHG^{-1})^2 = H^2$, и по лемме 6.2 $GHG^{-1} = H$, т. е. G и H коммутируют. Обратная импликация очевидна. □

7. Гильбертово пространство

Не вдаваясь в подробности, в этом параграфе мы обсудим, какие утверждения этой главы про евклидовы (эрмитовы) пространства выживают в бесконечномерных пространствах со скалярным произведением. Мы будем рассматривать только одно (с точностью до изометрии) бесконечномерное пространство, а именно то, которым в основном интересуется функциональный анализ – сепарабельное гильбертово пространство.

Пусть V – (бесконечномерное) векторное пространство над $\mathbb{R}$ или $\mathbb{C}$ с евклидовым (соотв. эрмитовым) скалярным произведением. Также как в предыдущих параграфах мы будем обозначать основное поле буквой F, а в вещественном случае черта будет тождественной инволюцией. Скалярное произведение задает норму в пространстве V, которая в соответствии с неравенством треугольника превращает его в нормированное, а, следовательно, и топологическое пространство. Оно называется полным, если любая фундаментальная последовательность имеет предел. Оно называется сепарабельным, если в нем существует плотное счетномерное подпространство. Гильбертово пространство – это полное пространство со скалярным произведением. Для простоты мы считаем наше гильбертово пространство сепарабельным. Все сепарабельные гильбертовы пространства над F изометричны.

Моделью сепарабельного гильбертова пространства является множество ℓ_2 бесконечных последовательностей $x = (x_n)_{n=0}^{\infty}$, для которых ряд $\sum_{n=0}^{\infty} |x_n|^2$ сходится, с естественными операциями

сложения и умножения на число и со скалярным произведением

$$(x, y) = \sum_{n=0}^{\infty} \overline{y_n} x_n.$$

Другой пример (приведем только вещественную версию) – $L_2([a, b])$. Пусть V – множество измеримых функций $f : [a, b] \to \mathbb{R}$ таких, что интеграл Лебега $\int_a^b f(t)^2\, dt$ сходится, с поточечными операциями. Зададим симметричную билинейную форму B на V формулой

$$B(f, g) = \int_a^b f(t)g(t)\, dt.$$

Ясно, что эта форма неотрицательно определена, но вырождена, скажем, функция, отличающаяся от нуля на множестве меры 0 (например, в конечном числе точек), ортогональна всему пространству. Положим $L_2([a, b]) = V/V^{\perp_B}$. Тогда B индуцирует скалярное произведение на этом пространстве. Можно доказать, что набор функций 1, $\sin nt$, $\cos nt$ ($n \in \mathbb{N}$) являются ортогональным базисом плотного подпространства в $L_2([-\pi, \pi])$, поэтому $L_2([a, b])$ является сепарабельным (чтобы перевести $[-\pi, pi]$ в $[a, b]$ достаточно сделать линейную замену переменных).

В анализе базисом гильбертова пространства называют базис плотного подпространства, т. е. базис – это набор v такой, что любой вектор представляется в виде бесконечной линейной комбинации базисных векторов $\sum_{n=0}^{\infty} v_n \alpha_n$ единственным образом (как обычно, сумма ряда – это предел последовательности частичных сумм, а понятие предела в нормированном пространстве существует). Тот базис, который изучается в курсе линейной алгебре, в гильбертовом пространстве, конечно же, также существует. Он называется алгебраическим базисом или базисом Гамеля. Заметим, что счетного базиса Гамеля в гильбертовом пространстве не существует.

Неравенство треугольника, КБШ, и процесс ортогонализации естественно имеют место и в гильбертовом пространстве, так как работают с конечномерными подпространствами. Процесс ортогонализации работает и с бесконечными базисами.

А вот дальше начинаются неожиданности, связанные с наличием незамкнутых линейных подпространств, не непрерывных линейных отображений и инъективных, но не сюръективных эндоморфизмов. В ℓ_2 незамкнутым линейным подпространством является множество ℓ_2^{fin} финитных последовательностей. Оно как раз является счетномерным плотным подпространством, наличие которого доказывает сепарабельность. Базис этого подпространства можно дополнить до алгебраического базиса всего гильбертова пространства, следовательно, $\ell_2 = \ell_2^{fin} \oplus U$ для некоторого подпространства U. Отображение $L : \ell_2 \to F$, заданное равенствами $L(x + y) = \sum_{n=0}^{\infty} n x_n$, где $x \in \ell_2^{fin}$, а $y \in U$, не является непрерывным. Инъективным, но не сюръективным является сдвиг $S(x)_1 = 0$, $S(x)_n = x_{n-1}$. Сдвиг в обратную сторону: $S'(x)_n = x_{n+1}$ является примером сюръективного, но не инъективного оператора в ℓ_2. Заметим, что $S' \circ S = \mathrm{id} \neq S \circ S'$, то есть в алгебре операторов существуют односторонне обратимые элементы.

В анализе под подпространством гильбертова пространства по умолчанию считают *замкнутое* подпространство, а под линейным оператором – непрерывный линейный оператор. С такой оговоркой разложение в прямую сумму подпространства и его ортогонального дополнения и равенство $(U^\perp)^\perp = U$ имеют место и в гильбертовых пространствах (для незамкнутого подпространства $(U^\perp)^\perp$ является замыканием U).

Формула для координат в ортогональном базисе, неравенство Бесселя и равенство Парсеваля имеют место в гильбертовых пространствах и для бесконечных ортогональных наборов. Выполнена также лемма о расстоянии от точки до (замкнутого) подпространства.

Двойственное пространство к гильбертову пространству H – это пространство *непрерывных* линейных функционалов, т. е. непрерывных линейных отображений $H \to F$. С таким соглашением можно доказать, что H^* канонически изоморфно H (это называется теоремой Рисса). Оба определения сопряженного оператора теперь переносятся и на гильбертовы пространства.

Изучение собственных чисел операторов становится существенно сложнее из за наличия инъективных, но не сюръективных операторов. Далее $F = \mathbb{C}$, потому что даже в конечномерных пространствах над $\mathbb{R}$ все доказательства проходят через $\mathbb{C}$. Множество тех чисел $\lambda \in \mathbb{C}$, для которых оператор $L - \lambda\mathrm{id}$ не обратим, называется спектром оператора L. При этом λ называется собственным числом, если (также как и в конечномерном случае) этот оператор не инъективен. Множество собственных чисел называется точечным спектром. Для компактного самосопряженного оператора верен аналог следствия 5.12 (с бесконечной линейной комбинацией проекторов). Если же самосопряженный оператор не компактен, то он может вообще не иметь собственных чисел и теорема о его строении гораздо сложнее.

Некоторые из матричных разложений имеют место и для операторов в гильбертовых пространствах с соответствующими оговорками. Таково, например, полярное разложение.

8. Кватернионы и движения трехмерного пространства

К теории квадратичных и эрмитовых форм примыкает теория конечномерных алгебр с делением. Алгебра A над полем F называется телом или алгеброй с делением, если любой ее ненулевой элемент обратим. Она называется центральной, если F является ее центром, т. е. множеством элементов, коммутирующих со всеми элементами A. Легко доказать, что над замкнутым полем любая такая алгебра совпадает с самим полем. Действительно, для любого элемента $a \in A$ алгебра $F[a]$ изоморфна $F[t]/(f)$ для некоторого многочлена f. Если f приводим, то $F[a]$ содержит делители нуля, что невозможно. В противном случае, так как F замкнуто, то $\deg f = 1$, откуда $a \in F$.

Над полем вещественных чисел существует ровно одна центральная алгебра с делением. Она называется алгеброй кватернионов.

Определение 8.1. Алгебра кватернионов $\mathbb{H}$ – это 4-мерное векторное пространство над $\mathbb{R}$ с базисом $e = (1, \mathbf{i}, \mathbf{j}, \mathbf{k})$ и таблицей умножения

$$\mathbf{i}^2 = \mathbf{j}^2 = \mathbf{k}^2 = -1,\ \mathbf{ij} = -\mathbf{ji} = \mathbf{k},\ \mathbf{jk} = -\mathbf{kj} = \mathbf{i},\ \mathbf{ki} = -\mathbf{ik} = \mathbf{j}.$$

Легко проверить, что эта алгебра ассоциативна (достаточно проверить это на базисных элементах). Кватернион $a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k}$ можно представить как сумму скалярной части a и векторной части $v = b\mathbf{i}+c\mathbf{j}+d\mathbf{k}$. Если отождествить векторную часть с трехмерным вектором, то умножение векторных частей выглядит следующим образом:

$$uv = u \times v - (u, v),$$

где $\times$ обозначает векторное произведение. Кватернион с нулевой скалярной частью называется чистым кватернионом.

Отображение $\bar{\cdot} : \mathbb{H} \to \mathbb{H},\ a + v \to a - v$, называется сопряжением и является антиинволюцией на $\mathbb{H}$. Действительно,

$$(b + u)(a + v) = ba - (u, v) + (u \times v + au + bv),\ \text{а}$$
$$(a - v)(b - u) = ab - (v, u) + (v \times u - au - bv) = \overline{(b + u)(a + v)},$$

а остальные свойства антиинволюции очевидны. Кроме того

$$(a + v)(a - v) = a^2 - vv = a^2 - v \times v + (v, v) = a^2 + (v, v).$$

Для ненулевого кватерниона это число является положительным вещественным числом и квадратный корень из него называется модулем кватерниона. Таким образом, мы видим, что $h\frac{\bar{h}}{|h|^2} = 1$, т. е. $\mathbb{H}$ – алгебра с делением.

Легко видеть, что модуль является гомоморфизмом мультипликативной группы кватернионов в мультипликативную группу положительных вещественных чисел. Действительно,

$$|h|^2|g|^2 = h\bar{h}g\bar{g} = gh\bar{h}\bar{g} = gh\overline{gh} = |gh|^2.$$

Сопоставление кватерниону $a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k}$ матрицы умножения на этот кватернион в стандартном базисе $(1,\mathbf{i},\mathbf{j},\mathbf{k})$

$$\begin{pmatrix} a & -b & -c & -d \\ b & a & -d & c \\ c & d & a & -b \\ d & -c & b & a \end{pmatrix}$$

задает гомоморфизм алгебры кватернионов в алгебру матриц $\mathrm{M}_4(\mathbb{R})$, а определитель этой матрицы – гомоморфизм мультипликативной группы кватернионов в $\mathbb{R}^*$, который является нормой[4] из $\mathbb{H}$ в $\mathbb{R}$. Вычисление показывает, что этот определитель равен $(a^2+b^2+c^2+d^2)^2$, т. е. четвертой степени модуля кватерниона.

Другое матричное представление кватернионов получается при рассмотрении $\mathbb{H}$, как векторного пространства над $\mathbb{C}$ с умножением $h(a+bi)=h(a+b\mathbf{i})$, где $h\in\mathbb{H}$, а $a,b\in\mathbb{R}$. Тогда $\mathbb{H}$ будет 2-мерным векторным пространством над $\mathbb{C}$ с базисом $(1,\mathbf{j})$: любой кватернион единственным образом представляется в виде $a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k}=1\cdot(a+b\mathbf{i})+\mathbf{j}\cdot(c-d\mathbf{i})$. Умножение на кватернион $a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k}$ слева является линейным оператором на этом векторном пространстве, матрица которого равна

$$\begin{pmatrix} a+bi & -c-di \\ c-di & a-bi \end{pmatrix}.$$

Нетрудно проверить, что это отображение задает вложение мультипликативной группы кватернионов в группу $\mathrm{GL}_2(\mathbb{C})$. Определитель указанной матрицы равен $a^2+b^2+c^2+d^2$ и называется редуцированной нормой $\mathrm{nrd}(a+b\mathbf{i}+c\mathbf{j}+d\mathbf{k})$.

Заметим, что кватернионы с модулем 1 отображаются в унитарные матрицы с определителем 1. Это совершенно не удивительно, потому что умножение на кватернион с модулем 1 является изометрией 4-мерного пространства кватернионов, норма кватерниона равна норме столбца его координат в выбранном базисе (относительно стандартного скалярного произведения в $\mathbb{C}^2$), а умножение на матрицу сохраняет нормы столбцов тогда и только тогда, когда она унитарна. Более удивительно, что это отображение биективно, чуть позже мы сформулируем это в виде изоморфизма групп. Заметим, что это не все изометрии 4-мерного вещественного пространства $\mathbb{H}$, а только $\mathbb{C}$-линейные изометрии.

Интересно также рассмотреть операцию сопряжения кватернионом: $c_h(z)=hzh^{-1}$. Эта операция тоже не меняет модуль кватерниона z, а кроме того оставляет инвариантным пространство чистых кватернионов. Это можно доказать вычислением, а можно воспользоваться идеями параграфа 5. Действительно, множество вещественных чисел инвариантно относительно оператора сопряжения любым кватернионом, а подпространство чистых кватернионов является его ортогональным дополнением (относительно евклидова скалярного произведения, ассоциированного с выбранной нормой). По лемме 5.8 оно инвариантно относительно сопряженного оператора. Так как оператор сопряжения является унитарным, то сопряженный с ним равен обратному. Таким образом, подпространство чистых кватернионов инвариантно относительно сопряжения обратным к любому кватерниону.

Рассмотрим сужение c_h на трехмерное пространство чистых кватернионов. Если $h=a+v$ – разложение кватерниона на вещественную часть и чистый кватернион, то $c_h(v)=v$, и нетрудно видеть, что $\langle v\rangle_{\mathbb{R}}$ – собственное подпространство оператора c_h, соответствующее 1. По леммам 5.7 и 5.8 ортогональное дополнение этого подпространства (плоскость в пространстве чистых кватернионов) c_h-инвариантно. Сейчас мы покажем, что любой поворот в этой плоскости реализуется, а отношение $a/|v|$ определяет угол этого поворота.

[4]Понятие нормы в алгебре отличается от понятия нормы в анализе. В алгебре нормой называют определитель матрицы умножения на элемент F-алгебры.

Можно считать, что $|h| = 1$. Тогда $a = \cos\frac{\alpha}{2}$, а $|v| = \sin\frac{\alpha}{2}$ для некоторого угла α. В ортогональном дополнении к v выберем ортогональный базис $e_1 = w$, $e_2 = \frac{v\times w}{|v|}$. Имеем

$$(a+v)w(a-v) = a^2w + a(vw - wv) - vwv = a^2w + 2av\times w - (v\times w)v =$$
$$a^2w + 2av\times w - v\times w\times v = a^2w + 2av\times w - (v,v)w =$$
$$w(\cos^2\frac{\alpha}{2} - \sin^2\frac{\alpha}{2}) + 2e_2\cdot\cos\frac{\alpha}{2}\sin\frac{\alpha}{2} = w\cos\alpha - \mathrm{pr}_v\, w\cdot\cos\alpha + \mathrm{pr}_v\, w + e_2\cdot\sin\alpha = e_1\cos\alpha + e_2\sin\alpha + e_3.$$

Заметим, что третья координата w в этом базисе не изменилась, а длины векторов e_1 и e_2 равны. Следовательно, $(a+v)w(a-v)$ – вектор, полученный поворотом w на угол α вокруг прямой, натянутой на вектор v. Таким образом, сопряжение кватернионом реализует любой поворот трехмерного пространства.

Несколько слов об обозначениях групп изометрий эрмитова пространства. В общем виде, если F – поле с инволюцией, V – конечномерное векторное пространство над F, а B – невырожденная эрмитова форма на V, то группа изометрий формы B называется унитарной группой и обозначается $\mathrm{U}(B)$ или $\mathrm{U}_n(F)$, где $n = \dim V$, если инволюция и форма известны из контекста. Если инволюция тривиальна, то эта группа называется ортогональной группой и обозначается через $\mathrm{O}(B)$ или $\mathrm{O}_n(F)$. Подгруппы этих групп, состоящие из преобразований с определителем 1, называются специальной унитарной и специальной ортогональной группой и обозначаются через SU и SO, соответственно.

Для стандартных скалярных произведений в $\mathbb{C}^n$ и $\mathbb{R}^n$ соответствующие группы обычно обозначаются просто через U_n, SU_n, O_n и SO_n. Как мы видели, определитель унитарной матрицы по модулю равен 1. Следовательно, определитель вещественной ортогональной матрицы равен ±1. Поэтому группа O_n раскладывается в произведение SO_n и циклической группы второго порядка, порожденной диагональной матрицей ±1 по диагонали и определителем -1. В случае нечетного n в качестве такой матрицы подойдет $-E$, и тогда произведение будет прямым.

Собственные числа матрицы из SO_3 по модулю равны 1, одно из них вещественно, а два других комплексно сопряженные (если все 3 вещественны, то одно равно 1, а два других равны между собой). Так как произведение комплексно сопряженных чисел равно 1, то одно из собственных чисел должно быть равно 1, т. е. ортогональное преобразование действует на некоторой прямой тривиально. Ортогональное дополнение этой прямой инвариантно. Легко видеть, что ортогональная матрица 2×2 с определителем 1 – это матрица поворота. Если же определитель преобразования из O_3 равен -1, то композиция его с любым зеркальным отражением будет лежать в SO_3, т. е. будет вращением. Таким образом, мы доказали хорошо известный из геометрии факт.

Предложение 8.2. *Любое движение трехмерного пространства с неподвижной точкой является поворотом или композицией поворота с (любой) зеркальной симметрией.*

Повороты называются собственными движениями, а композиция поворота и симметрии – несобственными. Таким образом, SO_3 – это группа собственных движений трехмерного пространства.

Теперь можно сформулировать связь кватернионов с унитарными и ортогональными группами.

Предложение 8.3. *Мультипликативная группа кватернионов с модулем* 1 *изоморфна группе* SU_2*, а всех кватернионов* $\mathbb{H}^* \cong \mathbb{R}^*_{>0}\times\mathrm{SU}_2$.

Факторгруппа $\mathbb{H}^*/\mathbb{R}^* \cong \mathrm{SU}_2/\{\pm E\} \cong \mathrm{SO}_3$.

Модуль кватерниона задает на $\mathbb{H}^*$ *структуру нормированного топологического пространства, которая индуцирует топологию на* $\mathbb{H}^*/\mathbb{R}^*$*. Ясно, что* $\mathbb{H}^*/\mathbb{R}^*$ *с этой топологией гомеоморфно проективному пространству* $\mathbb{RP}^3$*. Изоморфизм* $\mathbb{H}^*/\mathbb{R}^* \to \mathrm{SO}_3$ *является гомеоморфизмом и, следовательно,* SO_3 *гомеоморфно* $\mathbb{RP}^3$.

9. Теоремы Витта

Этот параграф посвящен основам классификации квадратичных пространств над произвольным полем F характеристики не 2. Мы увидим, что любая квадратичная форма является суммой нескольких экземпляров гиперболической плоскости и анизотропной формы. Начнем с определения суммы форм и гиперболической плоскости.

Будем говорить, что две формы изоморфны, если соответствующие квадратичные пространства изометричны. Пусть (V, Q) и (V', Q') – квадратичные пространства. Обозначим через $Q \oplus Q'$ квадратичную форму на $V \oplus V'$, действующую по правилу $Q \oplus Q'(x, x') = Q(x) + Q'(x')$ (проверьте, что это действительно квадратичная форма). Если подпространства V и V' невырождены, то в пространстве $V \oplus V'$ они являются ортогональными дополнениями друг друга относительно формы $Q \oplus Q'$. Поэтому в текстах про квадратичные формы чаще используют обозначение $Q \perp Q'$ вместо $Q \oplus Q'$ и называют такую форму ортогональной суммой форм Q и Q'.

Обозначим через $\langle \alpha_1 \rangle \oplus \cdots \oplus \langle \alpha_n \rangle = \langle \alpha_1, \ldots, \alpha_n \rangle$ форму, которая в некотором базисе имеет матрицу $\operatorname{diag}(\alpha_1, \ldots, \alpha_n)$. Форма $\mathbb{H} \cong \langle 1, -1 \rangle$ (точнее двумерное пространство с этой формой) называется гиперболической плоскостью. Ненулевой вектор x называется изотропным относительно формы Q, если $Q(x) = 0$. Напомним, что форма Q называется изотропной, если существует Q-изотропный вектор, и анизотропной в противном случае.

Лемма 9.1. *Любая невырожденная изотропная квадратичная форма Q изоморфна $\mathbb{H} \oplus Q'$ для некоторой невырожденной формы Q'.*

Доказательство. Так как Q изотропна, существует $x \in V \smallsetminus \{0\}$ такой, что $Q(x) = 0$. Обозначим через B симметричную билинейную форму, ассоциированную с Q. Поскольку Q невырождена, найдется $y \in V$ такой, что $B(x, y) \neq 0$. Домножая y на подходящий скаляр, можно считать, что $B(x, y) = 2$. Для любого $\lambda \in F$ вектора x и $z = y + \lambda x$ линейно независимы, причем $B(x, z) = B(x, y) + \lambda B(x, x) = 2$. Наконец, $Q(z) = B(y + \lambda x, y + \lambda x) = B(y, y) + 2\lambda B(x, y) + B(x, x) = Q(y) + 4\lambda$. Значит, если положить $\lambda = -Q(y)/4$, получим $Q(z) = 0$. Возьмем теперь $v_1 = (x + z)/2$, а $v_2 = (x - z)/2$. Тогда $B(v_1, v_2) = 0$, а $Q(v_1) = -Q(v_2) = 1$.

Дополнив v_1, v_2 до базиса $(v_1, v_2, u_3, \ldots, u_n)$ при $i \geqslant 3$ положим

$$v_i = u_i - B(v_1, u_i)v_1 + B(v_2, u_i)v_2.$$

Так как $B(v_1, v_i) = B(v_2, v_i) = 0$, то $Q \cong \mathbb{H} \oplus Q'$, где Q' – сужение формы Q на подпространство, порожденное векторами $v_3, \ldots, v_n$. □

Из леммы следует в частности, что все невырожденные изотропные двумерные формы изоморфны $\mathbb{H}$. Большая часть алгебраистов предпочитает выбирать базис $\mathbb{H}$ так, чтобы матрица квадратичной формы была равна $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, а не $\operatorname{diag}(1, -1)$, то есть в координатах: $Q(x) = 2x_1x_2$, а не $Q(x) = x_1^2 - x_2^2$.

Из леммы следует также, что любое квадратичное пространство представляется в виде суммы нескольких экземпляров гиперболической плоскости и анизотропного подпространства. Наша следующая цель – доказать, что анизотропное подпространство определено единственным образом с точностью до изометрии. Для этого мы докажем теорему о сокращении, которая говорит о том, что изометрию $U \oplus V \cong U \oplus W$ можно сокращать на U.

Пусть Q – квадратичная форма на V, соответствующая симметричной билинейной форме B, а x – неизотропный вектор. Определим отражение S_x относительно вектора[5] x формулой

$$S_x(y) = y - 2\frac{B(x, y)}{Q(x)}x.$$

Простое вычисление показывает, что отражение является изометрией, а $S_x^2 = \mathrm{id}$.

[5]Геометрически, это является отражением относительно гиперплоскости, ортогональной x.

Лемма 9.2. *Пусть $x_1, x_2 \in V$ и $Q(x_1) = Q(x_2) \neq 0$. Тогда существует композиция отражений, переводящая x_1 в x_2.*

Доказательство. Если $Q(x_1 - x_2) \neq 0$, то подойдет отражение относительно этого вектора:

$$S_{x_1-x_2}(x_1) = x_1 - 2\frac{B(x_1 - x_2, x_1)}{Q(x_1 - x_2)}(x_1 - x_2) = \\ x_1 - 2\frac{Q(x_1) - B(x_2, x_1)}{Q(x_1) + Q(x_2) - 2B(x_2, x_1)}(x_1 - x_2) = x_1 - (x_1 - x_2) = x_2.$$

Если $Q(x_1 + x_2) \neq 0$, то $S_{x_1+x_2}(x_1) = -x_2$, а $S_{x_2}(-x_2) = x_2$. Если же $Q(x_1 - x_2) = Q(x_1 + x_2) = 0$, то

$$0 = Q(x_1 - x_2) + Q(x_1 + x_2) = Q(x_1) + Q(x_2) + Q(x_1) + Q(x_2) = 4Q(x_1),$$

что противоречит предположению. □

Следствие 9.3. *Любая изометрия невырожденного пространства есть композиция отражений.*

Доказательство. Пусть $L : V \to V$ – изометрия невырожденного квадратичного пространства (V, Q). Доказываем индукцией по $n = \dim V$; база $n = 1$ очевидна. Пусть $n > 1$. Возьмем $x \in V$ такой, что $Q\big(L(x)\big) = Q\big(x\big) \neq 0$. По лемме найдется композиция отражений $S\colon V \to V$ такая, что $S(x) = L(x)$. Отображение $S^{-1}L$, таким образом, является изометрией и оставляет x на месте; значит, $S^{-1}L$ оставляет на месте и $W = x^\perp$ – подпространство размерности $n - 1$. По предположению индукции изометрия $S^{-1}L|_W$ является композицией отражений (относительно векторов из W). Заметим, что любое отражение относительно вектора из W оставляет на месте x, поскольку $x \perp W$. Значит, изометрия $S^{-1}L$ является композицией тех же самых отражений, рассматриваемых как преобразования всего пространства V. Перенося S в другую часть, получаем, что и L является композицией отражений. □

Теорема 9.4 (Витта о сокращении). *Если $Q \oplus Q_1 \cong Q \oplus Q_2$, то $Q_1 \cong Q_2$.*

Доказательство. Докажем сначала утверждение при $Q = \langle\alpha\rangle$. Пусть при $i = 1, 2$ форма $Q'_i = \langle\alpha\rangle \oplus Q_i$ задана на пространстве $W'_i = \langle x_i\rangle \oplus W_i$, причем $Q'_i(x_i) = \alpha$. Пусть $L : W'_1 \to W'_2$ – изометрия. Если $\alpha = 0$, то $(W'_i)^\perp = \langle x_i\rangle \oplus W_i^\perp$ (в этом доказательстве $W_i^\perp$ обозначает ортогональное дополнение к W_i внутри самого W_i). Пусть $W_1 = W_1^\perp \oplus U_1$, а $U_2 = L(U_1)$. Тогда $W'_1 = \langle x_1\rangle \oplus W_1^\perp \oplus U_1 = (W'_1)^\perp \oplus U_1$. Так как $L\big((W'_1)^\perp\big) = (W'_2)^\perp$, то $W'_2 = (W'_2)^\perp \oplus U_2 = \langle x_2\rangle \oplus W_2^\perp \oplus U_2$. Отсюда следует, что $\dim W_1^\perp = \dim W_2^\perp$, и, значит, существует изморфизм векторных пространств $\hat{L} : W_1^\perp \to W_2^\perp$. Тогда $\tilde{L} : W_1 \to W_2$ заданный формулой $\tilde{L}(w + u) = \hat{L}(w) + L(u)$, где $w \in W_1^\perp$, а $u \in U_1$, является требуемой изометрией.

Пусть теперь $\alpha \neq 0$. Так как L – изометрия, то $Q'_2\big(L(x_1)\big) = Q'_1\big(x_1\big) = Q'_2\big(x_2\big) = \alpha \neq 0$. По лемме 9.2 найдется изометрия $S\colon \langle x_2\rangle \oplus W_2 \to \langle x_2\rangle \oplus W_2$ такая, что $S(x_2) = L(x_1)$. Ясно, что $S^{-1}L$ является изометрией $\langle x_1\rangle \oplus W_1 \to \langle x_2\rangle \oplus W_2$ и $S^{-1}L(x_1) = x_2$. Так как изометрия сохраняет ортогональность векторов, то она сохраняет и ортогональные дополнения. Поэтому $S^{-1}L$ отображает $W_1 = x_1^\perp$ на $W_2 = x_2^\perp$. Это означает, что ограничение $S^{-1}L$ на W_1 и дает нужную изометрию между Q_1 и Q_2.

В общем случае по теореме 2.5 $Q \cong \langle\alpha_1\rangle \oplus \cdots \oplus \langle\alpha_n\rangle$, и доказательство проводится очевидной индукцией по n. □

Следствие 9.5. *Любая невырожденная форма Q представляется в виде*

$$Q \cong \underbrace{\mathbb{H} \oplus \cdots \oplus \mathbb{H}}_{r \text{ раз}} \oplus Q_{an},$$

где анизотропная часть Q_{an} *определена однозначно с точностью до изометрии, и* индекс Витта $i(Q) := r$ *определен однозначно.*

Доказательство. По лемме 9.1 если форма изотропна, из нее можно выделить $\mathbb{H}$. Продолжая этот процесс, дойдем до какой-то анизотропной формы (потому что размерность все время убывает). Единственность легко выводится из теоремы о сокращении. □

Теорема 9.6 (Теорема о продолжении изометрии). *Пусть (V, Q) – невырожденное квадратичное пространство, W_1, W_2 – подпространства в V такие, что существует изометрия $L\colon W_1 \to W_2$. Тогда существует изометрия $M\colon V \to V$ такая, что $M|_{W_1} = L$.*

Для доказательства этой теоремы в случае, когда мы продолжаем изометрию с вырожденного подпространства, нам понадобится следующее утверждение.

Лемма 9.7. *Пусть B невырожденная симметричная билинейная форма на V, а W вырожденное подпространство в V. Пусть $w = (w_1, \dots, w_k)$ – базис пространства $W \cap W^\perp$, а $v = (v_1, \dots, v_m)$ – дополнение w до базиса пространства W. Тогда существует набор векторов $u = (u_1, \dots, u_k)$ пространства V таких, что подпространство $U = \langle w \cup u\rangle$ B-ортогонально подпространству $\langle v\rangle$, а матрица сужения B на U в базисе $w \cup u$ равна $\left(\begin{smallmatrix} 0 & E \\ E & 0 \end{smallmatrix}\right)$.*

Заметим, что тогда подпространство $\langle w \cup u \cup v\rangle$ невырождено.

Следствие 9.8. *Размерность максимального изотропного подпространства равна индексу Витту.*

Доказательство теоремы о продложении изометрии. Так как B невырождена, то по лемме 4.7 размерность ортогонального дополнения подпространства W строго меньше, чем размерность ортогонального дополнения подпространства $W' = \langle w_1, \dots, w_{k-1}, v_1, \dots, v_m\rangle$. Поэтому существует вектор $x \in V$, ортогональный всем векторам из набора $w \cup v$, кроме w_k. Таким образом, подпространство, порожденное w_k и x двумерно, изотропно и ортогонально W'. Также как в доказательстве леммы 9.1, найдем в этом подпространстве изотропный вектор u_k такой, что $B(u_k, w_k) = 1$.

Теперь в пространстве $\overline{W} = W + \langle u_k\rangle$ имеем: $\overline{W} \cap \overline{W}^\perp = \langle w_1, \dots, w_{k-1}\rangle$. Индукцией по k найдем изотропные вектора $u_1, \dots, u_k$ такие, что u_i ортогонален w_j при всех $j \neq i$, всем u_j и всем v_j. При этом $B(u_i, w_i) = 1$, а это и означает, что u – искомый набор.

По лемме 2.2 подпространство $\langle v\rangle$ невырождено. Подпространство $\langle w \cup u\rangle$ невырождено, так как матрица сужения формы B на него невырождена. Наконец, ортогональная сумма двух невырожденных подпространств невырождена. □

Доказательство. В случае, когда W_1 невырождено, утверждение следует из теоремы Витта о сокращении. Действительно, в этом случае по предложению 4.7 V раскладывается в прямую сумму W_i и его ортогонального дополнения ($i = 1, 2$). По теореме о сокращении существует изометрия $L' : W_1^\perp \to W_2^\perp$. Тогда изометрия $M : V = W_1 \oplus W_1^\perp \to W_2 \oplus W_2^\perp = V$ задается формулой $M(x + y) = L(x) + L'(y)$ при $x \in W_1$, $y \in W_1^\perp$. Пусть теперь W_1 – вырожденное подпространство в невырожденном пространстве V. Выберем базис $w = (w_1, \dots, w_k)$ пространства $W_1 \cap W_1^\perp$ и дополним его до базиса $w \cup v$ пространства W_1.

Выберем набор $u = (u_1, \dots, u_k)$, удовлетворяющий условиям леммы 9.7. Так как L изометрия, то $L(w)$ является базисом пространства $W_2 \cap W_2^\perp$, а $L(w) \cup L(v)$ – базисом W_2. Выберем набор $u' = (u'_1, \dots, u'_k)$, удовлетворяющий условиям леммы 9.7 с $W = W_2$. Продолжим L на пространство $W_1' = W_1 + \langle u\rangle$, положив $L(u) = u'$ и продолжив по линейности. Так как матрицы сужений формы Q на $\langle w \cup u\rangle$ и $\langle L(w) \cup L(u)\rangle$ в базисах $w \cup u$ и $L(w) \cup L(u)$ совпадают, то сужение L на эти подпространства является изометрией. С другой стороны, сужение L на $\langle v\rangle$ и было изометрией этого пространства на $\langle L(v)\rangle$. Так как $\langle w \cup u\rangle \perp \langle v\rangle$, а $\langle L(w) \cup L(u)\rangle \perp \langle L(v)\rangle$, то определенное выше продолжение L является изометрией $W_1' \to W_2' = W_2 + \langle u'\rangle$. Наконец, по той же самой лемме 9.7 пространство W_1' невырождено, и по первому абзацу доказательства L можно продолжить с него на все пространство V. □

Следствие 9.9. *Индекс Витта равен размерности максимального полностью изотропного подпространства (т. е. подпространства, сужение формы на которое нулевое). Более того, все максимальные изотропные подпространства переводятся друг в друга изометриями.*

10. Симплектические формы

Симплектическая форма – это билинейная антисимметричная форма. В отличие от квадратичных форм, определение симплектической формы не зависит от характеристики поля, надо только правильно определять понятие антисимметричности, см. параграф 1. Некоторая часть теории эрмитовых форм переносится и на симплектические формы,[6] основным же отличием является то, что любой вектор изотропен и, следовательно, нет никаких шансов диагонализовать матрицу симплектической формы. К счастью, любая симплектическая форма приводится к очень простому каноническому виду, причем все симплектические пространства (т. е. пространства с симплектической формой) одной размерности изометричны. Группа линейных преобразований, сохраняющих симплектическую форму, называется симплектической группой. Наряду с полной линейной, унитарной, ортогональной группами и некоторыми их модификациями, симплектическая группа является *классической* группой. Классические группы являются примерами полупростых групп Ли (над $\mathbb{R}$ или $\mathbb{C}$) и алгебраических групп (над произвольным полем), и играют огромную роль во многих отраслях математики от чистой алгебры до механики. В частности, из классических групп над конечными полями строится большая часть простых конечных групп (обычно, факторгруппа коммутанта классической группы по центру проста), а симплектические многообразия (нечто, склеенное из кусков симплектических пространств) и симплектическая группа являются чуть ли не основой гамильтоновой механики.

В этом параграфе мы докажем только классификацию конечномерных невырожденных симплектических пространств. Заметим сначала, что матрица симплектической формы в любом базисе кососимметрическая, т. е. $A^\intercal = -A$. Пусть V – двумерное пространство с ненулевой симплектической формой B. Существует пара векторов v_1, v_2 для которых $B(v_1, v_2) \neq 0$. Домножая v_2 на подходящую константу, можно считать, что $B(v_1, v_2) = 1$. Так как $B(v_1, v_1) = 0$, то отсюда сразу следует линейная независимость v_1 и v_2. В базисе $v = (v_1, v_2)$ матрица формы B имеет вид

$$B_v = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Таким образом, все двумерные невырожденные симплектические пространства изометричны между собой. Такое пространство будем называть симплектической плоскостью.

Теорема 10.1. *Любое невырожденное симплектическое пространство имеет четную размерность и изометрично ортогональной сумме симплектических плоскостей.*

Доказательство. Из сказанного выше следует, что любая пара векторов пространства V, на которых симплектическая форма B не равна нулю, порождают симплектическую плоскость S. Также как в доказательстве леммы 4.7 легко видеть, что $\dim S^{\perp_B} = \dim V - 2$ и $S \cap S^{\perp_B} = \{0\}$. Поэтому $V = S \oplus S^{\perp_B}$ и $(S^{\perp_B})^{\perp_B} = S$. Так как $S^{\perp_B} \cap (S^{\perp_B})^{\perp_B} = \{0\}$, то подпространство $S^{\perp_B}$ невырождено, и можно применить индукцию по размерности пространства V. Так как размерность падает на 2, то база индукции – случай размерности 0 или 1. Первый из них тривиален, а во втором надо только заметить, что любая симплектическая форма на одномерном пространстве нулевая, откуда следует, что невырожденных симплектических форм нечетной размерности не бывает. □

[6] Кусок теории эрмитовых форм переносится на полуторалинейные формы с довольно слабым условием симметрии: $B(x, y) = 0 \iff B(y, x) = 0$.

Глава 8

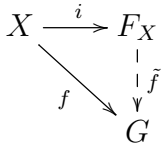
Теория групп

Мы возвращаемся к изучению теории групп и рассмотрим в этой главе несколько базовых конструкций.

1. Свободные группы, задание группы образующими и соотношениями

Универсальное свойство свободной группы аналогично определению свободного модуля.

Определение 1.1. Пусть X – множество. Группа F_X вместе с отображением $i : X \to F_X$ называется *свободной группой с множеством образующих* X (или порожденной X, или свободной группой множества X), если для любой функции f из X в группу G существует единственный гомоморфизм групп $\tilde{f} : F_X \to G$ такой, что диаграмма



коммутативна.

Как обычно, это определение ничего не говорит о существовании универсального объекта. Сейчас мы построим свободную группу и покажем, что она удовлетворяет сформулированному универсальному свойству. Пусть $\bar{X} = \{\bar{x} \mid x \in X\}$ – множество символов. Для $x \in X$ положим $\bar{\bar{x}} = x$.

Рассмотрим множество $W = W_X$, состоящее из всех слов (включая пустое слово) в алфавите $X \cup \bar{X}$. Пусть Q – подмножество в $W \times W$, состоящее из всех пар $(w_1 x \bar{x} w_2, w_1 w_2)$, $w_1, w_2 \in W$, $x \in X \cup \bar{X}$. Обозначим через $\sim$ наименьшее отношение эквивалентности на W, содержащее Q. Другими словами, $u \sim v$ тогда и только тогда, когда u приводится к v при помощи вставки и стирания фрагментов вида $x\bar{x}$.

Пусть $F_X = W/\sim$ – множество классов эквивалентности $\sim$. Определим операцию на $F(X)$, как конкатенацию слов. Точнее, $[w_1]\cdot[w_2] = [w_1 w_2]$, где $w_1, w_2 \in W$, а квадратные скобки означают класс эквивалентности, содержащий данное слово. Нетрудно проверить, что результат операции не зависит от выбора представителей классов эквивалентности. Очевидно, что операция ассоциативна, а нейтральным элементом является класс эквивалентности пустого слова. Обратный к $[x_1 \dots x_n]$ – это элемент $[\bar{x}_n \dots \bar{x}_1]$, где $x_1, \dots, x_n \in X \cup \bar{X}$. Таким образом, F_X – группа.

Теорема 1.2. *Группа F_X вместе с отображением $X \to F_X$, $x \mapsto [x]$ является свободной группой с множеством образующих X.*

Доказательство. Заметим, что по определению умножения в F_X для $x \in X$ имеем $[x]^{-1} = [\bar{x}]$. Для удобства обозначений, допуская вольность речи, будем писать x^{-1} вместо $\bar{x}$. Пусть $f : X \to G$ – функция из X в группу G. Зададим отображение $f' : W \to G$ формулой

$$f'(x_1^{\varepsilon_1} \dots x_n^{\varepsilon_n}) = f(x_1)^{\varepsilon_1} \cdot \dots \cdot f(x_n)^{\varepsilon_n}, \text{ где } x_1, \dots, x_n \in X,\ \varepsilon_1, \dots, \varepsilon_n \in \{\pm 1\}.$$

Так как $f'(w_1 x \bar{x} w_2) = f'(w_1 w_2)$, то отношение $u \sim_{f'} v \iff f'(u) = f'(v)$ на W содержит Q. А так как оно является отношением эквивалентности, то оно содержит и отношение $\sim$ на W, определенное выше. Другими словами, $f'(u) = f'(v)$ при любых $u \sim v$.

Поэтому функция

$$\tilde{f} : F_X \to G, \qquad \tilde{f}([w]) = f'(w)$$

задана корректно. Теперь очевидно, что $\tilde{f}$ является гомоморфизмом, причем $\tilde{f}([x]) = f(x)$ для любого $x \in X$, что равносильно коммутативности диаграммы. Так как любой гомоморфизм $F_X \to G$, делающий диаграмму коммутативной, должен отображать $[x]$ в $f(x)$, а множество $\{[x] \mid x \in X\}$ порождает F_X, то это отображение должно совпадать с $\tilde{f}$. □

Обычно элементы множества X отождествляют с их образами в F_X, т. е. с классами эквивалентности однобуквенных слов. Слово из W называется редуцированным, если оно не содержит вхождений $x\bar{x}$ при $x \in X \cup \bar{X}$. Нетрудно выбрать редуцированного представителя в каждом классе эквивалентности. [1]

Предложение 1.3. *В каждом классе эквивалентности $W/\sim$ есть ровно одно редуцированное слово. Оно имеет наименьшую длину среди всех слов из этого класса.*

Доказательство. Возьмем произвольное слово из фиксированного класса эквивалентности. Если оно не редуцированное, то выкинув из него вхождение $x\bar{x}$ получим более короткое слово из того же класса. Доказательство существования заканчивает индукция по длине слова.

Доказательству единственности мешает неоднозначность алгоритма удаления вхождений $x\bar{x}$ из нередуцированного слова. Значит, для однозначности надо придумать детерминированный алгоритм, который является проекцией на множество редуцированных слов, т. е. редуцированное слово переводит в себя, а нередуцированное – в редуцированное. Кроме того, надо доказать, что применение этого алгоритма к эквивалентным словам возвращает одно и то же. Так как эквивалентные слова связаны цепочкой вставок и удалений фрагментов $x\bar{x}$, то достаточно доказывать, что алгоритм возвращает одинаковые значения на словах w_1w_2 и $w_1x\bar{x}w_2$, где $w_1, w_2 \in W$, а $x \in X \cup \bar{X}$.[2] Определим функцию $f : W \times W \to W \times W$ следующими равенствами:

$$\begin{aligned} f(w, \varnothing) &= (w, \varnothing); \\ f(\varnothing, xw) &= (x, w); \\ f(vy, xw) &= (vyx, w), \text{ если } y \neq \bar{x}; \\ f(v\bar{x}, xw) &= (v, w), \end{aligned}$$

где $v, w \in W$, а $x, y \in X \cup \bar{X}$. Так как длина второго слова под действием f убывает ровно на 1, пока оно не пустое, то $f^k(\varnothing, w) = (w', \varnothing)$ для достаточно большого k (здесь f^k обозначает композицию f с собой k раз). В этом случае положим $g(w) = w'$. Это и есть наш алгоритм, он называется W-процессом. Легко проверить, что если v редуцированное слово, то первое слово в $f(v, w)$ также является редуцированным. Поэтому g всегда возвращает редуцированное слово. Также очевидно, что редуцированное слово отображается в себя под действием g, потому что последнее правило из определения f ни разу не применяется.

Осталось доказать, что $g(w_1w_2) = g(w_1x\bar{x}w_2)$. Ясно, что

$$f^k(\varnothing, w_1w_2) = (g(w_1), w_2) \text{ и } f^k(\varnothing, w_1x\bar{x}w_2) = (g(w_1), x\bar{x}w_2),$$

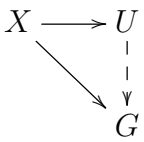
где k – длина слова w_1. Далее, если $g(w_1) = \varnothing$ или $g(w_1) = vy$ при $y \neq \bar{x}$, то $f^2(g(w_1), x\bar{x}w_2) = f(g(w_1)x, \bar{x}w_2) = (g(w_1), w_2)$. В случае $g(w_1) = v\bar{x}$, $f^2(v\bar{x}, x\bar{x}w_2) = f(v, \bar{x}w_2) = (g(w_1), w_2)$. Итак, в любом случае $f^{k+2}(\varnothing, w_1x\bar{x}w_2) = f^k(\varnothing, w_1w_2)$, следовательно, $g(w_1w_2) = g(w_1x\bar{x}w_2)$. □

[1] Следующее утверждение – из серии "очевидно, потому что очевидно". Доказывать такие утверждения бывает, однако, нелегко, а главное, очень противно. Зачем доказывать то, что и так понятно? К сожалению уровень абстракции современной математики таков, что некоторые утверждения из этой серии оказываются неверными. Почти любой математик испытал это на себе, найдя контрпример к "очевидному" утверждению в начале своего доказательства какой-нибудь классной теоремы.

[2] Другими словами, если $g(v) = g(w)$ при всех $(v, w) \in Q$, то эквивалентность $v \overset{g}{\sim} w \iff g(v) = g(w)$ содержит эквивалентность $\sim$. Это соображение мы уже использовали выше.

Соотношением в группе называется элемент w свободной группы на некотором множестве X или, в другой системе обозначений, равенство $w = 1$, Если соотношение выполняется в группе для всех элементов (как, например, $x^{-1}y^{-1}xy = 1$ в абелевой группе), то оно называется групповым тождеством. Мы же сейчас рассмотрим соотношения, которые выполнены для конкретных элементов. Пусть R – подмножество свободной группы на множестве X, а G группа. Говорят, что в G выполняются соотношения R, если задана функция $f : X \to G$, и при подстановке в элементы из R вместо элементов множества X их образов в G получается единица. Более строго, по универсальному свойству свободной группы f индуцирует единственный гомоморфизм $\tilde{f} : F_X \to G$ такой, что $f = \tilde{f} \circ i$, где i – вложение X в F_X. Говорят, что в G выполняются соотношения R, если $R \subseteq \operatorname{Ker} \tilde{f}$.[3]

Универсальная группа, удовлетворяющая соотношениям R, называется группой, заданной образующими X и соотношениями R, и обозначается через $\langle X \mid R\rangle$. Точнее, группа $U = \langle X \mid R\rangle$ вместе с функцией $X \to U$ называется группой заданной образующими X и соотношениями R, если в ней выполнены соотношения R, и для любой группы G с соотношениями R существует единственный гомоморфизм $U \to G$, для которого диаграмма

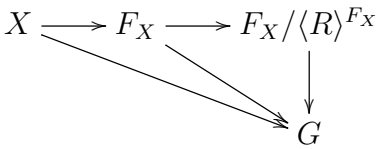


коммутативна.

Пусть $P \leqslant Q$ – группы, а S – подмножество в Q. Напомним, что подгруппа, порожденная S, – это наименьшая подгруппа $\langle S\rangle$ в Q, содержащая S. Наименьшая нормальная подгруппа группы Q, содержащая подгруппу P называется нормальным замыканием P в Q и обозначается P^Q. Нетрудно видеть, что она порождена всеми элементами вида p^q, $p \in P$, $q \in Q$.

Теорема 1.4. *$\langle X \mid R\rangle \cong F_X/\langle R\rangle^{F_X}$. Отображение из X в эту группу – это композиция канонического отображения $X \to F_X$ с канонической проекцией $F_X \to F_X/\langle R\rangle^{F_X}$.*

Доказательство. Пусть G – группа с соотношениями R. Функция $X \to G$ индуцирует единственный гомоморфизм $F_X \to G$. По определению R лежит в ядре этого гомоморфизма. Так как ядро – это нормальная подгруппа, то подгруппа $\langle R\rangle$ и ее нормальное замыкание также лежат в ядре. По универсальному свойству факторгруппы (теорема 5.3 главы 3) существует единственный гомоморфизм $F_X/\langle R\rangle^{F_X} \to G$, делающий следующую диаграмму коммутативной:



Таким образом, группа $F_X/\langle R\rangle^{F_X}$ вместе с указанным отображением из множества X удовлетворяет определению группы, заданной образующими X и сотношениями R. □

Примеры групп, заданных образующими и соотношениями.

(1) $\mathbb{Z} \times \mathbb{Z} \cong \langle x, y \mid [x, y]\rangle = \langle x, y \mid [x, y] = 1\rangle = \langle x, y \mid xy = yx\rangle$.
(2) $D_n \cong \langle x, y \mid x^2 = y^2 = (xy)^n = 1\rangle$.
(3) $S_n \cong \langle \sigma_1, \ldots, \sigma_{n-1} \mid \sigma_i^2 = (\sigma_i\sigma_{i+1})^3 = 1,\ \sigma_i\sigma_j = \sigma_j\sigma_i$ при $|i - j| > 1\rangle$.
(4) $A_n \cong \langle \sigma_3, \ldots, \sigma_n \mid \sigma_i^3 = (\sigma_i\sigma_j)^2 = 1\rangle$

[3]Второе определение, в отличие от первого, не апеллирует к конструкции свободной группы, а только к ее универсальному свойству. Также как и для многочленов, подстановка значений вместо переменных – это образ элемента свободной группы, являющегося аналогом многочлена, под действием некоторого канонического гомоморфизма.

2. Подгруппы свободной группы

Теорема 2.1 (Нильсен, Шрайер)**.** *Любая подгруппа свободной группы свободна.*

Замечание 2.2. При этом количество образующих подгруппы свободной группы с двумя образующими может быть любым от 1 до (счетной) бесконечности.

Существует несколько доказательств теоремы Нильсена–Шрайера. Мы приведем идеи двух геометрических доказательств и полностью изучим алгебраическое доказательство Шрайера, в котором явно строятся свободные образующие.

Топологическое доказательство теоремы Нильсена—Шрайера. Рассмотрим свободную группу F со свободной системой образующих X. Пусть, далее, Y — букет окружностей, занумерованных элементами множества X. Тогда F является фундаментальной группой Y. Каждая подгруппа $H \leqslant F$ является фундаментальной группой какого-то накрытия пространства Y. Однако каждое накрытие Y само гомотопно букету окружностей. Но это и значит, что H свободна. □

Комбинаторно-геометрическое доказательство, приведенное в книге J.-P. Serre "Trees", сразу следует из теоремы о свободном действии группы на дереве. Представим граф, как множество вершин V, множество ребер E, и функции $s, t : E \to V$, сопоставляющие ребру его начало (source) и конец (target). Говорят, что группа G действует на графе (V, E, s, t), если задан гомоморфизм из G в группу автоморфизмов графа. Другими словами, задано действие G на множествах V и E, удовлетворяющие условиям:

$$gs(v) = s(gv) \text{ и } gt(v) = t(gv) \text{ для любых } g \in G \text{ и } v \in V.$$

Граф называется неориентированным, если задан автоморфизм $\bar{\cdot}$ этого графа порядка 2 такой, что $s(\bar{v}) = t(v)$ для любого ребра $v \in V$ (в этом случае ребра v и $\bar{v}$ рассматриваются как одно неориентированное ребро). Будем говорить, что группа G действует свободно на неориентированном графе $(V, E, s, t, \bar{\cdot})$, если она действует свободно на множестве вершин и $gv \neq \bar{v}$ для всех $g \in G$ и $v \in V$. Неориентированный граф без циклов называется деревом.

Графом Кэли группы G относительно системы образующих S называется неориентированный граф, вершинами которого являются элементы группы, а неориентированные ребра соответствуют парам (g, gs) по всем $g \in G$ и $s \in S$. Группа очевидным образом действует на своем графе Кэли: действие на вершинах – это левое умножение, при этом ребро (g, gs) переходит в ребро (hg, hgs) под действием элемента h. Заметим, что это действие свободно тогда и только тогда, когда S не содержит инволюций (т. е. элементов порядка 2).

Лемма 2.3. *Граф Кэли группы G является деревом, на котором группа действует свободно, тогда и только тогда, когда S свободно порождает G.*

Доказательство. Пусть $s_1, \ldots, s_n \in S \cup S^{-1}$, а $w = s_1 \ldots s_n \in F_S$ – редуцированное слово. Канонический образ w в G равен e_G тогда и только тогда, когда ребра, соединяющие вершины $e_G, s_1, s_1 s_2, \ldots, s_1 \ldots s_n = e_G$ образуют цикл. Условие свободы действия необходимо для того, чтобы исключить образующие порядка 2, которые по нашему соглашению образуют не цикл, а определяют неориентированное ребро. □

Таким образом, свободная группа свободно действует на некотором дереве. Оказывается, верно и обратное.

Теорема 2.4. *Группа свободно действует на дереве тогда и только тогда, когда она свободная.*

Отсюда сразу следует теорема Нильсена–Шрайера, потому что любая подгруппа свободной группы свободно действует на том же дереве, на котором свободно действует вся группа.

Третье доказательство теоремы Нильсена–Шрайера основано на теореме Шрайера об образующих подгруппы и на выборе "хорошей" системы представителей свободной группы по ее подгруппе. Пусть $X \subseteq G$ — порождающее множество группы G, $H \leqslant G$, а Y — система представителей правых смежных классов G по H, т. е. $G = HY$, причем, если $Hy_1 = Hy_2$ для $y_1, y_2 \in Y$, то $y_1 = y_2$. Будем считать, что $H \cap Y = \{1\}$. Рассмотрим проекцию $G \to Y$, $g \mapsto \overline{g}$, которая каждому $g \in G$ сопоставляет тот единственный $\overline{g} \in Y$, для которого $H\overline{g} = Hg$. По нашему соглашению относительно представителя H для любого $h \in H$ имеем $\overline{h} = 1$. Положим

$$Z(X, Y) = \{yx \cdot (\overline{yx})^{-1} \mid y \in Y,\ x \in X\}.$$

(если Y или оба множества Y, X известны из контекста, будем писать $Z(X)$ или просто Z вместо $Z(X, Y)$).

Лемма 2.5. *Элемент $yx\cdot(\overline{yx})^{-1}$ всегда лежит в H. Он равен 1 тогда и только тогда, когда $yx \in Y$.*

Пусть $\tilde{y} = \overline{yx}$. Тогда $\left(yx \cdot (\overline{yx})^{-1}\right)^{-1} = \tilde{y}x^{-1}(\overline{\tilde{y}x^{-1}})^{-1} \in Z(X^{-1})$. Следовательно, $Z(X)^{-1} = Z(X^{-1})$.

Доказательство. Первое предложение леммы вытекает из равенства $Hyx = H\overline{yx}$, второе – из того, что $yx \in Y \iff xy = \overline{yx}$.

Далее, заметим, что для любых $g_1, g_2 \in G$ из $H\overline{g_1} = Hg_1$ следует, что $H\overline{g_1}g_2 = Hg_1g_2$, откуда $\overline{\overline{g_1}g_2} = \overline{g_1g_2}$. Таким образом, $\overline{\tilde{y}x^{-1}} = \overline{yxx^{-1}} = y$, откуда сразу получаем следующую формулу из леммы. Из этой формулы сразу вытекает включение $Z(X)^{-1} \subseteq Z(X^{-1})$, а обратное включение получается заменой X на X^{-1}. □

Теорема 2.6 (Шрайер). *Пусть X – система образующих группы G, $H \leqslant G$, а Y – система представителей правых смежных классов G по H, содержащая 1. Тогда подгруппа H порождается множеством $Z = Z(X, Y)$.*

Доказательство. Запишем элемент $h \in H$ в виде произведения образующих: $h = x_1 \dots x_m$, где $x_1, \dots, x_m \in X \cup X^{-1}$, и воспользуемся "трюком накопления". Так как $\overline{x_1 \dots x_m} = \overline{h} = 1$, то можно переписать выражение для h в виде

$$\left(1 \cdot x_1(\overline{1 \cdot x_1})^{-1}\right)\left(\overline{x_1}x_2(\overline{x_1x_2})^{-1}\right)\left(\overline{x_1x_2}x_3(\overline{x_1x_2x_3})^{-1}\right) \dots \left(\overline{x_1 \dots x_{m-1}}x_m(\overline{x_1 \dots x_m})^{-1}\right).$$

По предыдущей лемме все сомножители правой части лежат в $Z(X) \cup Z(X^{-1}) = Z \cup Z^{-1}$. □

Следствие 2.7. *Подгруппа конечного индекса в конечнопорожденной группе конечно порождена.*

Следующее утверждение, также принадлежащее Шрайеру, говорит о выборе "хорошего" множества представителей смежных классов по подгруппе H свободной группы F_X. Пусть $U \subseteq F_X$, а $g \in F_X$. Обозначим через $\ell(g)$ длину редуцированного слова в алфавите $X \cup X^{-1}$ равного g и положим $\ell(U) = \min\limits_{g \in U} \ell(g)$. Система представителей Y смежных классов F_X по H называется минимальной, если для любого $g \in F_X$ длина представителя $\overline{g}$ смежного класса Hg равна $\ell(Hg)$. Система Y называется шрайеровской трансверсалью, если любой начальный отрезок редуцированного слова из Y принадлежит Y (в частности, пустое слово принадлежит Y). Ясно, что достаточно требовать, чтобы для каждого редуцированного слова $x_1 \dots x_n \in Y$ слово $x_1 \dots x_{n-1}$ также лежало бы в Y (т. е. брать отрезки на 1 меньшей длины).

Лемма 2.8. *Для любой подгруппы $H \leqslant F_X$ существует минимальная шрайеровская трансверсаль Y.*

Доказательство. Строим представитель $Y \cap Hg$ индукцией по $n = \ell(Hg)$. При $n = 0$ выбираем нейтральный элемент (пустое слово). Пусть теперь $n > 0$, а $g' = x_1 \dots x_n$, $x_i \in X \cup X^{-1}$, – слово наименьшей длины в Hg. Пусть $u = x_1 \dots x_{n-1}$ — начальный отрезок этого слова. Так как $\ell(Hu) < n$, то по индукционному предположению мы уже выбрали представителя Hu в Y. Пусть,

скажем, $H\overline{u} = Hu$, где $\overline{u}$ – такой представитель, $\ell(\overline{u}) = m \leqslant n-1$, а $\overline{u} = t_1 \dots t_m$, $t_i \in X \cup X^{-1}$, – его приведенное разложение. Тогда $H\overline{u}x_n = Hux_n = Hg' = Hg$. Так как g' – элемент наименьшей длины в своем смежном классе, то $m+1 \geqslant \ell(\overline{u}x_n) \geqslant \ell(g') = n$. Таким образом, $m = n-1$ и, значит, $y = \overline{u}x_n = t_1 \dots t_{n-1}x_n$ является приведенным разложением y. Выберем y в качестве представителя смежного класса Hg. Очевидно, что для него выполняется условие, фигурирующее в определении минимальной шрайеровской системы. □

Пусть H произвольная группа, а Z – произвольная система образующих H такая, что $Z \cap Z^{-1} = \varnothing$. Для того чтобы доказать, что $H \cong F_Z$, достаточно проверить, что непустое редуцированное слово в алфавите $Z \cup Z^{-1}$ не равно 1 в H. Действительно, по универсальному свойству свободной группы существует единственный гомоморфизм $\varphi : F_Z \to H$, отображающий однобуквенные слова в соответствующие элементы множества Z. Так как система образующих группы H лежит в образе φ, то φ сюръективно. Ядро φ – это множество редуцированных слов (точнее, их классов эквивалентности), которые отображаются в 1_H, а приведенное выше условие как раз и говорит, что таких непустых слов не существует. Это простое соображение вместе с двумя предыдущими леммами лежит в основе нашего доказательства теоремы Нильсена–Шрайера.

Доказательство теоремы Нильсена–Шрайера. Пусть H – подгруппа свободной группы F_X, Y – шрайеровская трансверсаль к H (не обязательно минимальная), существование которой обеспечивает предыдущая лемма, а $\bar{\cdot} : F_X \to Y$ – проекция из теоремы 2.6. Так же как в теореме 2.6, возьмем систему образующих

$$Z = Z(X) = \{yx(\overline{yx})^{-1} \mid y \in Y,\ x \in X\} \smallsetminus \{1\}$$

группы H (естественно, единичный элемент можно выбросить из системы образующих) и докажем, что множество Z свободно порождает H. Начнем с доказательства того, что естественное представление элемента из $Z(X \cup X^{-1})$ несократимо. Пусть $x \in X \cup X^{-1}$, а $y \in Y$, причем $yx \notin Y$ (иначе $yx(\overline{yx})^{-1} = 1$). Пусть $y = p_1 \dots p_n$ и $\tilde{y} = \overline{yx} = q_1 \dots q_m$ представлены редуцированными словами, где $p_i, q_i \in X \cup X^{-1}$. Если $x = p_n^{-1}$, то $yx = p_1 \dots p_{n-1} \in Y$, так как система представителей шрайеровская, а если $x = q_m$, то $\tilde{y}x^{-1} = q_1 \dots q_{m-1} \in Y$, по той же причине. Но тогда $\tilde{y}x^{-1} = \overline{\tilde{y}x^{-1}} = \overline{yxx^{-1}} = y$, откуда опять $yx = \tilde{y} \in Y$, что противоречит предположению. В других же местах сокращения произойти не могут.

Далее рассмотрим произведение двух элементов $z = yx(\overline{yx})^{-1}$ и $z' = uv(\overline{uv})^{-1}$ из $Z \cup Z^{-1} = Z(X \cup X^{-1})$ (т. е. $x, v \in X \cup X^{-1}$, а $y, u \in Y$). Положим, $\tilde{u} = \overline{uv}$ и предположим, что $z' \neq z^{-1}$. Докажем, что при этом условии

$$v\tilde{u}^{-1}yx = vwx, \qquad \text{т. е.} \qquad z'z = uvwx(\overline{yx})^{-1},$$

где w и vwx несократимые слова (при этом w может быть пустым). Пусть $\tilde{u} = q_1 \dots q_k$ и $y = p_1 \dots p_n$ – редуцированные слова (здесь $q_i, p_i \in X \cup X^{-1}$). Так как z и z' не равны 1, то их естественные представления несократимы, т. е. $v \neq q_k$, а $x \neq p_n^{-1}$. Для определенности предположим, что $k \geqslant n$ (ситуация $n < k$ аналогична случаю $n > k$). Если $y \neq q_1 \dots q_n$, то $v\tilde{u}^{-1}yx = vwx$, где w – несократимое слово начинающееся с q_k^{-1} и заканчивающееся на p_n. Следовательно, слово vwx также редуцировано.

Пусть теперь $y = q_1 \dots q_n$. Если $k = n$, то $v\tilde{u}^{-1}yx = vx$. Если $v = x^{-1}$, то $(\overline{yx})^{-1} = (\overline{\tilde{u}v^{-1}})^{-1} = u^{-1}$, следовательно, $z' = z^{-1}$, что противоречит предположению. В противном случае возьмем в качестве w пустое слово. Наконец, если $k > n$, то $v\tilde{u}^{-1}yx = vq_k^{-1} \dots q_{n+1}^{-1}x$. Если $x \neq q_{n+1}$, то последнее слово несократимо. В противном случае $yx = q_1 \dots q_{n+1}$, что лежит в Y, так как это начало слова $\tilde{u} \in Y$, а Y шрайеровская. Но это противоречит предположению о том, что $z \neq 1$.

Из доказательства следует, что если $z' = z^{-1}$, то $v = x^{-1}$, а это невозможно при $v, x \in X$. Поэтому пересечение $Z(X) \cap Z(X)^{-1}$ пусто.

Доказательство заканчивает индукция по длине произведения элементов из $Z \cup Z^{-1}$. Пусть $z_i = y_i x_i (\overline{y_i x_i})^{-1} \in Z \cup Z^{-1}$, где $y_i \in Y$, а $x_i \in X \cup X^{-1}$. Индукцией по m докажем, что редуцированная форма элемента $z_1 \dots z_m$ оканчивается на $x_m(\overline{y_m x_m})^{-1}$ и, следовательно, это произведение не равно

1. База $m = 1$ уже доказана там, где говорится, что $yx(\overline{yx})^{-1}$ либо равно 1, либо несократимо (естественно, везде предполагается, что $(\overline{yx})^{-1}$ записано редуцированным словом).

По индукционному предположению $z_1 \dots z_{m-1} = tx_{m-1}(\overline{y_{m-1}x_{m-1}})^{-1}$. По доказанному ранее $x_{m-1}(\overline{y_{m-1}x_{m-1}})^{-1}y_m x_m = x_{m-1}wx_m$, причем последняя запись несократима. Следовательно,

$$z_1 \dots z_m = tx_{m-1}(\overline{y_{m-1}x_{m-1}})^{-1}y_m x_m(\overline{y_m x_m})^{-1} = tx_{m-1}wx_m(\overline{y_m x_m})^{-1}.$$

□

Рассмотрим следующий пример. Пусть $X = \{t, x\}$ и $F = F_X$ – свободная группа с двумя образующими. Построим множество шрайеровских образующих коммутанта группы F. Ясно, что факторгруппа по коммутанту – это свободная абелева группа (т. е. свободный $\mathbb{Z}$-модуль) с двумя образующими. Короче, $F/[F, F] \cong \mathbb{Z} \oplus \mathbb{Z}$. Естественный представитель класса эквивалентности, отображающегося в (n, k) – элемент $t^n x^k$. Ясно, что трансверсаль $Y = \{t^n x^k \mid n, k \in \mathbb{Z}\}$ является шрайеровской. Далее, $(t^n x^k)x \in Y$, поэтому из таких произведений не возникают элементы шрайеровской системы образующих. Зато $(t^n x^k)t \notin Y \iff k \neq 0$, и $\overline{(t^n x^k)t} = t^{n+1}x^k$. Таким образом,

$$(t^n x^k)t(t^{n+1}x^k)^{-1} = t^n x^k t x^{-k} t^{-n-1} \in Z$$

(заметим, что последний элемент является коммутатором $[t^n x^k t^{-n}, t]$). Легко видеть, что все такие слова являются редуцированными и, следовательно, различны в F. По доказательству теоремы Нильсена–Шрайера получаем, что подгруппа $[F, F]$ свободно порождена множеством $Z = \{t^n x^k t x^{-k} t^{-n-1} \mid k, n \in \mathbb{Z},\ k \neq 0\}$ и является свободной группой со счетным числом образующих.

3. Действие группы на множестве и лемма Бернсайда

Определение 3.1. Пусть G – группа, а X – множество. Будем говорить, что G действует на X и писать $G \curvearrowright X$, если задана операция $G \times X \to X$, $(g, x) \mapsto gx$, обладающая для любого $x \in X$ и $g, h \in G$ следующими свойствами:

(1) $g(hx) = (gh)x$ (внешняя ассоциативность);
(2) $1 \cdot x = x$ (унитальность).

Напомним, что для множества X множество всех биективных функций $X \to X$ с операцией композиции называется *симметрической группой* на множестве X и обозначается через S_X. Заметим, что любой гомоморфизм $\theta : G \to S_X$ задает действие группы G на множестве X по правилу $gx = \theta(g)(x)$ (проверьте, что эта операция действительно удовлетворяет условиям определения 3.1). Обратно, если задано действие G на X, то можно задать гомоморфизм $\theta : G \to S_X$ формулой $\theta(g)(x) = gx$ (проверьте, что $\theta(g)$ – биекция, и что θ – гомоморфизм). Таким образом, можно считать, что действие группы на множестве – это гомоморфизм $G \to S_X$, что является равносильным определением действия группы на множестве.

На самом деле мы определили левое действие $G \curvearrowright X$. Правое действие $X \curvearrowleft G$ определяется аналогично.

Определение 3.2. Будем говорить, что G действует справа на X и писать $X \curvearrowleft G$, если задана операция $X \times G \to X$, $(x, g) \mapsto xg$, обладающая для любого $x \in X$ и $g, h \in G$ следующими свойствами:

(1) $(xg)h = x(gh)$;
(2) $x \cdot 1 = x$.

Правому действию соответствует антигомоморфизм $\eta_X : G \to S_X$. Из правового действия довольно просто сделать левое действие, взяв композицию

$$G \overset{inv}{\to} G \overset{\eta_X}{\to} S_X, \text{ где } inv(g) = g^{-1}.$$

Пример. Пусть X, Y – множества, G – группа, а Y^X – множество функций из X в Y. Если $G \curvearrowright X$, то $Y^X \curvearrowleft G$ по правилу: $fg = f \circ \theta_X(g)$. Но тогда формула $gf = f \circ \theta_X(g^{-1})$ задает левое действие. Другими словами, $(gf)(x) := f(g^{-1}x)$.

Введем теперь некоторые понятия, связанные с действием группы G на множестве X.

Определение 3.3. *Орбитой* элемента $x \in X$ под действием G называется множество $Gx = \{gx \mid g \in G\}$. Количество элементов в данной орбите называется *длиной орбиты* (в разных орбитах может быть разное количество элементов).

Лемма 3.4. *Любые две орбиты либо не пересекаются, либо совпадают. Таким образом, множество X разбивается в дизъюнктное объединение орбит.*

Доказательство этого утверждения практически совпадает с доказательством аналогичного утверждения для смежных классов.

Определение 3.5. Неподвижными точками элемента $g \in G$ называются те $x \in X$, для которых $gx = x$. *Множество неподвижных точек* элемента g мы будем обозначать через $\mathrm{Fix}_X(g)$.

Определение 3.6. Множество элементов группы G, оставляющих на месте данный элемент $x \in X$ называется *стабилизатором* элемента x и обозначается через G_x. Другими словами, $G_x = \{g \in G \mid gx = x\}$. Очевидно, что стабилизатор является подгруппой в G.

Замечание 3.7. Обратите внимание на то, что количество пар $(g, x) \in G \times X$, для которых $gx = x$ можно вычислить двумя способами, которые указаны в разных частях следующего равенства:

$$\sum_{x \in X} |G_x| = \sum_{g \in G} |\mathrm{Fix}_X(g)|.$$

Последнее равенство, несмотря на свою очевидность, играет важную роль при доказательстве важного комбинаторного приложения теории групп, леммы Бернсайда. Второе ключевое соображение приведено в следующей лемме. Здесь G/G_x обозначает *множество* левых смежных классов (оно не обязано быть подгруппой, потому что G_x, вообще говоря, не является нормальной подгруппой).

Лемма 3.8. *Пусть $x \in X$. Отображение $f : G/G_x \to Gx$, заданное формулой $f\big(gG_x\big) = gx$, является биекцией. В частности, длина орбиты элемента x равна индексу стабилизатора этого элемента: $|Gx| = |G : G_x|$.*

Доказательство. Очевидно, $gx = gg'x$ для любого $g' \in G_x$, поэтому f задана корректно (определение не зависит от выбора представителя смежного класса). Сюръективность f сразу следует из определения орбиты. Предположим, что $f\big(gG_x\big) = f\big(hG_x\big)$, т. е. $gx = hx$. Но тогда $h^{-1}gx = x$, откуда $h^{-1}g \in G_x$, а из этого сразу следует, что $gG_x = hG_x$. □

Определение 3.9. Пусть $G \curvearrowright X$.

- Действие называется точным, если $\mathrm{Ker}(\theta_X) = \{1\}$, другими словами, если из того что $\forall x \in X : gx = x$ следует, что $g = 1$.
- Действие называется свободным, если $gx = x \implies g = 1$, другими словами, если $\forall x \in X$: $G_x = \{1\}$.
- Действие называется транзитивным, если $\forall x, y \in X\ \exists g \in G : gx = y$, другими словами, $\forall x \in X : Gx = X$ (квантор не имеет значения, равносильно можно написать $\exists x \in X : Gx = X$).

Примеры.

- $S_n \curvearrowright \{1, \dots, n\}$.
- $\mathrm{GL}_n(R) \curvearrowright R^n$.

- $G \curvearrowright G$, gx – умножение в группе (регулярное действие или действие левыми трансляциями) Оно является свободным и транзитивным.
- $G \curvearrowright G$, $(g,x) \mapsto {}^g x = gxg^{-1}$ – действие левым сопряжением.
  $\operatorname{Ker}(G \to S_G) = C(G)$ – центр группы G. Орбита называется классом сопряженных элементов.
- $G \curvearrowright G$, $(x,g) \mapsto x^g = g^{-1}xg$ – действие правым сопряжением. Это правое действие.
- $G \times G \curvearrowright G$: $(g,h)x := gxh^{-1}$.
- $G \curvearrowright X$, $H \leqslant G \implies H \curvearrowright X$.
  Если $H \curvearrowright G$ левыми трансляциями, то орбиты – правые смежные классы.

Лемма Бернсайда вычисляет количество орбит действия группы на множестве с помощью суммы по всем элементам группы. Она применяется в том случае, когда порядок множества X намного больше, чем порядок группы G.

Теорема 3.10 (лемма Бернсайда). *Количество орбит действия группы G на множестве X равно*

$$\frac{1}{|G|}\sum_{g\in G} |\operatorname{Fix}_X(g)|.$$

Доказательство. Обозначим число орбит через N. Каждый элемент $x \in X$ лежит в орбите Gx. Сопоставим ему число $\frac{1}{|Gx|}$. Сумма этих чисел по всем x из данной орбиты $\mathcal{O}$ очевидно равна 1 (мы просто $|\mathcal{O}|$ раз складываем число $\frac{1}{|\mathcal{O}|}$ с самим собой). Поэтому количество орбит можно вычислить по формуле $N = \sum_{x\in X} \frac{1}{|Gx|}$. Подставляя сюда формулу для длины орбиты из леммы 3.8 получим $N = \sum_{x\in X} \frac{|G_x|}{|G|} = \frac{1}{|G|}\sum_{x\in X} |G_x|$. Используя формулу из замечания 3.7 получим $N = \frac{1}{|G|}\sum_{g\in G} |\operatorname{Fix}_X(g)|$, что и требовалось доказать. □

4. Классификация G-множеств

Перейдем теперь к классификации действий данной группы G на множестве. При этом удобно будет говорить, что если G действует на множестве X, то X является G-множеством (по аналогии с R-модулем). Во-первых, надо понять, с точностью до чего будет происходить классификация, т. е. что такое изоморфизм G-множеств.

Определение 4.1. Пусть X, Y – G-множества. Отображение $\varphi : X \to Y$ называется G-эквивариантным, если $\varphi(gx) = g \cdot \varphi(x)$ для любых $x \in X$ и $g \in G$.

Изоморфизмами G-множеств являются G-эквивариантные биекции.

В этом параграфе мы дадим классификацию G-множеств с точностью до изоморфизма.

Лемма 4.2. *Стабилизаторы точек из одной орбиты сопряжены.*

Доказательство. Пусть $x \in X$, а $y \in Gx$, т. е. существует $g \in G$ такое, что $y = gx$. Тогда

$$h \in G_y \iff hy = y \iff hgx = gx \iff g^{-1}hgx = x \iff g^{-1}hg \in G_x.$$

Таким образом, $G_x = g^{-1}G_y g$, что и требовалось. □

Определение 4.3. Если $G \curvearrowright X$ транзитивно, то X называется однородным G-множеством.

Если H – подгруппа в G, то G действует на множестве левых смежных классов G/H по формуле $g(xH) = (gx)H$, $g, x \in G$. Такое действие называется стандартным однородным G-множеством.

Теорема 4.4. *Любое однородное G-множество X изоморфно стандартному однородному G-множеству. Точнее, $X \cong G/G_x$ для любой точки $x \in X$.*

G-множества G/H и G/F изоморфны тогда и только тогда, когда подгруппы H и F сопряжены.

Доказательство. Рассмотрим функцию $\varphi : G/G_x \to Gx = X,\ \ \varphi(gG_x) = gx$, заданную в лемме 3.8, где проверено, что это отображение биективно. Проверка того, что это отображение G-эквивариантно, не составляет труда.

Пусть $F = gHg^{-1}$. Докажем, что $F = G_{gH}$ – стабилизатор элемента gH при действии G на G/H. Действительно,

$$f(gH) = gH \iff g^{-1}fgH = H \iff g^{-1}fg \in H \iff f \in gHg^{-1} = F.$$

Следовательно, по первой части доказательства $G/H \cong G/F$.

Обратно, пусть φ – изоморфизм G-множеств $G/H \xrightarrow{\sim} G/F$, а $\varphi(H) = gF$. Для любого $h \in H$ имеем: $gF = \varphi(H) = \varphi(hH) = h\varphi(H) = hgF$, откуда $g^{-1}hg \in F$. Таким образом, $g^{-1}Hg \subseteq F$. Для доказательства обратного включения заметим, что $\varphi(H) = gF \iff F = \varphi(g^{-1}H) \iff \varphi^{-1}(F) = g^{-1}H$. Следовательно, для любого $f \in F$ имеем $g^{-1}H = \varphi^{-1}(fF) = f\varphi^{-1}(F) = fg^{-1}H$, откуда $gfg^{-1} \in H$. □

Следствие 4.5 (классификация G-множеств)**.** *Любое G-множество изоморфно $\bigsqcup\limits_{i\in\mathscr{I}} G/H_i$, где $\mathscr{I}$ – некоторое множество индексов, а H_i – подгруппа в G. Такое представление единственно, с точностью до перестановки элементов множества $\mathscr{I}$ и замены каждой H_i на сопряженную. Точнее, если*

$$\bigsqcup_{i\in\mathscr{I}} G/H_i \cong \bigsqcup_{j\in\mathscr{J}} G/F_j,$$

то существует биекция $\sigma : \mathscr{I} \to \mathscr{J}$ и элементы $g_i \in G$ такие, что $g_iH_ig_i^{-1} = F_{\sigma(i)}$.

5. Несколько приложений действия группы на множестве

Естественное приложение леммы Бернсайда – задачи о раскрасках. Оно подробно описаны в моем тексте `http://alexei.stepanov.spb.ru/students/algebra3/Bernside.pdf`. Такие задачи естественно разбирать на практических занятиях. В этом же параграфе мы рассмотрим менее очевидные приложения действия групп на множествах. На самом деле, все утверждения этого параграфа являются несложными задачами, которые показывают, как можно использовать изученный в предыдущих двух параграфах материал.

Определение 5.1. Пусть $H \leqslant G$. *Сердцевиной* подгруппы H называется наибольшая нормальная подгруппа $\operatorname{Core} H \trianglelefteq G$, содержащаяся в H.

Легко видеть, что $\operatorname{Core} H = \bigcap_{g\in G} H^g$.

Предложение 5.2. *Ядро транзитивного действия $G \curvearrowright X$ равно $\operatorname{Core} G_x = \cap_{y\in X} G_y$, где $x \in X$.*

Доказательство. Ядро действия по определению есть пересечение стабилизаторов всех точек из X. По лемме 4.2 все эти стабилизаторы сопряжены в G. Легко проверить также, что $gG_xg^{-1} = G_{gx}$, т. е. все подгруппы, сопряженные с G_x являются стабилизаторами точек. □

Теорема 5.3. *Пусть H – подгруппа индекса n в G. Тогда индекс $\operatorname{Core} H$ конечен и делит $n!$. В частности, если в бесконечной группе есть подгруппа конечного индекса, то есть и нормальная подгруппа конечного индекса.*

Доказательство. Стандартное однородное пространство G/H задает гомоморфизм $\theta : G \to S_n$. Порядок образа этого гомоморфизма равен индексу ядра, т. е. $|G : \operatorname{Core} H|$. □

Следующее утверждение – задача, которую я услышал от А. С. Меркурьева, показывает еще один трюк, связанный с гомоморфизмом $G \to S_n$, который возникает при действии G на n-элементном множестве.

Упражнение 5.4. Пусть G – группа четного порядка. Зададим функцию $\varphi : G \to \mathbb{Z}_2$ формулой $\varphi(g) = |G|/\operatorname{ord} g \mod 2$. Докажите, что эта функция – гомоморфизм.

Решение. Пусть $|G| = 2n$. Рассмотрим действие G на себе левыми трансляциями, соответствующее отображение $\theta : G \to S_{2n}$ и его композицию с четностью перестановки $\varepsilon : S_{2n} \to \mathbb{Z}_2$. Если $k = \operatorname{ord} g$, то g отображается в перестановку, состоящую из k-циклов (эти циклы – соответствуют правым смежным классам по $\langle g \rangle$). Четность такой перестановки равна $(k-1) \cdot \frac{2n}{k} = 2n - \frac{2n}{k} \equiv \frac{2n}{k} \mod 2$. Таким образом, функция из условия задачи – это и есть гомоморфизм $\varepsilon \circ \theta$. □

Вот еще одна задача, которую в разных вариациях любят давать на мат-меховских экзаменах.

Упражнение 5.5. Пусть p – наименьшее простое число, делящее порядок группы G. Тогда подгруппа индекса p нормальна в G. В частности, подгруппа индекса 2 всегда нормальна, подгруппа индекса 3 нормальна в группе нечетного порядка и т. п.

Решение. Пусть $|G : H| = p$ – простое, а $|G|$ не делится на простые, меньшие p. Тогда $|G|$ взаимно просто с $(p-1)!$, а индекс $|G : \operatorname{Core} H|$ делит $|G|$ и $p!$. Так как $\gcd(|G|, p!) = p$, то $|G : H| \leqslant |G : \operatorname{Core} H| = p$, откуда $\operatorname{Core} H = H$. □

6. Теоремы Силова

Далее в этом параграфе G – конечная группа, а p – простое число. Группа порядка p^k называется p-группой.

Определение 6.1. p-Подгруппа $S \leqslant G$ называется силовской p-подгруппой, если ее индекс взаимно прост с p.

Целью этого параграфа является доказательство существования силовских подгрупп и их основных свойств. Для этого нам потребуется несколько вспомогательных утверждений.

Лемма 6.2. *Если A, B – p-подгруппы в G, причем A нормализует B, то AB также является p-подгруппой.*

Доказательство. По второй теореме об изоморфизме (теорема 5.6 главы 3) $\frac{AB}{B} \cong \frac{A}{A \cap B}$, отуда $|AB| = |B| \cdot |\frac{A}{A \cap B}|$, что очевидно является степенью числа p. □

Лемма 6.3. *Пусть $G \curvearrowright X$. Предположим, что индекс любой собственной подгруппы G делится на p. Тогда количество неподвижных точек под действием G сравнимо с $|X|$ по модулю p.*

Доказательство. Длина орбиты равна индексу стабилизатора точки. Если этот стабилизатор – собственная подгруппа, т. е. если точка не является неподвижной, то по условию длина ее орбиты делится на p. Количество элементов в X равно сумме длин орбит, следовательно, по модулю p оно сравнимо с суммой длин одноэлементных орбит, т. е. с количеством неподвижных точек. □

Напомним, что центром группы G называется множество элементов, коммутирующих со всеми элементами группы G:

$$\operatorname{Center}(G) = \{c \in G \mid cg = gc \, \forall g \in G\}.$$

Следствие 6.4. *Любая p-группа имеет нетривиальный центр.*

Доказательство. Рассмотрим действие p-группы G на себе сопряжениями. По лемме 6.3 количество неподвижных точек сравнимо с порядком группы по модулю p, а значит, делится на p. Но оно ненулевое, так как одна из неподвижных точек – нейтральный элемент группы. □

Теорема 6.5 (теоремы Силова).

E$_p$: *В G существует силовская p-подгруппа.*
C$_p$: *Все силовские p-подгруппы в G сопряжены.*
D$_p$: *Любая p-подгруппа содержится в силовской p-подгруппе.*
F$_p$: *Количество силовских p-подгрупп сравнимо с 1 по модулю p.*

Доказательство. $\mathbf{E}_p$. Индукция по $|G|$. База тривиальна. Если $|G|$ не делится на p, то доказывать нечего. Поэтому считаем, что $|G| = p^n m$, где m не делится на p.

Пусть существует собственная подгруппа H в G, индекс которой взаимно прост с p. Тогда по индукционному предположению в H существует силовская p-подгруппа, которая будет силовской p-подгруппой в G. В противном случае рассмотрим действие G на себе сопряжениями. По лемме 6.3 количество неподвижных точек этого действия делится на p, т. е. порядок центра группы G делится на p. Предположим, что он равен $p^k\ell$, где ℓ не делится на p. По теореме о строении конечнопорожденных абелевых групп в центре существует подгруппа C порядка p^k. По индукционному предположению в группе G/C, имеющей порядок $p^{n-k}m$, существует силовская p-подгруппа $\overline{P}$, которая имеет порядок p^{n-k}. Обозначим через P полный прообраз группы $\overline{P}$ под действием гомоморфизма редукции $\rho : G \to G/C$. Пусть $\pi : P \to \overline{P}$ – сужение ρ на P. Тогда $|\operatorname{Ker}\pi| = |C| = p^k$, а $|\operatorname{Im}\pi| = |\overline{P}| = p^{n-k}$. Следовательно, $|P| = p^n$, и P является силовской p-подгруппой в G.

$\mathbf{C}_p$+$\mathbf{D}_p$. Пусть H – p-подгруппа, а S – силовская p-подгруппа. Рассмотрим действие H на G/S левыми трансляциями. По лемме 6.3 существует неподвижная точка этого действия, скажем, $HxS = xS$. Тогда $x^{-1}Hx \subseteq S$, откуда H содержится в силовской p-подгруппе xSx^{-1}. С другой стороны, если H была силовской, то мы доказали, что H и S сопряжены.

$\mathbf{F}_p$. Рассмотрим действие силовской p-подгруппы S на множестве X всех силовских p-подгрупп сопряжением. Если $\{P\}$ – неподвижный элемент этого действия, то S нормализует P. Тогда по лемме 6.2 PS является p-подгруппой и, следовательно, совпадает с S. Таким образом, у этого действия ровно одна неподвижная точка. Теперь, по лемме 6.3 $|X| \equiv 1 \mod p$, что и требовалось доказать. □

Лемма 6.6 (аргумент Фраттини). *Пусть H – конечная нормальная подгруппа группы G, а P – силовская подгруппа в H. Тогда $G = N_G(P) \cdot H$.*

Доказательство. Для любого $g \in G$ рассмотрим подгруппу P^g. Так как $P \leqslant H \trianglelefteq G$, то $P^g \leqslant H$, и, следовательно, является силовской подгруппой в H. Так как силовские подгруппы сопряжены, существует элемент $h \in H$ такой, что $P^h = P^g$, откуда $P^{gh^{-1}} = P$, т. е. $gh^{-1} \in N_G(P)$. Таким образом, $g \in N_G(P)h \subseteq N_G(P) \cdot H$. □

Следствие 6.7. *Любая подгруппа, содержащая нормализатор силовской подгруппы, самонормализуема.*

Доказательство. Пусть P – силовская подгруппа группы F. Пусть H содержит ее нормализатор $N_F(P)$, и положим $G = N_F(H)$. Заметим, что силовская подгруппа P группы F является силовской подгруппой и в любой промежуточной между P и F подгруппой, в частности, P – силовская в H. В соответствии с аргументом Фраттини $G = N_G(P) \cdot H \leqslant N_F(P) \cdot H = H$. □

Упражнение 6.8. Для подгруппы B группы G следующие условия эквивалентны.

(1) Любая надгруппа группы B самонормализуема, и никакие 2 различные надгруппы не сопряжены.
(2) $x \in \langle B, B^x\rangle$ для любого $x \in G$.
(3) $B \not\subseteq H^x$ для любых $B \leqslant H \leqslant G$ и $x \in G \smallsetminus H$.

Подгруппа B, удовлетворяющая условиям последнего упражнения, называется абнормальной.

Упражнение 6.9. Докажите, что нормализатор силовской подгруппы абнормален.

Упражнение 6.10. Докажите, что в полной линейной группе $\mathrm{GL}_n(F)$ над конечным полем F характеристики p одна из силовских p-подгрупп – это группа верхних унитреугольных матриц $U_n(F)$,[4] а ее нормализатор – это группа всех обратимых верхнетреугольных матриц $B_n(F)$.

[4]Все остальные силовские p-подгруппы с ней сопряжены, т. е. становятся группой верхних унитреугольных матриц после замены базиса.

Из двух последних утверждений следует, что $B_n(F)$ абнормальна. Оказывается, что это верно над любым полем.

Упражнение 6.11. Пусть K – произвольное поле. Докажите, что $B_n(K)$ абнормальна в $\mathrm{GL}_n(K)$.

7. Полупрямое произведение

В категории $\mathscr{C}$ рассмотрим морфизмы $A \xrightarrow{\varphi} B \xrightarrow{\psi} A$, композиция которых равна id_A. В этом случае морфизм ψ называется ретракцией, а φ называется сечением морфизма ψ. Заметим, что в этом случае φ обязано быть мономорфизмом, а ψ – эпиморфизмом. Приведенная ситуация очень хороша тем, что сохраняется под действием любого функтора. Например, рассмотрим гомоморфизмы коммутативных колец $R \hookrightarrow R[t] \twoheadrightarrow R$ (при втором отображении t переходит в 0) и применим к этой диаграмме функтор GL_n. Получим ретракцию групп $\mathrm{GL}_n(R[t]) \twoheadrightarrow \mathrm{GL}_n(R)$. В настоящем параграфе мы выясним, как устроена любая ретракция групп.

Предложение 7.1. *Пусть H, K – подгруппы в G, причем K нормальна. Следующие условия эквивалентны.*

(1) *Существует ретракция $\psi : G \to H$ группы G на подгруппу H (сечение – это вложение H в G), а $K = \mathrm{Ker}\,\psi$.*
(2) *$G = KH$ и $H \cap K = \{1\}$ (заметим, что KH – подгруппа по теореме 5.6 главы 3).*
(3) *Любой элемент группы G единственным образом представляется в виде произведения kh, $h \in H$, $k \in K$.*

Доказательство. (1) $\Longrightarrow$ (2). По условию композиция вложения $H \hookrightarrow G$ и ψ – тождественное отображение. Другими словами, $\psi(h) = h$ для любого $h \in H$, в частности, $\psi\big(\psi(g)\big) = \psi(g)$. Для любого $g \in G$ имеем: $g = g\psi(g)^{-1} \cdot \psi(g)$, причем $\psi\big(g\psi(g)^{-1}\big) = \psi(g)\psi(g)^{-1} = 1$, т. е. $g\psi(g)^{-1} \in K$, откуда $g \in KH$. Если $g \in H \cap K$, то $g = \psi(g) = 1$, следовательно, пересечение тривиально.

(2) $\Longrightarrow$ (3). Существование очевидно. Если $kh = k'h'$ для некоторых $h, h' \in H$ и $k, k' \in K$, то $hh'^{-1} = k^{-1}k' \in H \cap K = \{1\}$, откуда $h = h'$ и $k = k'$.

(3) $\Longrightarrow$ (1). Для любого $g = kh \in G$, где $k \in K$, $h \in H$, положим $\psi(g) = \psi(kh) = h$. При $h, h' \in H$ и $k, k' \in K$ имеем $khk'h' = k(hk'h^{-1}) \cdot hh'$, $k(hk'h^{-1}) \in K$ и $hh' \in H$. Поэтому $\psi(khk'h') = hh' = \psi(kh)\psi(k'h')$, т. е. ψ – гомоморфизм. Так как $\psi(kh) = 1 \iff h = 1$, то $\mathrm{Ker}\,\psi = K$. С другой стороны, $\psi(h) = h$ для любого $h \in H$, т. е. композиция вложения $H \hookrightarrow G$ и ψ тождественна. □

Определение 7.2. Если выполнены условия предложения 7.1, то G называется (внутренним) полупрямым произведением подгрупп H и K; это обозначается через $G = K \rtimes H$.

Пусть $G = K \rtimes H$, а $\theta : H \to S_K$, $h \mapsto \theta_h$, – гомоморфизм, определяющий действие H на K левым сопряжением, т. е. $\theta_h(k) = hkh^{-1}$. Легко проверить, что сопряжение является автоморфизмом, т. е. образ θ лежит в группе автоморфизмов $\mathrm{Aut}(K)$ группы K. При этом, для $h, h' \in H$ и $k, k' \in K$ имеем $khk'h' = k(hk'h^{-1}) \cdot hh' = k\theta_h(k') \cdot hh'$. В соответствии с последним равенством мы определим внешнее полупрямое произведение произвольных групп A и B, соответствующее гомоморфизму $\theta : B \to \mathrm{Aut}(A)$.

Определение 7.3. Пусть A и B – группы, а $\theta : B \to \mathrm{Aut}(A)$, $b \mapsto \theta_b$ – гомоморфизм. Определим умножение на $G = A \times B$ формулой

$$(a, b) \cdot (a', b') = (a \cdot \theta_b(a'), bb').$$

Тогда G называется (внешним) полупрямым произведением групп A и B, соответствующим θ и обозначается $G = A \rtimes_\theta B$.

Также как и в случае прямого произведения, внешнее полупрямое произведение после некоторых отождествлений становится внутренним.

Предложение 7.4. *Пусть $G = A \rtimes_\theta B$. Положим $A' = A \times \{1_B\}$ и $B' = \{1_A\} \times B$. Тогда G является внутренним полупрямым произведением $A' \rtimes B'$.*

Обратно, если G – внутреннее полупрямое произведение своих подгрупп $K \rtimes H$, а $\theta : H \to \mathrm{Aut}(K)$ – действие H на K левыми сопряжениями, то $G \cong K \rtimes_\theta H$.

Одним из простых геометрических примеров полупрямого произведения является группа автоморфизмов аффинного пространства.

Определение 7.5. Пусть V – векторное пространство над произвольным полем F. Если аддитивная группа V свободно и транзитивно действует на множестве A, то пара (A, V) называется аффинным пространством над F. Действие V на A обычно обозначается сложением: $(a, v) \mapsto a+v$.

Аффинным отображением $\varphi : (A, V) \to (A', V')$ называется пара (φ_a, φ_v), состоящая из линейного отображения $\varphi_v : V \to V'$ и функции $ph_a : A \to A'$, для которых имеет место тождество $\varphi_a(a + v) = \varphi_a(a) + \varphi_v(v)$. Нетрудно видеть, что φ_a однозначно определяется заданием образа одной точки $a \in A$ и линейным отображением φ_v. Действительно, по условию для любой $b \in A$ существует единственный вектор $v \in V$ такой, что $b = a+v$, следовательно, $\varphi_a(b) = \varphi_a(a)+\varphi_v(v)$. Аффинное отображение φ называется изоморфизмом, если φ_v изоморфизм (тогда φ_a автоматически является биекцией).

Пара (B, W) называется аффинным подпространством пространства (A, V), если $B \subseteq A$, $W \leqslant V$, и эта пара сама является аффинным пространством относительно той же операции. Пара (B, W), $B \subseteq A$, $W \leqslant V$, является аффинным подпространством тогда и только тогда, когда $b + w \in B$ для любых $b \in B$ и $w \in W$ и W действует транзитивно на B.

Если мы рассмотрим регулярное действие V на себе, то получим *стандартное* аффинное пространство (V, V).[5] Нетрудно видеть, что любое аффинное пространство изоморфно стандартному. Если $\varphi_v : V \to V$ тождественное отображение, то изоморфизм φ однозначно определяется выбором базовой точки множества A, т. е. образом нуля при отображении φ_a. Давайте изучим строение группы автоморфизмов стандартного аффинного пространства.

Предложение 7.6. *Группа автоморфизмов аффинного пространства (V, V) изоморфна полупрямому произведению $V \rtimes \mathrm{Aut}(V)$. Если $\dim V = n < \infty$, то эта группа изоморфна подгруппе в $\mathrm{GL}_{n+1}(F)$, состоящей из всех матриц, у которых последняя строка совпадает с последней строкой единичной матрицы.*

Доказательство. Сопоставим автоморфизму φ его векторную часть φ_v. Ясно, что это гомоморфизм групп $\pi : \mathrm{Aut}(V, V) \to \mathrm{Aut}(V)$. Ядро этого гомоморфизма состоит из аффинных изоморфизмов вида (ψ, id), которые, как мы заметили выше, однозначно определяются образом нуля. Так как $\psi(v) = \psi(0 + v) = \psi(0) + v$(здесь v в выражении $\psi(v)$ играет роль точки, а в остальных – роль вектора), то ψ – это сдвиг на $\psi(0)$. Поэтому отображение $(\psi, \mathrm{id}) \mapsto \psi(0)$ задает изоморфизм $\mathrm{Ker}\,\psi \cong V$.

С другой стороны, $\mathrm{Aut}(V)$ вкладывается в $\mathrm{Aut}(V, V)$ по правилу $\theta \mapsto (\theta, \theta)$. Таким образом, имеем последовательность гомоморфизмов $\mathrm{Aut}(V) \to \mathrm{Aut}(V, V) \to \mathrm{Aut}(V)$, композиция которых тождественная. Кроме того,

$$(\theta, \theta) \circ (\psi, \mathrm{id}) \circ (\theta^{-1}, \theta^{-1}) = (\theta \circ \psi \circ \theta^{-1}, \mathrm{id}), \text{ а } \theta \circ \psi \circ \theta^{-1}(0) = \theta\big(\psi(0)\big),$$

так что образ автоморфизма θ действует на сдвиг, соответствующий вектору $\psi(0)$ естественным образом. □

На основании теорем Силова и строения полупрямого произведения можно классифицировать все группы порядка pq, где p и q – простые числа. Напомним, что $C_k \cong \mathbb{Z}/k\mathbb{Z}$ обозначает циклическую группу порядка k. Докажем сначала две простые леммы.

Лемма 7.7. *Группа автоморфизмов аддитивной группы $\mathbb{Z}/p\mathbb{Z}$ равна $(\mathbb{Z}/p\mathbb{Z})^* \cong C_{p-1}$.*

[5] Рассмотрение стандартного аффинного пространства вместо векторного пространства – это способ формализовать отождествление точек и векторов.

Доказательство. Автоморфизм φ аддитивной группы $\mathbb{Z}/p\mathbb{Z}$ однозначно определен образом элемента $1+p\mathbb{Z}$: если $\varphi(1+p\mathbb{Z}) = m+p\mathbb{Z}$, то $\varphi(x+p\mathbb{Z}) = mx+p\mathbb{Z}$. Для того чтобы заданный таким образом гомоморфизм φ был автоморфизмом, необходимо и достаточно, чтобы m было взаимно просто с p, т. е. $m \in (\mathbb{Z}/p\mathbb{Z})^*$ (таким образом, элемент группы $(\mathbb{Z}/p\mathbb{Z})^*$ отождествляется с автоморфизмом умножения на этот элемент). В этой части доказательства p не обязательно простое число. Если же p простое, то кольцо $\mathbb{Z}/p\mathbb{Z}$ является полем, следовательно, группа $(\mathbb{Z}/p\mathbb{Z})^*$ циклическая. □

Лемма 7.8. *Множество элементов циклической группы, порядок которых делит фиксированное число q, является циклической подгруппой, порядок которой делит q.*

Если q простое число, то любой неединичный элемент порождает эту подгруппу.

Доказательство. Возведение в степень q является эндоморфизмом абелевой группы, а рассматриваемая подгруппа – его ядро. Любая подгруппа циклической группы циклическая. Порядок циклической группы равен порядку образующей, который по условию делит q. □

Предложение 7.9. *Пусть $p > q$ – простые числа, а G – группа порядка pq. Тогда $G \cong (\mathbb{Z}/p\mathbb{Z}) \rtimes_\theta (\mathbb{Z}/q\mathbb{Z})$.*

Если гомоморфизм θ тривиален, то $G \cong (\mathbb{Z}/p\mathbb{Z}) \times (\mathbb{Z}/q\mathbb{Z}) \cong \mathbb{Z}/pq\mathbb{Z}$. В частности, если $p-1$ не делится на q, то существует единственная группа порядка pq (с точностью до изоморфизма).

Если гомоморфизмы θ, η оба нетривиальны, то $(\mathbb{Z}/p\mathbb{Z}) \rtimes_\theta (\mathbb{Z}/q\mathbb{Z}) \cong (\mathbb{Z}/p\mathbb{Z}) \rtimes_\eta (\mathbb{Z}/q\mathbb{Z})$. В частности, если $p-1$ делится на q, то существует ровно 2 группы порядка pq (с точностью до изоморфизма).

Доказательство. По теореме Силова количество силовских p-подгрупп равно $pk+1$ для некоторого целого неотрицательного k. Так как в группе порядка p нет собственных подгрупп, две различных силовских p-подгруппы пересекаются по нейтральному элементу. Тогда количество элементов во всех силовских p-подгруппах равно $1+(p-1)(pk+1) = p(pk-k+1) \leqslant pq$, откуда $k(p-1) \leqslant q-1$. Но по условию $q < p$, поэтому это неравенство возможно только при $k = 0$. Таким образом, существует только одна силовская p-подгруппа $P \cong \mathbb{Z}/p\mathbb{Z}$. Так как P^g является силовской p-подгруппой, то $P^g = P$ для любого $g \in G$. Следовательно, P нормальна в G. Если $S \cong \mathbb{Z}/q\mathbb{Z}$ – силовская q-подгруппа, то SP является группой, а $S \cap P = \{1_G\}$. По второй теореме об изоморфизме $SP/P \cong S$, откуда $|SP| = pq$ и $SP = G$. Таким образом, $G = P \rtimes S \cong (\mathbb{Z}/p\mathbb{Z}) \rtimes (\mathbb{Z}/q\mathbb{Z})$.

По предложению 7.4 $G \cong (\mathbb{Z}/p\mathbb{Z}) \rtimes_\theta (\mathbb{Z}/q\mathbb{Z})$. Если θ отображает любой элемент $\mathbb{Z}/q\mathbb{Z}$ в тождественный автоморфизм, то полупрямые сомножители поэлементно коммутируют, т. е. произведение прямое. В случае, если $p-1$ не делится на q, никаких других гомоморфизмов из $\mathbb{Z}/q\mathbb{Z}$ в $\mathrm{Aut}(\mathbb{Z}/p\mathbb{Z}) \cong C_{p-1}$ нет, что доказывает второе утверждение нашего предложения.

Пусть теперь $p-1$ делится на q, а $\theta, \eta : \mathbb{Z}/q\mathbb{Z} \to \mathrm{Aut}(\mathbb{Z}/p\mathbb{Z}) = (\mathbb{Z}/p\mathbb{Z})^*$ – два нетривиальных гомоморфизма. Так как группа $(\mathbb{Z}/p\mathbb{Z})^*$ циклическая, то по лемме 7.8 существует $k \in \mathbb{Z}/q\mathbb{Z} \setminus \{0\}$ такое, что $\eta(k) = \eta(1)^k = \theta(1)$. Тогда $\theta(y) = \eta(ky)$ при любом $y \in \mathbb{Z}/q\mathbb{Z}$. Зададим отображение

$$\varphi : (\mathbb{Z}/p\mathbb{Z}) \rtimes_\theta (\mathbb{Z}/q\mathbb{Z}) \to (\mathbb{Z}/p\mathbb{Z}) \rtimes_\eta (\mathbb{Z}/q\mathbb{Z}) \text{ по формуле } \varphi(x,y) = (x, ky).$$

Тогда

$$\varphi\big((x_1,y_1)(x_2,y_2)\big) = \varphi\big(x_1 + x_2 \cdot \theta(y_1), y_1 + y_2\big) = \big(x_1 + x_2 \cdot \theta(y_1), k(y_1+y_2)\big)$$
$$\varphi(x_1,y_1)\varphi(x_2,y_2) = (x_1, ky_1)(x_2, ky_2) = (x_1 + x_2 \cdot \eta(ky_1), ky_1 + ky_2)$$

Эта выкладка доказывает, что φ – гомоморфизм, а его биективность очевидна. □

8. Лемма о бабочке

В этом параграфе мы докажем техническое утверждение, несложно вытекающее из теорем о гомоморфизме.

Напомним, что подгруппа A нормализует подгруппу B, если $a^{-1}Ba = B$ для любого $a \in A$. Другими словами, A нормализует B, если она содержится в нормализаторе

$$N_G(B) := \{g \in G \mid B^g = B\}$$

подгруппы B в группе G. В этом случае $AB = \{ab \mid a \in A, b \in B\}$ является подгруппой в G, в которой B нормальна.

Теорема 8.1 (лемма о бабочке). *Пусть $A' \trianglelefteq A$ и $B' \trianglelefteq B$ – подгруппы некоторой группы G. Тогда*

$$\frac{B'(B \cap A)}{B'(B \cap A')} \cong \frac{A \cap B}{(A \cap B')(B \cap A')} \cong \frac{A'(A \cap B)}{A'(A \cap B')}.$$

Доказательство. Легко видеть, что $A \cap B$ нормализует A', B', $A \cap B'$ и $B \cap A'$, поэтому все произведения подгрупп и все факторгруппы в формуле, которую требуется доказать, корректно определены. Докажем первый изоморфизм, второй следует из него перестановкой букв A и B.

Если C – подгруппа группы D, то $D = CD$. Поэтому с использованием второй теоремы о гомоморфизме получаем

$$\frac{B'(B \cap A)}{B'(B \cap A')} = \frac{B'(B \cap A')(B \cap A)}{B'(B \cap A')} = \frac{(B \cap A)}{B'(B \cap A') \cap B \cap A}.$$

Так как $B'(B \cap A')$ содержится в B, знаменатель последней дроби равен $B'(B \cap A') \cap A$. Ясно, что $B \cap A'$ и $A' \cap B$ содержатся и в A, и в $B'(B \cap A')$. Таким образом, $(A \cap B')(B \cap A') \leqslant B'(B \cap A') \cap A$. Обратно, любой элемент группы $B'(B \cap A') \cap A$ имеет вид ba, где $b \in B'$, $a \in B \cap A'$, а $ba \in A$. Но тогда $b = (ba)a^{-1} \in A$, откуда $ba \in (A \cap B')(B \cap A')$. □

9. Субнормальные ряды

Определение 9.1. Цепочка подгрупп $\{1\} = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_{n-1} \leqslant H_n = G$ называется нормальным (субнормальным) рядом, если H_i нормально в G (соответственно в H_{i+1}) для любого $i = 1, \ldots, n-1$. Длиной ряда называется количество включений (в данном случае n). Факторгруппы H_{i+1}/H_i называются факторгруппами ряда.

Два ряда называются эквивалентными, если их длины равны, а факторгруппы изоморфны с точностью до перестановки, т. е. существует перестановка $\sigma \in S_n$ такая, что факторгруппа первого ряда с номером i изоморфна факторгруппе второго ряда с номером $\sigma(i)$.

Ряд не имеет повторов, если $H_i \neq H_{i+1}$ при всех $i = 1, \ldots, n-1$.

Уплотнение ряда подгрупп – это другой ряд подгрупп, содержащий каждый элемент первоначального ряда.

Субнормальный ряд без повторов, для которого любое его уплотнение без повторов совпадает с ним самим, называется композиционным рядом.

Напомним, что группа называется простой, если она не содержит нетривиальных нормальных подгрупп (т. е. нормальных подгрупп, не совпадающих с самой группой и с единицей). Эквивалентное определение композиционного ряда – это ряд без повторов, у которого все факторгруппы просты.

Ясно, что композиционный ряд существует не для любой группы, но для конечных групп он, конечно, существует. Более общо, композиционный ряд существует тогда и только тогда, когда выполнены условия обрыва возрастающих и убывающих цепей субнормальных подгрупп данной группы. Сейчас мы докажем, что любые 2 композиционных ряда изоморфны (если существуют). Этот факт легко следует из теоремы Шрайера о том, что любые 2 субнормальных ряда имеют эквивалентные уплотнения.

Теорема 9.2 (теорема Шрайера об уплотнении). *Любые 2 субнормальных ряда имеют эквивалентные уплотнения.*

Доказательство. Пусть

$$\{1\} = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_n = G \text{ и } \{1\} = K_0 \leqslant K_1 \leqslant \cdots \leqslant K_m = G$$

субнормальные ряды группы G. Для всех допустимых i, j положим

$$H_{ij} = H_i(H_{i+1} \cap K_j) \text{ и } K_{ji} = K_j(K_{j+1} \cap H_i)$$

(так как $H_i \trianglelefteq H_{i+1}$ и $K_j \trianglelefteq K_{j+1}$, произведения являются подгруппами). Получаем композиционные ряды

$$\begin{aligned}\{1\} = H_{00} \leqslant H_{01} \leqslant \cdots \leqslant H_{0\,m-1} \leqslant \ldots \\ \leqslant H_{i-1\,m} = H_i = H_{i0} \leqslant H_{i1} \leqslant \cdots \leqslant H_{i\,m-1} \leqslant \ldots \\ \leqslant H_{n-2\,m} = H_{n-1} = H_{n-1\,0} \leqslant H_{n-1\,1} \leqslant \cdots \leqslant H_{n-1\,m-1} \leqslant H_{n-1\,m} = G\end{aligned}$$

и аналогично

$$\begin{aligned}\{1\} = K_{00} \leqslant K_{01} \leqslant \cdots \leqslant K_{0\,n-1} \leqslant \ldots \\ \leqslant K_{j-1\,n} = K_j = K_{j0} \leqslant K_{j1} \leqslant \cdots \leqslant K_{j\,n-1} \leqslant \ldots \\ \leqslant K_{m-2\,n} = K_{m-1} = K_{m-1\,0} \leqslant K_{m-1\,1} \leqslant \cdots \leqslant K_{m-1\,n-1} \leqslant K_{m-1\,n} = G\end{aligned}$$

Очевидно, эти ряды являются уплотнениями исходных рядов. Для доказательства того, что они эквивалентны, докажем, что $H_{i\,j+1}/H_{ij} \cong K_{j\,i+1}/K_{ji}$. Другими словами,

$$\frac{H_i(H_{i+1} \cap K_{j+1})}{H_i(H_{i+1} \cap K_j)} \cong \frac{K_j(K_{j+1} \cap H_{i+1})}{K_j(K_{j+1} \cap H_i)}.$$

Но последний изоморфизм – это в точности лемма о бабочке. □

Из теоремы следует, что любые 2 композиционных ряда имеют эквивалентные уплотнения. Если из этих уплотнений выкинуть повторения, то получатся исходные композиционные ряды, которые, естественно, также будут эквивалентны.

Следствие 9.3. *Любые 2 композиционных ряда группы эквивалентны.*

Факторы композиционного ряда называются композиционными факторами группы. Как мы только что доказали, набор (но не последовательность!) композиционных факторов определен однозначно. Это и означает, что любая конечная группа строится из кирпичиков, которыми являются простые группы. При этом набор кирпичиков, но не их порядок, определен группой однозначно. И, как мы уже видели при изучении групп порядка pq, из одного набора кирпичиков, даже сложенного в одном порядке, можно получить разные группы.

10. Примеры простых групп

Простейшими примерами простых групп являются группы A_n, $n \geqslant 5$, и $\mathrm{PSL}_n(F)$, где $n \geqslant 3$ или $n = 2$, а поле F содержит больше 3 элементов. Покажем без подробностей путь доказательства простоты A_n.

Лемма 10.1. *Группа A_n порождена 3-циклами.*

Следствие 10.2. *Если подгруппа $H \trianglelefteq A_n$ содержит 3-цикл, но она совпадает с A_n.*

Лемма 10.3. *Пусть $n > 3$. Для любой $\sigma \in S_n \smallsetminus \{\mathrm{id}\}$ существует $\tau \in A_n$ такая, что $[\sigma, \tau] \neq \mathrm{id}$ и имеет не менее $n - 5$ неподвижных точек.*

Теорема 10.4. *Группа A_n является простой при $n \geqslant 5$.*

Доказательство теоремы состоит в получении 3-цикла из четной перестановки, двигающей не более 5 элементов, при помощи коммутирования.

Другой пример простой группы – проективная специальная линейная группа. Напомним, что $\mathrm{PSL}_n(F)$ обозначает факторгруппу группы $\mathrm{SL}_n(F)$, состоящей из матриц с определителем 1, по центру, т. е. по подгруппе скалярных матриц.

Теорема 10.5. *Пусть F – поле, а $n \geqslant 2$. При $n = 2$ предположим дополнительно, что $|F| > 3$. Тогда группа $\mathrm{PSL}_n(F)$ проста.*

Если поле конечно, то $\mathrm{PSL}_n(F)$ – конечная простая группа. Большая часть конечных простых групп – это проективные линейные группы, т.е. факторгруппы по центру матричных групп над конечными полями. Они называются группами типа Ли, потому что в каком-то смысле они похожи на группы Ли. Группы типа Ли классифицируются некоторыми комбинаторными структурами, называемыми системами корней (система корней – это конечный набор точек в евклидовом пространстве, обладающий большой группой симметрий). Теорема классификации простых конечных групп утверждает, что кроме знакопеременных групп и групп типа Ли существует ровно 26 простых конечных групп, которые называются спорадическими группами.

11. Разрешимые группы

Определение 11.1. *Разрешимая группа* – группа, обладающая субнормальным рядом подгрупп с абелевыми факторами.

Как видно из следующего утверждения, разрешимая группа всегда обладает *нормальным* рядом с абелевыми факторами. Определим *производный ряд* группы G следующим образом:

$$D^0G = G,\ D^1G = [G,G],\ D^{k+1}G = [D^kG, D^kG].$$

Ясно, что производный ряд является нормальным рядом с абелевыми факторами.

Предложение 11.2. *Группа разрешима тогда и только тогда, когда ее производный ряд обрывается на $\{1\}$.*

Доказательство. Если производный ряд конечен, то он и является субнормальным (и даже нормальным) рядом с абелевыми факторами. Обратно, если $\{1\} = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_{n-1} \leqslant H_n = G$ – субнормальный ряд с абелевыми факторами, то $[H_k, H_k] \leqslant H_{k-1}$. По индукции легко получаем, что $D^kG \leqslant H_{n-k}$, откуда $D^nG = \{1\}$. □

Длина производного ряда, т. е. наименьшее n, для которого $D^nG = \{1\}$, называется ступенью разрешимости группы.

Теорема 11.3. *Пусть $H \triangleleft G$. Группа G разрешима тогда и только тогда, когда разрешимы H и G/H.*

Доказательство. Ясно, что $D^kH \leqslant D^kG$, а D^kG/H является образом D^kG под действием канонической проекции $G \to G/H$. Следовательно разрешимость G влечет разрешимость H и G/H. Обратно, если $D^kG/H = \{1\}$, то $D^kG \leqslant H$. Теперь, если $D^mH = \{1\}$, то $D^{m+k}G = D^m(D^kG) = \{1\}$. □

Следствие 11.4. *Пусть H_1 и H_2 – разрешимые подгруппы в G, и H_1 нормализует H_2. Тогда H_1H_2 разрешима.*

Доказательство. По второй теореме о гомоморфизме 5.6 произведение H_1H_2 является подгруппой, H_2 нормальна в этой подгруппе, и $\frac{H_1H_2}{H_2} \cong \frac{H_1}{H_1 \cap H_2}$. Так как H_1 разрешима, то правая часть последнего изоморфизма разрешима. Следовательно, разрешима факторгруппа H_1H_2/H_2 и, по условию, подгруппа H_2. Теперь результат следует из теоремы. □

Разрешимость группы Галуа поля разложения неприводимого сепарабельного многочлена p равносильно разрешимости уравнения $p(t) = 0$ в радикалах. Это одна из основных теорем теории Галуа, которая дала название разрешимым группам. Исходя из этих соображений интересно изучать разрешимость конечных групп.

Из последнего следствия вытекает, что в конечной группе существует наибольшая разрешимая нормальная подгруппа – произведение всех разрешимых подгрупп данной группы. Она называется разрешимым радикалом группы. Фактор по разрешимому радикалу уже не содержит разрешимых подгрупп.

Для группы, обладающей композиционным рядом, например конечной, существует альтернатива: либо группа разрешима, либо хотя бы один из композиционных факторов является неабелевой простой группой. Самая маленькая неабелева простая группа – знакопеременная группа A_5, порядок которой равен 60. Поэтому все группы меньших порядков разрешимы.

12. Нильпотентные группы

Определение 12.1. Цепочка нормальных подгрупп $\{1\} = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_{n-1} \leqslant H_n = G$ называется *центральным рядом*, если H_{i+1}/H_i лежит в центре G/H_i (равносильно: $[G, H_{i+1}] \leqslant H_i$) для любого $i = 1, \ldots, n$.

Нильпотентная группа – группа, обладающая центральным рядом подгрупп.

Определим *нижний центральный ряд* группы G по индукции следующим образом. Положим $C_0(G) = G$ и $C_n(G) = [C_{n-1}(G), G]$ для $n \geqslant 1$. Из определения сразу вытекает, что

$$G = C_0(G) \geqslant C_1(G) \geqslant C_2(G) \geqslant \ldots \geqslant C_n(G) \geqslant \ldots$$

Ясно, что этот ряд действительно является центральным. Нижний центральный ряд известен также под именем *убывающий* центральный ряд. Очень часто эпитет “нижний” или “убывающий” здесь опускается и говорят просто о центральном ряде.

Упражнение 12.2. Докажите, что $[C_m(G), C_n(G)] \leqslant C_{m+n+1}(G)$.

Определим теперь *верхний центральный ряд* группы G по индукции следующим образом. Положим $C^0(G) = 1$, а для $n \geqslant 1$ обозначим через $C^n(G)$ полный прообраз центра фактор-группы $G/C^{n-1}(G)$ по предыдущему члену ряда, относительно канонической проекции. Иными словами, $C^n(G)$ это такая подгруппа $C^{n-1}(G) \leqslant C^n(G) \leqslant G$, что ее фактор-группа по $C^{n-1}(G)$ совпадает с центром фактор-группы $G/C^{n-1}(G)$:

$$C^n(G)/C^{n-1}(G) = C(G/C^{n-1}(G)).$$

Таким образом, $C^1(G) = C(G)$ – центр группы G, $C^2(G)$ – *гиперцентр*, т. е. такая подгруппа в G, что $C^2(C)/C(G) = C(G/C(G))$ и так далее. Верхний центральный ряд часто называется также *возрастающим* центральным рядом, а его член $C^n(G)$ обычно называется также *n-м гиперцентром* группы G.

Приведем классический пример вычисления центральных рядов, часто встречающийся в приложениях. Напомним, что

$$U_n^{(k)} = U_n^{(k)}(F) = \{a \in \mathrm{M}_n(F) \mid a_{ii} = 1,\ a_{ij} = 0 \text{ при всех } i \neq j,\ j - i < k\},$$

см. параграф 8 главы 3.

Упражнение 12.3. Докажите, что $C_m(U(n, F)) = U_n^{(m)}(F) = C^{n-1-m}(G)$.

Теорема 12.4. *Пусть $\{1\} = H_0 \leqslant H_1 \leqslant \cdots \leqslant H_{n-1} \leqslant H_n = G$ – центральный ряд группы G. Тогда $C_{n-k}(G) \leqslant H_k \leqslant C^k(G)$.*

Длины нижнего и верхнего центральных рядов нильпотентной группы равны.

Доказательство. По определению центрального ряда $[G, H_k] \leqslant H_{k-1}$. Теперь простая индукция по $n-k$ показывает, что $C_{n-k}(G) = [G, C_{n-k-1}(G)] \leqslant [G, H_{k+1}] \leqslant H_k$. Таким образом, длина любого центрального ряда не меньше длины нижнего.

Докажем теперь второе включение индукцией по k. База $k = 0$ очевидна. Далее, $[G, H_{k+1}] \leqslant H_k \leqslant C^k$ по индукционному предположению. Следовательно, H_{k+1}/C^k лежит в центре G/C^k. Так как C_{k+1} – полный прообраз этого центра, то $H_{k+1} \leqslant C_{k+1}$. В частности, длина любого центрального ряда не меньше длины верхнего. Таким образом и нижний, и вехний центральные ряды имеют наименьшую возможную длину, следовательно их длины равны. □

Длина нижнего центрального ряда нильпотентной группы называется *классом* или *степенью нильпотентости*. Любая нильпотентная группа класса n удовлетворяет тождеству

$$\big[[\dots[x_1, x_2], \dots], x_n\big] = 1,$$

Таким образом, имеется универсальный прием доказательства всех результатов о нильпотентных группах – индукция по классу нильпотентности. Классифицировать нильпотентные группы невозможно, но индивидуально каждая из них устроена чрезвычайно незатейливо.

Далее мы докажем, что класс нильпотентных групп замкнут относительно перехода к подгруппам, фактор-группам и прямым произведениям.

Предложение 12.5. *Класс нильпотентных групп обладает следующими свойствами:*

- *Подгруппа нильпотентной группы класса n нильпотентна класса $\leqslant n$;*
- *Фактор-группа нильпотентной группы класса n нильпотентна класса $\leqslant n$;*
- *Прямое произведение двух нильпотентных групп классов m и n нильпотентно класса $\max(m, n)$.*

Доказательство абсолютно аналогично первой части доказательства теоремы 11.3.

В противоположность классу разрешимых групп, класс нильпотентных групп не замкнут относительно расширений. Иными словами, если $H \trianglelefteq G$ нормальная подгруппа в G, причем как H, так и G/H нильпотентны, то отсюда еще совершенно не следует, что G тоже нильпотентна. Однако имеет место следующий результат.

Теорема 12.6 (Холла). *Если $H \trianglelefteq G$ нормальная подгруппа в G, причем как H, так и $G/[H, H]$ нильпотентны, то G тоже нильпотентна.*

Подгруппа H в G называется *субнормальной*, если существует субнормальный ряд $H = H_0 \leqslant H_1 \leqslant \dots H_d = G$. Длина самого короткого такого ряда называется глубиной подгруппы H. Часто условие нильпотентности используется в следующей форме.

Теорема 12.7. *Каждая подгруппа H нильпотентной группы G субнормальна, причем глубина d подгруппы H не превосходит класс нильпотентности группы G.*

Доказательство. По определению гиперцентров имеем

$$[C^{i+1}(G), C^i(G) \cdot H] \leqslant [C^{i+1}(G), G] \leqslant C^i(G) \leqslant C^i(G) \cdot H,$$

так что $C^{i+1}(G)$ нормализует $C^i(G) \cdot H$, откуда $C^i(G) \cdot H \lhd C^{i+1}(G) \cdot H$. Пусть k – наибольшее целое такое, что $C^k(G) \leqslant H$. Получаем субнормальный ряд

$$H = C^k(G) \cdot H \leqslant C^{k+1}(G) \cdot H \leqslant \dots \leqslant C^n(G) \cdot H = G,$$

длина которого не больше длины верхнего центрального ряда. □

Очевидно, что из этой теоремы сразу вытекает такое следствие, называемое *нормализаторным условием*.

Следствие 12.8. *Если группа G нильпотентна, то $H \lneqq N_G(H)$ для любой собственной подгруппы $H \lneqq G$.*

Группа унитреугольных матриц является самым важным и самым типичным примером нильпотентной группы. Во многих важных классах групп никаких других нильпотентных групп, кроме подгрупп в $U(n,R)$, не бывает. Приведем два примера.

Каждая конечная группа вкладывается в подходящую полную линейную группу $\mathrm{GL}(n,K)$ над произвольным полем K. Если $\operatorname{char} K = p$, то группа $U = U(n,K)$ является силовской p-подгруппой в $\mathrm{GL}(n,K)$. Тем самым, каждая конечная p-группа вкладывается в группу $U(n,K)$. Из доказанной ниже теоремы Бернсайда–Виландта вытекает, что каждая конечная нильпотентная группа изоморфна подгруппе в $U(n,R)$, где $R = K_1 \oplus \ldots \oplus K_s$, где характеристики $p_1,\ldots,p_s$ полей $K_1,\ldots,K_s$ пробегают множество простых делителей порядка G.

Так как

$$\begin{pmatrix} 1 & * & * & * \\ 0 & 1 & x & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & * & * & * \\ 0 & 1 & nx & * \\ 0 & 0 & 1 & * \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

то группа $U(n,\mathbb{Z})$ не имеет кручения. Как мы знаем, каждая подгруппа нильпотентной группы является нильпотентной. Среди конечно порожденных групп никаких других нильпотентных групп без кручения нет.

Теорема 12.9 (Холла)**.** *Конечно порожденная нильпотентная группа без кручения изоморфна подгруппе в $U(n,\mathbb{Z})$ для достаточно большого n.*

Как утверждает следствие 6.4, центр конечной p-группы нетривиален. Поэтому любая конечная p-группа нильпотентна. Тем самым, нильпотентно и конечное прямое произведение конечных p-групп. В заключении параграфа мы докажем, что никаких других нильпотентных конечных групп не бывает.

Из сопряженности силовских подгрупп вытекает следующий результат.

Лемма 12.10. *Если все силовские подгруппы группы G нормальны, то G является их декартовым произведением.*

Таким образом, в этом случае группа G нильпотентна.

Доказательство. Так как силовские p-подгруппы сопряжены, для каждого простого p существует ровно одна силовская подгруппа. Так как силовские p-подгруппы с разными p имеют тривиальное пересечение, то их призведение прямое. Из сравнения количества элементов в этом произведении с порядком группы следует, что оно равно всей группе. □

Оказывается, верно и обратное. Иными словами, *конечная* группа G тогда и только тогда нильпотентна, когда она раскладывается в прямое произведение силовских подгрупп по всем простым делителям p порядка G. Или, иначе, конечная группа G тогда и только тогда нильпотентна, когда все ее силовские подгруппы нормальны.

Теорема 12.11 (Бернсайда–Виландта)**.** *Для конечной группы G следующие условия эквивалентны.*

(1) *G нильпотентна.*
(2) *любая подгруппа $H \leqslant G$ субнормальна.*
(3) *G является прямым произведением своих силовских p-подгрупп.*
(4) *G изоморфно декартову произведению p-групп.*

Доказательство. Выше мы доказали, что первое условие влечет второе для произвольных нильпотентных групп, а не только конечных. С другой стороны, мы уже убедились, что группа, являющаяся прямым произведением своих силовских p-подгрупп, нильпотентна. Эквивалентность последних двух условий очевидна.

Таким образом, нам остается только доказать, что второе условие влечет третье, причем согласно предыдущей лемме для этого нам достаточно проверить, что все силовские p-подгруппы

нормальны в G. В самом деле, пусть P – какая-то силовская p-подгруппа. По следствию 6.7 из аргумента Фраттини ее нормализатор $N_G(P)$ самонормализуем. Но по условию (2) он субнормален, следовательно, $N_G(P) = G$, что и требовалось. □

Следствие 12.12. *Конечная группа нильпотентна в том и только том случае, когда любые два элемента взаимно простых порядков коммутируют.*

Доказательство. Если группа G нильпотентна, то она изоморфна (будем считать, что равна) прямому произведению p_i-подгрупп P_i по $i = 1, \ldots, n$, где p_i – различные просты числа. Любой элемент $g \in G$ имеет вид $(g_1, \ldots, g_n)$, где $g_i \in P_i$. Тогда $\operatorname{ord} g = \operatorname{ord} g_1 \cdot \ldots \cdot \operatorname{ord} g_n$, причем $\operatorname{ord} g_i$ является степенью p_i. Пусть $h = (h_1, \ldots, h_n)$. Порядки g и h взаимно просты тогда и только тогда, когда для любого $k = 1, \ldots, n$ либо $g_k = 1$, либо $h_k = 1$ (иначе порядки обоих делятся на p_k). В любом случае g_k и h_k коммутируют, следовательно, g коммутирует с h.

Обратно, предположим, что любые два элемента группы G взаимно простых порядков коммутируют. Выберем по одной силовской p-подгруппе S_p для каждого p, делящего $|G|$. По условию подгруппы S_p и S_q поэлементно коммутируют для любых p, q. Ясно также, что они пересекаются по 1. Следовательно, $\langle S_p \cup S_q \rangle = S_p \times S_q$. По индукции получим, что подгруппа, порожденная всеми выбранными силовскими подгруппами равна их прямому произведению. Так как количество элементов в этом прямом произведении равно $|G|$, то оно равно G. По теореме Бернсайда–Виландта из этого следует, что G нильпотентна. □

Глава 9

Начала теории категорий

Чем раньше математик начинает пользоваться языком теории категорий, тем проще ему потом изучать любые области математики, хотя бы потому что он видит формальные взаимосвязи между ними. Весь курс алгебры буквально пронизан категориями, функторами, естественными (и неестественными!) преобразованиями, универсальными конструкциями и сопряженными функторами. Поэтому прочитав настоящую главу вы сможете лучше увидеть взаимосвязи между различными понятиями и утверждениями.

1. Категория, универсальные объекты, типы морфизмов

Определение 1.1. *Категорией* $\mathscr{C}$ называется набор следующих данных:

- класс *объектов* $\operatorname{Obj}\mathscr{C}$;
- для каждых двух объектов $X, Y \in \operatorname{Obj}\mathscr{C}$ множество $\operatorname{Mor}(X, Y)$, называемое множеством *морфизмов* из X в Y;
- для каждых трех объектов $X, Y, Z \in \operatorname{Obj}\mathscr{C}$ функция $\operatorname{Mor}(Y, Z) \times \operatorname{Mor}(X, Y) \to \operatorname{Mor}(X, Z)$, $(\varphi, \psi) \mapsto \varphi\psi$, называемую *законом композиции* морфизмов;

удовлетворяющих следующим условиям:

(1) Если $X \neq U$ или $Y \neq V$, то $\operatorname{Mor}(X, Y) \cap \operatorname{Mor}(U, V) = \varnothing$.
(2) закон композиции морфизмов ассоциативен;
(3) для каждого объекта $X \in \operatorname{Obj}\mathscr{C}$ существует *тождественный морфизм* $\mathrm{id}_X \in \operatorname{Mor}(X, X)$ такой, что для любых морфизмов $\alpha \in \operatorname{Mor}(X, Y)$ и $\beta \in \operatorname{Mor}(Y, X)$ выполнены равенства $\mathrm{id}_X \beta = \beta$ и $\alpha \mathrm{id}_X = \alpha$.

Категория называется *малой*, если класс объектов является множеством.

Часто мы будем писать:

- $X \in \mathscr{C}$ вместо $X \in \operatorname{Obj}\mathscr{C}$;
- $\operatorname{Mor}_{\mathscr{C}}(X, Y)$ вместо $\operatorname{Mor}(X, Y)$, если из контекста неясно, про какую категорию идет речь;
- $\varphi : X \to Y$ вместо $\varphi \in \operatorname{Mor}(X, Y)$;
- $\varphi \in \operatorname{Mor}\mathscr{C}$ для любого морфизма категории $\mathscr{C}$ (т. е. $\operatorname{Mor}\mathscr{C}$ – это класс всех морфизмов категории $\mathscr{C}$);
- $X = \operatorname{source}\varphi$ и $Y = \operatorname{target}\varphi$ для морфизма $\varphi : X \to Y$.

Определение 1.2. Категория $\mathscr{B}$ называется подкатегорией категории $\mathscr{C}$, если $\operatorname{Obj}\mathscr{B} \subseteq \operatorname{Obj}\mathscr{C}$ и $\operatorname{Mor}_{\mathscr{B}}(X, Y) \subseteq \operatorname{Mor}_{\mathscr{C}}(X, Y)$ для любых $X, Y \in \mathscr{B}$. Подкатегория $\mathscr{B}$ называется полной, если последнее включение всегда является равенством.

Определение 1.3. Пусть $\mathscr{C}$ – категория. Противоположной категорией к $\mathscr{C}$ называется категория $\mathscr{C}^{op}$:

- $\operatorname{Obj}\mathscr{C}^{op} = \operatorname{Obj}\mathscr{C}$;
- $\operatorname{Mor}_{\mathscr{C}^{op}}(X, Y) = \operatorname{Mor}_{\mathscr{C}}(Y, X)$ для каждых двух объектов $X, Y \in \mathscr{C}$;
- закон композиции в $\mathscr{C}^{op}$ отличается от закона композиции в $\mathscr{C}$ порядком, т. е. $(\alpha\beta)_{\mathscr{C}^{op}} = (\beta\alpha)_{\mathscr{C}}$.

Определение 1.4. Декартовым произведением $\mathscr{C} \times \mathscr{B}$ категорий $\mathscr{C}$ и $\mathscr{B}$ называется следующая категория:

- $\mathrm{Obj}\,(\mathscr{C}\times\mathscr{B}) = \mathrm{Obj}\,\mathscr{C}\times\mathrm{Obj}\,\mathscr{B}$;
- $\mathrm{Mor}((X,Y),(Z,W)) = \mathrm{Mor}(X,Z)\times\mathrm{Mor}(Y,W)$;
- $(\alpha,\beta)(\gamma,\delta) = (\alpha\gamma,\beta\delta)$.

Определение 1.5. Морфизм φ называется *мономорфизмом*, если равенство $\varphi\alpha = \varphi\beta$ влечет равенство $\alpha = \beta$, и эпиморфизмом, если $\alpha\varphi = \beta\varphi \implies \alpha = \beta$. Морфизм, являющийся одновременно мономорфизмом и эпиморфизмом называется биморфизмом. Морфизм $\varphi \in \mathrm{Mor}(X,Y)$ называется *изоморфизмом*, если существует $\varphi^{-1} \in \mathrm{Mor}(Y,X)$ такой, что $\varphi\varphi^{-1} = \mathrm{id}_Y$ и $\varphi^{-1}\varphi = \mathrm{id}_X$.

Замечание 1.6. Очевидно, что любой изоморфизм является биморфизмом. Обратное вообще говоря неверно.

Определение 1.7. Объект $*$ называется *инициальным*, если для любого объекта X множество $\mathrm{Mor}(*,X)$ состоит ровно из одного элемента. Объект $*$ называется *финальным*, если для любого объекта X множество $\mathrm{Mor}(X,*)$ состоит ровно из одного элемента.

Предложение 1.8. *Инициальный (финальный) объект категории определен единственным образом с точностью до единственного изоморфизма, т. е. между двумя инициальными (финальными) объектами есть единственный изоморфизм.*

Доказательство. Если $*$ и $*'$ – инициальные объекты категории $\mathscr{C}$, то существуют единственные морфизмы $\alpha : * \to *'$ и $\alpha' : *' \to *$. Тогда композиции $\alpha\alpha'$ и $\alpha'\alpha$ – это единственные эндоморфизмы $*'$ и $*$ соответственно. Но этими единственными морфизмами по определению категории должны быть тождественные. Следовательно, α и α' – взаимно обратные изоморфизмы. □

Примеры.

(1) Категория множеств **Set**, пунктированных множеств $\mathbf{Set}_*$.
(2) Категория множеств с инъективными (сюръективными) отображениями.
(3) Алгебраические и геометрические структуры.
   - Категория моноидов **Mon**.
   - Категория групп **Grp**.
   - Категория абелевых групп **Ab**.
   - Категория F-векторных пространств **Vect** $= F$-**Vect**, конечномерных векторных пространств F-$\mathbf{Vect}_{f.d.}$.
   - Категория R-модулей R-**Mod**.
   - Категория R-алгебр R-**Alg**.
   - Категория колец с 1 **Ring**, без 1 – **Rng**
   - Категория коммутативных колец с 1 **CRing**, без 1 **CRng**
   - Категория полей (автоматом все морфизмы – мономорфизмы).
   - Категория топологических пространств **Top**, пунктированных топологических пространств $\mathbf{Top}_*$.
(4) Моноид – категория с одним объектом.
(5) Группоид – малая категория, все морфизмы которой являются изоморфизмами. Обобщение понятия группы. Полезно, например, при изучении фундаментальной группы несвязных пространств. Строится фундаментальный группоид: объекты – точки данного топологического пространства, Морфизм из точки a в точку b – класс гомотопных путей из a в b. Композиция – конкатенация путей, как она определяется в топологии (с точности до гомотопии).
(6) Категория матриц над ассоциативным кольцом R с 1: объекты – натуральные числа, $\mathrm{Mor}(m,n) = \mathrm{M}_{n\times m}(R)$. Композиция морфизмов – произведение матриц.
(7) Частично упорядоченное множество, из a в b есть ровно 1 морфизм $\iff a \leqslant b$.
(8) Категории, связаные с ориентированным графом:

- категория путей: объекты – вершины, морфизмы – пути, включая пустой путь из точки в себя, композиция – конкатенация.
- категория достижимости: из a в b есть ровно 1 морфизм $\iff$ из a существует путь в b.

(9) Категория морфизмов $\mathrm{Mor}\,\mathscr{C}$. Объекты – морфизмы в категории $\mathscr{C}$, морфизмы из $\varphi : X \to Y$ в $\varphi' : X' \to Y'$ – коммутативные квадраты

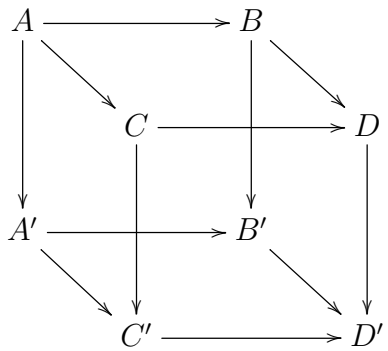
$$\begin{array}{ccc} X & \xrightarrow{\varphi} & Y \\ \downarrow & & \downarrow \\ X' & \xrightarrow{\varphi'} & Y' \end{array}$$

(10) Γ – ориентированный граф. Определим категорию коммутативных $\mathscr{C}_\Gamma$ диаграмм в $\mathscr{C}$ типа Γ. Диаграмма в $\mathscr{C}$ типа Γ – это функция, сопоставляющая вершинам Γ объекты $\mathscr{C}$, а ребрам Γ – морфизмы $\mathscr{C}$. Диаграмма называется коммутативной, если для любых двух вершин и для любых двух путей, соединяющих эти вершины, композиции морфизмов вдоль этих путей совпадают.

Морфизм η диаграмм F и G – набор морфизмов $\eta_a : F(a) \to G(a)$, где a пробегает вершины графа Γ, такой что для любого ребра α из вершины a в вершину b квадрат

$$\begin{array}{ccc} F(a) & \xrightarrow{F(\alpha)} & F(b) \\ {\scriptstyle \eta_a}\downarrow & & \downarrow{\scriptstyle \eta_b} \\ G(a) & \xrightarrow{G(\alpha)} & G(b) \end{array}$$

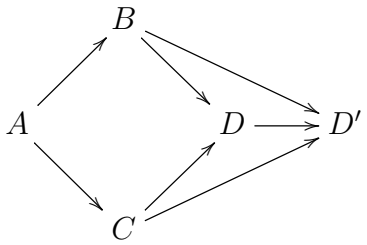
коммутативен. Например, если Γ – это квадрат со стрелками слева направо и сверху вниз, то морфизм диаграмм типа Γ – это коммутативный куб



(11) Часто рассматривают подкатегорию в категории коммутативных диаграмм, состоящую из диаграмм, у которых объекты в нескольких вершинах и несколько морфизмов между ними фиксированы. В этом случае морфизмы – это морфизмы диаграмм, тождественные на зафиксированных вершинах. Например, для данных объектов $A, B, C \in \mathscr{C}$ и морфизмов $C \leftarrow A \to B$ можно рассмотреть категорию квадратов вида

$$\begin{array}{ccc} A & \longrightarrow & B \\ \downarrow & & \downarrow \\ C & \longrightarrow & D \end{array}$$

морфизмами в которой являются коммутативные диаграммы



Инициальный объект в этой категории называется пушаутом диаграммы $C \leftarrow A \rightarrow B$ или копроизведением B и C над A. Позже мы рассмотрим несколько важных примеров пушаутов, а также двойственного понятия – пулбэка ([расслоенного] произведения).

2. Функторы

Определение 2.1. Функтором $\mathscr{F}$ из категории $\mathscr{B}$ в категорию $\mathscr{C}$ называется набор следующих отображений:

- $\mathscr{F} : \operatorname{Obj}\mathscr{B} \to \operatorname{Obj}\mathscr{C}$;
- $\mathscr{F}_{X,Y} : \operatorname{Mor}_{\mathscr{B}}(X,Y) \to \operatorname{Mor}_{\mathscr{C}}(\mathscr{F}(X),\mathscr{F}(Y))$ для каждой пары объектов $X, Y \in \mathscr{B}$,

удовлетворяющих свойствам $\mathscr{F}_{X,Z}(\alpha\beta) = \mathscr{F}_{Y,Z}(\alpha)\mathscr{F}_{X,Y}(\beta)$ и $\mathscr{F}_{X,X}(\mathrm{id}_X) = \mathrm{id}_{\mathscr{F}(X)}$.

Индексы в обозначении отображения $\mathscr{F}_{X,Y}$ обычно опускают, потому что они однозначно восстанавливаются по аргументу. В таких обозначениях свойства в определении означают, что $\mathscr{F}$ сохраняет композицию морфизмов и тождественные морфизмы. Для функторов используется такое же обозначение, что и для функций: запись $\mathscr{F} : \mathscr{B} \to \mathscr{C}$ означает, что $\mathscr{F}$ является функтором из категории $\mathscr{B}$ в категорию $\mathscr{C}$.

Функтор $\mathscr{B}^{op} \to \mathscr{C}$ называется контравариантным функтором из $\mathscr{B}$ в $\mathscr{C}$ (обычный функтор, если хочется подчеркнуть, что он не меняет направление стрелок, называется ковариантным). Контравариантный функтор можно рассматривать и как функтор из $\mathscr{B} \to \mathscr{C}^{op}$.

Функтор называется строгим (полным), если он действует инъективно (сюръективно) на каждом множестве морфизмов. Образ полного функтора – полная подкатегория.

Примеры.

(1) Забывающие функторы – строгие функторы, “забывающие” часть структуры на объекте категории. Неформальное понятие.
(2) Функтор вложения подкатегории в категорию. Очевидно он строгий. Он является полным $\iff$ подкатегория полная. Как вы видите, строгий и полный функтор не обязан быть изоморфизмом и даже, как мы увидим позже, эквивалентностью категорий.
(3) M_n, GL_n, обратимые элементы моноида.
(4) Центр группы $Z(G) = \{a \in G \mid ag = ga \,\forall g \in G\}$ не определяет функтор, потому что образ центра не обязательно лежит в центре. Он будет функтором, если в качестве морфизмов в категории групп рассматривать только сюръективные гомоморфизмы.
(5) Отображение, сопоставляющее группе ее коммутант, естественным образом определяет функтор, так как образ коммутанта очевидно содержится в коммутанте. Действие этого функтора на морфизмах – это просто сужение гомоморфизма групп на коммутанты (уменьшается как область определения, так и множество значений).
(6) Отображение, которое множеству ставит в соответствие дискретное (антидискретное) топологическое пространство, свободную группу, свободный модуль над фиксированным кольцом.
(7) Гомотопические группы: $\pi_0 : \mathbf{Top}_* \to \mathbf{Set}_*$ – множество компонент связности, $\pi_1 : \mathbf{Top}_* \to \mathbf{Grp}$ – фундаментальная группа, $\pi_i : \mathbf{Top}_* \to \mathbf{Ab}$ – высшие гомотопические группы.

(8) В произвольной категории $\mathscr{C}$ отображение Aut, отображающее объект в множество (моноид) его автоморфизмов, не является функтором, так как нет никакого разумного способа по морфизму $\varphi : A \to B$ задать отображение $\operatorname{Aut} A \to \operatorname{Aut} B$ или обратно.

(9) Пусть Γ – ориентированный граф, а $\mathscr{C}_\Gamma$ – категория достижимости, связанная с этим графом. Функтор из $\mathscr{C}_\Gamma$ в произвольную категорию $\mathscr{B}$ называется коммутативной диаграммой типа Γ в категории $\mathscr{B}$.

(10) $\operatorname{Mor} : \mathscr{C}^{op} \times \mathscr{C} \to \mathbf{Set}$. Пусть задан морфизм $(\alpha, \beta) \in \operatorname{Mor}_{\mathscr{C}^{op}\times\mathscr{C}}\big((X,Y),(X',Y')\big)$, т. е. пара морфизмов $\alpha : X' \to X$ и $\beta : Y \to Y'$ в категории $\mathscr{C}$. Тогда отображение

$$\operatorname{Mor}(\alpha,\beta) : \operatorname{Mor}(X,Y) \to \operatorname{Mor}(X',Y')$$

отображает морфизм $\varphi : X \to Y$ в композицию $\beta\varphi\alpha : X' \to Y'$.

Если $\mathscr{C}$ – категория R-модулей, где R – кольцо, то вместо категории множеств можно написать категорию R-модулей. Можно зафиксировать первый или второй аргумент.

В частности, $\operatorname{Mor}(\_, F) : (R\text{-}\mathbf{Mod})^{op} \to R\text{-}\mathbf{Mod}$ – контравариантный функтор на категории R-модулей. Обычно его действие на объектах обозначается звездочкой. Модуль $M^* = \operatorname{Mor}(M, R)$ называется двойственным (или сопряженным) к M. Если R – поле, а V – векторное пространство над F, то элементы V^* часто называются ковекторами.

(11) Предпучок на топологическом пространстве. Пусть X – топологическое пространство, а Ω – множество его открытых подмножеств упорядоченное по включению. Обозначим через $\mathscr{C}_X$ категорию, связанную с частично упорядоченным множеством Ω. Тогда предпучком объектов категории $\mathscr{B}$ на пространстве X называется контравариантный функтор из $\mathscr{C}_X$ в $\mathscr{B}$.

Другими словами, каждому открытому множеству в X предпучок ставит в соответствие объект категории $\mathscr{B}$, а вложению открытых множеств – морфизм в категории $\mathscr{B}$ в обратную сторону. Пучок – это предпучок, удовлетворяющий некоторым “условиям склейки”, которые выходят за рамки нашего беглого знакомства с теорией категорий.

Стандартный пример пучка множеств на топологическом пространстве X строится следующим образом. Каждому отрытому множеству A сопоставим множество непрерывных функций $\mathbb{R}^A$. Если $A \subseteq B$, то отображение $\mathbb{R}^B \to \mathbb{R}^A$ – это просто сужение функции $B \to \mathbb{R}$ на подмножество A.

Множество непрерывных функций $\mathbb{R}^A$ является коммутативным кольцом с 1 относительно поточечных операций. Обычно считают, что пучок из этого примера – это пучок коммутативных колец с 1. Естественно, $\mathbb{R}$ можно заменить на $\mathbb{C}$ или на что-то другое (правда тогда может оказаться, что множество функций не является кольцом), а свойство непрерывности на какое-нибудь другое свойство, скажем, дифференцируемость (если X – нормированное пространство).

Теперь мы можем привести еще один пример категории: категория категорий! Но, все же категории *всех* категорий не может быть, потому что мы не имеем право образовать класс всех категорий. Поэтому “категория категорий” – это на самом деле категория всех малых категорий. Морфизмами в этой категории являются функторы. Она обозначается **Cat**.

Философски, тот факт, что **Cat** знает только малые категории, не играет никакой роли, потому что категории из какого-то конкретного набора, который нас в данный момент интересует, всегда можно сделать малыми, изменив универсум, в котором мы работаем.

3. Естественные преобразования

Определение 3.1. Пусть $\mathscr{F}, \mathscr{G} : \mathscr{B} \to \mathscr{C}$ – функторы. Естественным преобразованием функторов $\eta : \mathscr{F} \to \mathscr{G}$ называется набор морфизмов $\eta_X \in \operatorname{Mor}(\mathscr{F}(X), \mathscr{G}(X))$ по всем объектам X категории $\mathscr{B}$, удовлетворяющих условию

$$\eta_Y \mathscr{F}(\alpha) = \mathscr{G}(\alpha)\eta_X$$

для любых объектов $X, Y \in \mathscr{B}$ и любого морфизма $\alpha \in \operatorname{Mor}(X, Y)$.

Последнее условие в определении означает коммутативность следующей диаграммы:

$$\begin{array}{ccc} \mathscr{F}(X) & \xrightarrow{\mathscr{F}(\alpha)} & \mathscr{F}(Y) \\ {\scriptstyle \eta_X}\downarrow & & {\scriptstyle \eta_Y}\downarrow \\ \mathscr{G}(X) & \xrightarrow{\mathscr{G}(\alpha)} & \mathscr{G}(Y). \end{array}$$

Функторы $\mathscr{F},\mathscr{G}:\mathscr{B}\to\mathscr{C}$ называются естественно изоморфными, если существует естественное преобразования $\eta:\mathscr{F}\to\mathscr{G}$ такое, что η_X является изоморфизмом для любого $X\in\mathscr{B}$. Очевидно, что в этом случае существует и обратное естественное преобразование $\eta^{-1}:\mathscr{G}\to\mathscr{F}$.

Категории $\mathscr{B}$ и $\mathscr{C}$ называются эквивалентными, если существуют функторы $\mathscr{F}:\mathscr{B}\to\mathscr{C}$ и $\mathscr{F}':\mathscr{C}\to\mathscr{B}$ такие, что композиция $\mathscr{F}\circ\mathscr{F}'$ естественно изоморфна $\mathrm{id}_{\mathscr{C}}$, а $\mathscr{F}'\circ\mathscr{F}$ естественно изоморфна $\mathrm{id}_{\mathscr{B}}$. Другими словами, существуют естественные изоморфизмы $X\cong\mathscr{F}\big(\mathscr{F}'(X)\big)$ и $Y\cong\mathscr{F}'\big(\mathscr{F}(Y)\big)$, где $X\in\mathscr{C}$, а $Y\in\mathscr{B}$. При этом функторы $\mathscr{F}$ и $\mathscr{F}'$ называются квазиобратными друг другу.

Примеры.

(1) Вложение в тождественный функтор: $M^*\hookrightarrow M$ (здесь M^* – *моноид* обратимых элементов), вложение коммутанта в группу и т.п.

(2) Тривиальные естественные изоморфизмы, например $(X\times Y)\times Z\cong X\times(Y\times Z)$.

(3) $\det_n:\mathrm{GL}_n\to\mathrm{GL}_1,\ \ A\mapsto\det A$.

(4) $\mathrm{Mor}(X,\mathrm{Mor}(Y,Z))\cong\mathrm{Mor}(X\times Y,Z)$ в категории множеств. Это называется "экспоненциальный закон для множеств" и обычно выражается формулой $(Z^Y)^X=Z^{X\times Y}$. При этом естественном отображении функция $f:X\to Z^Y$ переходит в функцию $g:X\times Y\to Z$, заданную формулой $g(x,y)=f(x)(y)$.

Аналог этого естественного изоморфизма выполнен в разных других категориях $\mathscr{C}$, если множество $\mathrm{Mor}(Y,Z)$ естественным образом превращается в объект категории $\mathscr{C}$. При этом вместо прямого произведения возникают другие универсальные конструкции.

(5) В категории векторных пространств над полем F множество $\mathrm{Mor}(Y,Z)$ имеет естественную структуру векторного пространства. Поэтому можно считать, что $\mathrm{Mor}_{F\text{-}\mathbf{Vect}}(\_\,,\_)$ является функтором $(F\text{-}\mathbf{Vect})^{op}\times F\text{-}\mathbf{Vect}\to F\text{-}\mathbf{Vect}$. Тогда $\mathrm{Mor}\big(X,\mathrm{Mor}(Y,Z)\big)$ естественно изоморфно пространству $\mathrm{Bil}(X\times Y,Z)$ билинейных отображений из $X\times Y$ в Z. В следующем параграфе мы определим тензорное произведение $X\otimes Y$ так, чтобы $\mathrm{Bil(X\times Y,Z)}$ было бы естественно изоморфно $\mathrm{Mor}(X\otimes Y,Z)$.

(6) Рассмотрим контравариантный функтор $V\mapsto V^*$ на категории векторных пространств из примера 10. Ясно, что его композиция с самим собой будет ковариантным функтором. Построим естественное преобразование η тождественного функтора в функтор $**$. Для этого для любого векторного пространства V необходимо определить линейное отображение $\eta_V:V\to V^{**}$. Для $x\in V$ положим $\eta_V(x)(f)=f(x)$. Проверка того, что такие отображения линейны, а η – естественное преобразование, является рутинной. Заметим, что отображение η_V является инъективным (нетрудно посчитать его ядро).

Пусть теперь $\mathscr{V}=F\text{-}\mathbf{Vect}_{f.d.}$ обозначает категорию конечномерных векторных пространств и их линейных отображений. Рассмотрим сужения функторов $*:\mathscr{V}\to\mathscr{V}^{op}$ и $**:\mathscr{V}\to\mathscr{V}$. Так как размерности пространств V, V^* и V^{**} совпадают, то построенные выше отображения η_V являются изоморфизмами векторных пространств. Это доказывает, что функтор $**:\mathscr{V}\to\mathscr{V}$ естественно изоморфен тождественному, а функтор $*$ – квазиобратен сам себе. В частности, категория конечномерных векторных пространств эквивалентна своей противоположной.

(7) Пусть Γ – ориентированный граф, $\mathscr{C}_\Gamma$ – категория достижимости в этом графе, а $\mathscr{B}$ – произвольная категория. Тогда естественное преобразование функторов $\mathscr{F}\to\mathscr{G}$, где $\mathscr{F},\mathscr{G}:\mathscr{C}_\Gamma\to\mathscr{B}$, – это просто морфизм соответствующих диаграмм (см. пример функторов номер 9).

(8) Пусть $\mathscr{F} : \mathscr{B} \to \mathscr{C}$ – функтор. Он определяет естественное преобразование функторов $\eta^{\mathscr{F}} : \mathrm{Mor}_{\mathscr{B}}(\_,\_) \to \mathrm{Mor}_{\mathscr{C}}\big(\mathscr{F}(\_),\mathscr{F}(\_)\big)$ по правилу: $\eta^{\mathscr{F}}_{(X,Y)}(\alpha) = \mathscr{F}(\alpha)$, где $\alpha \in \mathrm{Mor}(X,Y)$.

(9) Пусть $\mathscr{M}$ – категория матриц, а $\mathscr{V}$ – категория конечномерных векторных пространств (все над одним и тем же полем F). Пусть $\mathscr{F} : \mathscr{M} \to \mathscr{V}$ – функтор, который натуральному числу n ставит в соответствие пространство F^n, а матрице $A \in \mathrm{M}_{m\times n}(F) = \mathrm{Mor}_{\mathscr{M}}(n,m)$ – линейное отображение $L_A : F^n \to F^m$ умножения на матрицу A. Нетрудно проверить, что $\mathscr{F}$ действительно является функтором.

Если мы верим в аксиому выбора для классов, то можно построить и квазиобратный функтор. Зафиксируем в каждом векторном пространстве V базис b_V (для простоты в пространстве F^n выберем стандартный базис) и зададим функтор $\mathscr{G} : \mathscr{V} \to \mathscr{M}$, который векторному пространству сопоставляет его размерность, а линейному отображению – его матрицу в выбранных базисах. Тогда композиция $\mathscr{G} \circ \mathscr{F}$ просто тождественна (матрица отображения L_A в стандартных базисах равна A), а композиция в обратном порядке естественно изоморфна тождественному. Действительно, зададим естественное отображение $\eta : \mathscr{F} \circ \mathscr{G} \to \mathrm{id}_{\mathscr{V}}$ по правилу $\eta_V : F^{\dim V} \to V$ – это единственное линейное отображение, переводящее стандартный базис в выбранный базис пространства V. Условие естественности η – это коммутативность диаграммы:

$$\begin{array}{ccc} F^m & \longrightarrow & F^n \\ {\scriptstyle \eta_U}\big\downarrow & & \big\downarrow{\scriptstyle \eta_V} \\ U & \xrightarrow{L} & V \end{array}$$

где $m = \dim U$, $n = \dim V$, а верхняя стрелка $\mathscr{F}\big(\mathscr{G}(L)\big)$ – умножение на матрицу оператора L в выбранных базисах. А коммутативность эта выполнена просто по определению матрицы линейного отображения.

Если мы не верим в аксиому выбора для классов,[1] то надо рассмотреть категорию конечномерных векторных пространств с выбранными базисами, морфизмами в которой являются любые линейные отображения никак не связанные с выбранными базисами.

Функторы можно применять к морфизмам, что наводит на мысль, что можно применять их и набору морфизмов, в частности, к естественному преобразованию.

Предложение 3.2. *Пусть η – естественное преобразование функторов $\mathscr{F},\mathscr{G} : \mathscr{A} \to \mathscr{B}$, а $\mathscr{H} : \mathscr{B} \to \mathscr{C}$ – функтор. Тогда набор морфизмов $(\mathscr{H}\eta)_X := \mathscr{H}(\eta_X)$ является естественным преобразованием $\mathscr{H}\eta : \mathscr{H} \circ \mathscr{F} \to \mathscr{H} \circ \mathscr{G}$.*

Если $\mathscr{E}$ – функтор $\mathscr{C} \to \mathscr{A}$, то набор морфизмов $(\eta\mathscr{E})_X := \eta_{\mathscr{E}(X)}$ является естественным преобразованием $\eta\mathscr{E} : \mathscr{F} \circ \mathscr{E} \to \mathscr{G} \circ \mathscr{E}$.

Если η изоморфизм, то $\mathscr{H}\eta$ и $\eta\mathscr{E}$ – изоморфизмы.

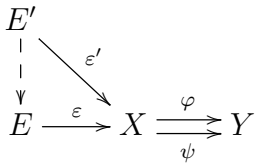
4. Эквалайзеры, произведения и универсальные квадраты

В этом параграфе мы изучим несколько важных универсальных конструкций, являющихся частым случаем (ко)пределов. Пусть $\alpha : A \to C$ и $\beta : B \to C$ – морфизмы в категории $\mathscr{C}$. Говорят, что β пропускается через α, если $\beta = \alpha\gamma$ для некоторого морфизма $\gamma : B \to A$.

Пусть $\varphi, \psi \in \mathrm{Mor}(X,Y)$ – морфизмы в категории $\mathscr{C}$. Морфизм $\varepsilon : E \to X$ называется *эквалайзером* пары (φ,ψ), если $\varphi\varepsilon = \psi\varepsilon$, и для любой $\varepsilon' : E' \to X$, обладающий свойством $\varphi\varepsilon' = \psi\varepsilon'$, пропускается через ε единственным образом. Это определение можно выразить коммутативной

[1] В отличие от аксиомы выбора для множеств, без которой никакой содержательной математики не получается, аксиома выбора для классов практически не используется, поэтому в некоторых аксиоматиках она отсутствует.

диаграммой



(обычно используется соглашение, что пути длины 1 из X в Y не должны быть равны на коммутативной диаграмме). Другими словами, рассмотрим категорию коммутативных диаграмм вида

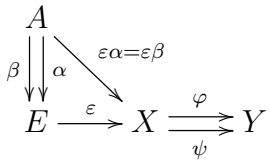
$$E \xrightarrow{\varepsilon} X \underset{\psi}{\overset{\varphi}{\rightrightarrows}} Y,$$

где X, Y, φ и ψ фиксированы, а E и ε меняются. Тогда эквалайзер – это финальный объект в этой категории. Как обычно, финальный объект единственный с точностью до единственного изоморфизма.

Двойственное понятие называется коэквалайзер. В категории множеств эквалайзер – это множество решений уравнения $\varphi(x) = \psi(x)$ вместе с отображением вложения. Коэквалайзер – каноническая проекция Y на фактормножество по наименьшему отношению эквивалентности, содержащему все пары $\big(\varphi(x), \psi(x)\big)$.

Предложение 4.1. *Эквалайзер всегда является мономорфизмом, а коэквалайзер – эпиморфизмом.*

Доказательство. Пусть $\varepsilon\alpha = \varepsilon\beta$ для некоторых морфизмов $\alpha, \beta : A \to E$ и эвалайзера $\varepsilon : E \to X$ морфизмов $\varphi, \psi : X \to Y$. Рассмотрим коммутативную диаграмму



По утверждению о единственности из определения эквалайзера получим, что $\alpha = \beta$. Как обычно, двойственное утверждение можно не доказывать (принцип “2 по цене 1”). □

Нетрудно проверить, что в категории множеств любой мономорфизм является эквалайзером каких-то двух морфизмов. В общем случае это не так, свойство “быть эквалайзером” существенно сильнее свойства “быть мономорфизмом”, что видно, в частности, из следующего утверждения.

Предложение 4.2. *Эквалайзер, являющийся эпиморфизмом – изоморфизм. Двойственно: коэквалайзер, являющийся мономорфизмом – изоморфизм.*

Доказательство. Если ε является эквалайзером морфизмов $\varphi, \psi : X \to Y$, то $\varphi\varepsilon = \psi\varepsilon$, а если ε – эпиморфизм, то из этого равенства следует, что $\varphi = \psi$. Но в этом случае легко видеть, что id_X удовлетворяет определению эквалайзера. Так как два разных эквалайзера отличаются друг от друга на изоморфизм, то ε – изоморфизм. □

Следующие универсальные конструкции, которые мы изучим – произведение и копроизведение. Пусть $A, B \in \mathscr{C}$. Рассмотрим категорию диаграмм вида $A \leftarrow C \to B$. Финальный объект в этой категории называется произведением объектов A, B и часто обозначается $A \times B$. Обратите внимание, что произведение – это объект вместе с морфизмами в A и в B, хотя, допуская вольность речи, про морфизмы часто забывают. В категории множеств, как и во многих других *конкретных* категориях[2], произведение – это действительно декартово произведение (морфизмы – проекции).

[2]Категория называется конкретной, если из нее существует строгий “забывающий” функтор в категорию множеств.

Упражнение 4.3. Пусть $A, B, C \in \mathscr{C}$. Докажите, что следующие условия эквивалентны.

(1) $C = A \times B$.
(2) Существует естественная биекция $\mathrm{Mor}(D, A) \times \mathrm{Mor}(D, B) \to \mathrm{Mor}(D, C)$.

Двойственное понятие, копроизведение, – это инициальный объект в категории диаграмм вида $A \to C \leftarrow B$ с фиксированными A и B. В категории множеств копроизведение – это дизъюнктное объединение, в других конкретных категориях – нечто, что "свободно" порождается A и B, конкретика в следующем параграфе.

Масса полезных конструкций в математике возникает, как универсальные объекты в категории квадратов с тремя фиксированными вершинами. Пусть $\mathscr{C}$ – категория, $A, B, C \in \mathscr{C}$, $\alpha \in \mathrm{Mor}(A, C)$, $\beta \in \mathrm{Mor}(B, C)$. Обозначим через $\mathscr{S}$ категорию коммутативных квадратов вида

$$\begin{array}{ccc} \bullet & \longrightarrow & B \\ \downarrow & & \downarrow{\scriptstyle \beta} \\ A & \xrightarrow{\alpha} & C \end{array}$$

(см. двойственный пример (11) из параграфа 1). Тогда финальный объект в этой категории называется пулбэком морфизмов α и β или, чаще, пулбэком диаграммы

$$A \xrightarrow{\alpha} C \xleftarrow{\beta} B.$$

Двойственным образом определяется пушаут, т. е. пушаут в категории $\mathscr{C}$ – это пулбэк в категории $\mathscr{C}^{op}$.

Предложение 4.4. *Пулбэк диаграммы $A \xrightarrow{\alpha} C \xleftarrow{\beta} B$ – это эквалайзер композиций $A \times B \xrightarrow{\pi_A} A \xrightarrow{\alpha} C$ и $A \times B \xrightarrow{\pi_B} B \xrightarrow{\beta} C$.*

Аналогично, пушаут диаграммы $A \leftarrow C \rightarrow B$ – это коэквалайзер двух естественных морфизмов из C в копроизведение A и B.

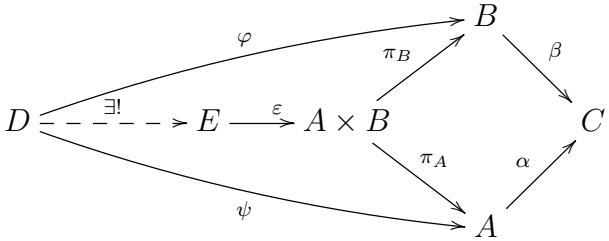
Доказательство. Рассмотрим коммутативную диаграмму

$$\begin{array}{ccc} D & \xrightarrow{\varphi} & B \\ {\scriptstyle \psi}\downarrow & & \downarrow{\scriptstyle \beta} \\ A & \xrightarrow{\alpha} & C \end{array}$$

По универсальному свойству произведения существует единственный морфизм $\theta : D \to A \times B$, чьи композиции с проекциями π_A и π_B равны ψ и φ соответственно. Получаем диаграмму

$$D \xrightarrow{\theta} A \times B \underset{\beta\pi_B}{\overset{\alpha\pi_A}{\rightrightarrows}} C,$$

которая является коммутативной, так как $\alpha\pi_A\theta = \alpha\psi = \beta\varphi = \beta\pi_B\theta$. Следовательно, существует морфизм из D в эквалайзер (E, ε) морфизмов $\alpha\pi_A$ и $\beta\pi_B$. Единственность морфизма из D в эквалайзер сразу следует из единственности в определениях произведения и эквалайзера.



(на последней диаграмме правый квадрат некоммутативен!). □

Упражнение 4.5. Рассмотрим коммутативную диаграмму

$$\begin{array}{ccccc} A & \longrightarrow & B & \longrightarrow & C \\ \downarrow & & \downarrow & & \downarrow \\ D & \longrightarrow & E & \longrightarrow & F \end{array}$$

Предположим, что квадрат $BCFE$ является пулбэком. Докажите, что $ABED$ – пулбэк тогда и только тогда, когда $ACFD$ пулбэк. Приведите пример, когда $ABED$ и $ACFD$ – пулбэки, а $BCFE$ – нет.

4.1. (ко)Ядро как (ко)эквалайзер или пулбэк (пушаут). Предположим, что в некоторой категории $\mathscr{C}$ инициальный объект $*$ совпадает с финальным, как например в в категории групп, R-модулей, колец без 1 или пунктированных множеств. В этом случае он обычно называется нулевым объектом. Тогда для любых двух объектов $A, B \in \mathscr{C}$ существуют единственные морфизмы $A \to * \to B$. Композиция этих морфизмов называется нулевым морфизмом из A в B. Ядром морфизма $\varphi : A \to B$ называется эквалайзер φ и нулевого морфизма. Как вы знаете, эквалайзер состоит из объекта и морфизма. Объект "ядро" будет обозначаться через Ker, а морфизм – ker (такая договоренность используется в литературе, но далеко не во всех текстах).

Другими словами, ядро – это пулбэк диаграммы $A \xrightarrow{\varphi} B \leftarrow *$. Действительно, нетрудно видеть, что $A \times * \cong A$ и, следовательно, по предложению 4.4 пулбэк равен эквалайзеру из предыдущего абзаца.

Двойственно, коядро $\operatorname{coker}\varphi$ – это коэквалайзер φ и нулевого морфизма. В категории модулей коядро – это фактор по образу, в категории групп – фактор по нормальному замыканию образа, в категории пунктированных множеств – склейка образа функции φ в одну точку.

$$\begin{array}{ccc} \operatorname{Ker}\varphi & \longrightarrow & * \\ {\scriptstyle \ker\varphi}\downarrow & & \downarrow \\ A & \xrightarrow{\varphi} & B \end{array} \qquad\qquad \begin{array}{ccc} A & \longrightarrow & * \\ {\scriptstyle \varphi}\downarrow & & \downarrow \\ B & \xrightarrow{\operatorname{coker}\varphi} & \operatorname{Coker}\varphi \end{array}$$

Упражнение 4.6. Определите образ морфизма в произвольной категории с помощью универсального свойства (для категории множеств он должен совпадать с вложением образа в множество значений функции). Формулируйте двойственное определение.

4.2. Удвоение кольца вдоль идеала, группы вдоль нормальной подгруппы. Естественно, не любой идеал кольца является ядром ретракции. Следующая конструкция позволяет исправить это неудобство. Удвоением кольца R вдоль (двустороннего) идеала I называется пулбэк диаграммы $R \to R/I \leftarrow R$. Другими словами, удвоение кольца вдоль идеала – это подкольцо в прямой сумме $R \oplus R$, состоящее из пар (r, s), где $r \equiv s \mod I$. Удвоение R вдоль I обозначается $R \ltimes I$.

Теперь мы имеем гомоморфизмы колец $R \to R \ltimes I$, $r \mapsto (r, r)$ и $\pi : R \ltimes I \to R$, $(r, s) \to r$ с тождественной композицией. Следовательно, для любого функтора $\mathscr{F} : \mathbf{Ring} \to \mathbf{Grp}$ имеем $\mathscr{F}(R \ltimes I) \cong \mathscr{F}(R) \ltimes K$, где $K = \operatorname{Ker}\mathscr{F}(\pi)$. Стандартный прием: если функтор определен на категории колец, а хочется определить его значение на идеалах, то по определению полагают $\mathscr{F}(R, I) = \operatorname{Ker}\mathscr{F}(\pi)$. При этом расширенное таким образом отображение становится функтором на категории пар (R, I), где I – идеал в R. Морфизмами в категории пар являются гомоморфизмы колец, отображающие идеал в идеал.

Например, для функтора $\mathscr{F} = \mathrm{GL}_n$ получим, что $\mathrm{GL}_n(R, I)$ состоит из обратимых матриц, сравнимых с единичной по модулю I. Заметим, что $\mathrm{GL}_n(R, I)$ не зависит от кольца R, потому что при вычислении произведения таких матриц используются только операции сложения и умножения внутри идеала I, рассматриваемого как кольцо без 1.

Аналогично можно определить удвоение группы вдоль нормальной подгруппы, как пулбэк диаграммы $G \to G/H \leftarrow G$. Это удвоение будет изоморфно полупрямому произведению $G \ltimes H$, где G действует на H сопряжениями.

4.3. Копроизведение в категории множеств. Для непересекающихся множеств копроизведение – это объединение. Однако, если $A \cap B \neq \varnothing$, то в объединении склеиваются элементы из A и B, чего, очевидно, не может быть в копроизведении. Так что копризведение – это так называемое дизъюнктное объединение, которое может быть построено как подмножество в $A \cup B \times \{0,\, 1\}$ состоящее из пар $(a, 0)$ и $(b, 1)$ по всем $a \in A$ и $b \in B$ с очевидными отображениями из A и B. Дизъюнктное объединение обозначается $A \coprod B$.

4.4. Копроизведение в категории модулей. Обычно, конструкция копроизведения в конкретных категориях сложнее, чем произведения. Но с модулями это не так. А именно, копроизведение в категории модулей, также как и произведение, равно прямой сумме (гомоморфизмы – канонические вложения). Действительно, задать диаграмму $M \to P \leftarrow N$ это то же самое, что задать гомоморфизм $M \oplus N \to P$, при этом этот гомоморфизм делает соответствующую диаграмму коммутативной.

Соответственно, пушаут диаграммы $M \xleftarrow{\varphi} Q \xrightarrow{\psi} N$ – это фактормодуль прямой суммы $M \oplus N$ по подмодулю $\{\big(\varphi(q), -\psi(q)\big) \mid q \in Q\}$.

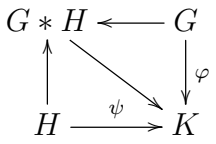
4.5. Копроизведение в категории групп. Копроизведение в категории групп называется свободным произведением групп. Свободное произведение групп G и H обозначается $G * H$ и строится следующим образом. На множестве W слов в алфавите $G \coprod H$ определим отношение $\sim$, как наименьшее отношение эквивалентности, содержащее все пары вида $(w_1 1_G w_2, w_1 w_2)$, $(w_1 1_H w_2, w_1 w_2)$, $(w_1 x y w_2, w_1 z w_2)$, где $w_1, w_2 \in W$, $x, y, z \in G$ или $x, y, z \in H$, и $xy = z$. Множество $W/\!\sim$ с операцией конкатенации слов – это и есть группа $G*H$. Также как и для свободной группы проверяется, что класс эквивалентности конкатенации не зависит от выбора представителей, а также, что в любом классе эквивалентности есть ровно одно самое короткое "редуцированное слово". В редуцированном слове буквы из G и H чередуются и не равны 1. Отображения $G \to G * H$ и $H \to G * H$ очевидны: элементам сопоставляются соответствующие однобуквенные слова. Обычно считают, что $G, H \subseteq G * H$, а отображения являются вложениями. Заметим, что множество $G \coprod H$ порождает группу $G * H$.

Для диаграммы

$$G \xrightarrow{\varphi} K \xleftarrow{\psi} H \tag{33}$$

определим отображение $\theta : W \to K$ по правилу $\theta(x_1 \dots x_n) = \varepsilon(x_1) \dots \varepsilon(x_n)$, где $x_i \in G \cup H$, а $\varepsilon : G \cup H \to K$ – отображение, совпадающее с φ на G и с ψ на H. Нетрудно видеть, что θ отображает каждый класс эквивалентности в один элемент, и поэтому индуцирует отображение $W/\!\sim = G * H \to K$, которое очевидно является гомоморфизмом.

Из коммутативности диаграммы



следует, что этот гомоморфизм однозначно определен на $G \coprod H$. Но гомоморфизм однозначно определяется своими значениями на множестве образующих группы, отсюда получаем единственность гомоморфизма $G * H \to K$.

Можно дать и другую конструкцию свободного произведения, с которой проще доказать, что она является свободным произведением, но сложнее работать в приложениях. А именно, определим $G * H$ как факторгруппу свободной группы с образующими $G \coprod H$ по наименьшей нормальной подгруппе N, содержащей все элементы вида xyz^{-1}, где $x, y, z \in G$ или $x, y, z \in H$, и $xy = z$. Действительно, для диаграммы (33) по универсальному свойству свободной группы существует гомоморфизм групп $\alpha : F_{G \coprod H} \to K$. Так как φ и ψ – гомоморфизмы, то все упомянутые выше элементы лежат в ядре α, следовательно, α пропускается через факторизацию по N. Естественные

отображения $G, H \to F_{G \coprod H}/N$ теперь уже являются гомоморфизмами, так что существование доказано. Единственность следует из единственности выбранных отображений.

$$G, H \to K \implies G \coprod H \to K \implies F_{G \coprod H} \to K \implies F_{G \coprod H}/N = G * H \to K.$$

Пушаут в категории групп называется амальгамированным (свободным) произведением. Теперь, зная, что такое копроизведения и коэквалайзеры, нетрудно его построить, пользуясь предложением 4.4. А именно, пушаут диаграммы $G \xleftarrow{\lambda} A \xrightarrow{\mu} H$ – это факторгруппа свободного произведения $G * H$ по наименьшей нормальной подгруппе, содержащей элементы $\lambda(a)\mu(a)^{-1}$ по всем $a \in A$.

Амальгамированное произведение G и H над A (или свободное произведение G и H с амальгамированной подгруппой A, если отображения λ и μ являются вложениями) обозначается через $G *_A H$. Как обычно, подразумевается, что гомоморфизмы λ и μ известны из контекста.

4.6. Копроизведение в категории коммутативных колец с 1. В категории **CRing** для разнообразия построим сразу пушаут диаграммы $A \xleftarrow{\varphi} R \xrightarrow{\psi} B$. Кольца A и B являются R-алгебрами, в частности, R-модулями. Поэтому существует тензорное произведение $A \otimes_R B$. Мы хотим превратить модуль $A \otimes_R B$ в кольцо так, чтобы канонические отображения $A, B \to A \otimes_R B$ стали бы гомоморфизмом алгебр. Это можно сделать коструктивно: на разложимых тензорах определить $(a \otimes b) \cdot (a' \otimes b') = aa' \otimes bb'$, а дальше продолжить по линейности. Но тогда надо проверять корректность, потому что элемент $A \otimes_R B$ не единственным образом представляется в виде суммы разложимых.

Альтернативно, умножение в A является R-билинейным, следовательно индуцирует линейное отображение $A \otimes_R A \to A$. Аналогично получаем линейное отображение $B \otimes_R B \to B$. Так как тензорное произведение является функтором R-**Mod** $\times$ R-**Mod** $\to$ R-**Mod**, эти отображения индуцируют линейное отображение $(A \otimes_R A) \otimes_R (B \otimes_R B) \to A \otimes_R B$. Так как $(A \otimes_R A) \otimes_R (B \otimes_R B)$ естественно изоморфно $(A \otimes_R B) \otimes_R (A \otimes_R B)$, то можно рассматривать последнее отображение, как отображение $(A \otimes_R B) \otimes_R (A \otimes_R B) \to A \otimes_R B$, которое соответствует билинейному отображению $(A \otimes_R B) \times (A \otimes_R B) \to A \otimes_R B$, а это и есть умножение в R-алгебре $A \otimes_R B$.

Пусть теперь

$$\begin{array}{ccc} R & \longrightarrow & A \\ \downarrow & & \downarrow{\scriptstyle \alpha} \\ B & \xrightarrow{\beta} & C \end{array}$$

коммутативная диаграмма в категории колец (это то же самое, что и диаграмма $A \xrightarrow{\alpha} C \xleftarrow{\beta} B$ в категории коммутативных R-алгебр). Пользуясь функториальностью тензорного произведения, зададим R-линейное отображение $A \otimes_R B \to C \otimes_R C \to C$, где вторая стрелка индуцирована умножением в C. Проверка того, что это отображение является гомоморфизмом R-алгебр и единственно оставляется в качестве упражнения.

Следующее утверждение показывает, на каком универсальном примере можно проверять, является ли морфизм эпиморфизмом. Формулировка двойственного утверждения оставляется читателю в качестве упражнения.

Упражнение 4.7. Предположим, что коммутативный квадрат

$$\begin{array}{ccc} A & \xrightarrow{\theta} & B \\ {\scriptstyle \theta}\downarrow & & \downarrow{\scriptstyle \psi} \\ B & \xrightarrow{\varphi} & C \end{array}$$

является пушаутом. Тогда следующие условия эквивалентны.

(1) θ – эпиморфизм.

(2) φ – изоморфизм.

(3) ψ – изоморфизм.
(4) $\varphi = \psi$.

Решение. (1) $\Longrightarrow$ (2). Если θ эпиморфизм, то для любого коммутативного квадрата

$$\begin{array}{ccc} A & \xrightarrow{\theta} & B \\ {\scriptstyle\theta}\downarrow & & \downarrow{\scriptstyle\beta} \\ B & \xrightarrow{\alpha} & D \end{array}$$

$\alpha = \beta$. Поэтому $C = B$, $\varphi = \psi = \mathrm{id}_B$ будут удовлетворять определению пушаута. Так как пушаут определен с точностью до изоморфизма, то $\varphi = \psi$ – изоморфизмы.

(2) $\Longrightarrow$ (4). По определению пушаута для тождественного морфизма id_B существует $\varepsilon : C \to B$ такой, что $\varepsilon\varphi = \varepsilon\psi = \mathrm{id}_B$. Если φ – изоморфизм, то $\varepsilon = \varphi^{-1}$ также является изоморфизмом (обратные слева и справа должны совпадать). Следовательно, на него можно сократить, откуда $\varphi = \psi$.

(4) $\Longrightarrow$ (1). Предположим, что $\varphi = \psi$. Пусть $\alpha, \beta : B \to D$ такие морфизмы, что $\alpha\theta = \beta\theta$. По универсальному свойству пушаута существует единственный морфизм $\gamma : C \to D$ такой, что $\alpha = \gamma\varphi = \beta$. Следовательно, θ – эпиморфизм.

Так как φ и ψ входят в условие симметрично, то импликации (1) $\Longrightarrow$ (3) $\Longrightarrow$ (4) доказывать не надо. □

Как это ни удивительно, но доказательство отрицательного результата, т. е. того, что морфизм не является эпиморфизмом, часто очень сложно доказать, пользуясь последним предложением. Проще оказывается построить не универсальный пример морфизмов $\alpha \neq \beta$, для которых $\alpha\theta = \beta\theta$. Так делается, например, для доказательства того, что в категории групп несюръективное отображение не может быть эпиморфизмом.

5. Сопряженные функторы

5.1. Категория запятой. Существенная часть универсальных конструкций в алгебре возникает как инициальный объект в “категории запятой” (comma category).

Пусть $\mathscr{F} : \mathscr{B} \to \mathscr{C}$ – функтор, а $X \in \mathscr{C}$. Построим категорию $(X \downarrow \mathscr{F})$ следующим образом.[3] Объекты этой категории – пары (B, φ), где $B \in \mathscr{B}$, а $\varphi \in \mathrm{Mor}_{\mathscr{C}}(X, \mathscr{F}(B))$. Морфизм между (B, φ) и (B', φ') – это морфизм $\alpha \in \mathrm{Mor}_{\mathscr{B}}(B, B')$ такой, что $\mathscr{F}(\alpha)\varphi = \varphi'$. Композиция морфизмов – это их композиция в категории $\mathscr{B}$.

Аналогично определяется категория $(\mathscr{F} \downarrow X)$. А именно, объекты – это пары (B, φ), где $B \in \mathscr{B}$, а $\varphi \in \mathrm{Mor}_{\mathscr{C}}(\mathscr{F}(B), X)$, а морфизмы – $\alpha \in \mathrm{Mor}_{\mathscr{B}}(B, B')$ такие, что $\varphi = \varphi'\mathscr{F}(\alpha)$.

5.2. Сопряженные функторы.

Определение 5.1. Пусть $\mathscr{F} : \mathscr{B} \to \mathscr{C}$ и $\mathscr{G} : \mathscr{C} \to \mathscr{B}$ – функторы. $\mathscr{G}$ называется левым сопряженным к $\mathscr{F}$, если существует естественный изоморфизм между функторами $\mathrm{Mor}_{\mathscr{C}}\big(\_, \mathscr{F}(\_)\big)$ и $\mathrm{Mor}_{\mathscr{B}}\big(\mathscr{G}(\_), \_\big)$ (как нетрудно заметить, эти функторы действуют из $\mathscr{C}^{op} \times \mathscr{B}$ в **Set**). При этом $\mathscr{F}$ называется правым сопряженным к $\mathscr{G}$.

Прежде, чем изучать свойства сопряженных функторов, приведем несколько уже известных нам примеров.

Примеры.

5.1. Левые сопряженные к забывающим функторам $\mathscr{F} :$ **Mon**, **Grp**, R-**Mod**, R-**Alg** $\to$ **Set** – это сопоставление множеству X свободного объекта, построенного на этом множестве. На примере категории групп: каждой функции из множества X в $\mathscr{F}(G)$ сопоставляется

[3] Вероятно, изначально эта категория обозначалась через $(X, \mathscr{F})$, и запятая, стоящая между $\mathscr{F}$ и X, дала этим категориям такое странное название.

единственный гомоморфизм групп $F_X \to G$, и обратно: каждому гомоморфизму $F_X \to G$ сопоставляется его композиция с вложением $X \to F_X$. Таким образом,

$$\mathrm{Mor}_{\mathbf{Set}}(X, \mathscr{F}(G)) \cong \mathrm{Mor}_{\mathbf{Grp}}(\mathscr{G}(X), G),$$

где $\mathscr{G}$ – функтор, сопоставляющий множеству свободную группу, построенную на нем.

5.2. Левый сопряженный к забывающему функтору $\mathscr{F} : R\text{-}\mathbf{Alg} \to \mathbf{Mon}$.

$$\mathrm{Mor}_{\mathbf{Mon}}(M, \mathscr{F}(A)) \cong \mathrm{Mor}_{R\text{-}\mathbf{Alg}}(RM, A),$$

где RM – это то, что мы хотим построить. По аналогии с предыдущим пунктом, нам нужна R-алгебра, содержащая M с минимумом соотношений (чем больше соотношений в RM, тем меньше гомоморфизмов из RM). Возьмем в качестве RM свободный R-модуль с базисом M и определим умножение при помощи умножения, заданного в M:

$$(\sum_i r_i m_i)(\sum_j s_j n_j) = \sum_{i,j} (r_i s_j)(m_i n_j).$$

Легко проверить, что RM теперь является R-алгеброй. Ясно, что у нас есть гомоморфизм моноидов $M \to RM$. Гомоморфизму R-алгебр $RM \to A$ соответствует гомоморфизм моноидов $M \to RM \to A$. Обратно, если есть гомоморфизм моноидов $M \to A$ то он единственным образом превращается в гомоморфизм R-модулей $RM \to A$, который при ближайшем рассмотрении оказывается гомоморфизмом R-алгебр.

Аналогично определяется левый сопряженный к функтору "группа обратимых элементов R-алгебры". Группе G он сопоставляет *групповую алгебру* RG (G рассматривается как моноид, но так как это группа, то она отображается в группу обратимых элементов).

5.3. Единица сопряжения. Можно заметить, что каждый раз у нас появляется (как бы ниоткуда) морфизм $X \to \mathscr{G}(X)$ (на самом деле, $X \to \mathscr{F}\big(\mathscr{G}(X)\big)$) в категории с менее структурированными объектами, где $\mathscr{G}$ – левый сопряженный к забывающему. Оказывается, что это совсем не случайно и имеет место для любой пары сопряженных функторов.

По определению сопряженных функторов

$$\mathrm{Mor}_{\mathscr{C}}\big(X, \mathscr{F}(Y)\big) \cong \mathrm{Mor}_{\mathscr{B}}\big(\mathscr{G}(X), Y\big).$$

Подставим $Y = \mathscr{G}(X)$ и возьмем тождественный морфизм в правой части равенства. Ему будет соответствовать единственный морфизм $f_X : X \to \mathscr{F}\big(\mathscr{G}(X)\big)$ в левой части. Из естественности биекции в вынесенной формуле следует что набор морфизмов f_X по всем $X \in \mathscr{C}$ является естественным преобразованием функторов $\mathrm{id} \to \mathscr{F} \circ \mathscr{G}$. Это естественное преобразование называется *единицей сопряжения*.

Аналогично, подставляя $X = \mathscr{F}(Y)$ и беря тожественный морфизм в левой части, получаем соответствующий ему естественный морфизм $\mathscr{G}\big(\mathscr{F}(Y)\big) \to Y$, который называется *коединицей сопряжения*.

Теорема 5.2. *Для функтора $\mathscr{F} : \mathscr{B} \to \mathscr{C}$ существует левый сопряженный тогда и только тогда, когда для любого $X \in \mathscr{C}$ существует инициальный объект в категории $\mathscr{M}_X$, определенной следующим образом:*

- $\mathrm{Obj}\,\mathscr{M}_X = \big\{(Y, f) \mid Y \in \mathscr{B},\ f \in \mathrm{Mor}_{\mathscr{C}}\big(X, \mathscr{F}(Y)\big)\big\}$;
- $\mathrm{Mor}\big((Y, f), (Z, g)\big) = \big\{h \in \mathrm{Mor}_{\mathscr{B}}(Y, Z) \mid \mathscr{F}(h)f = g\big\}$;
- *композиция морфизмов – это их композиция в $\mathscr{B}$.*

Этот инициальный объект и будет единицей сопряжения.

Доказательство. Пусть $\mathscr{G}$ – левый сопряженный к $\mathscr{F}$, f_X – единица сопряжения, а

$$\eta : \mathrm{Mor}_{\mathscr{C}}\big(\_, \mathscr{F}(\_)\big) \longrightarrow \mathrm{Mor}_{\mathscr{B}}\big(\mathscr{G}(\_), \_\big)$$

– естественный изоморфизм. Докажем, что пара $(\mathscr{G}(X), f_X)$ является инициальным объектом в категории $\mathscr{M}_X$. Действительно, пусть $(Y, f) \in \mathscr{M}_X$. Положим $g = \eta_{X,Y}(f) : \mathscr{G}(X) \to Y$. Рассмотрим коммутативную диаграмму, связанную с естественным преобразованием η^{-1} и морфизмом $(\mathrm{id}_X, g) : (X, \mathscr{G}(X)) \to (X, Y)$:

$$\begin{array}{ccc} \mathrm{Mor}_{\mathscr{C}}(X, \mathscr{F}(\mathscr{G}(X))) & \xrightarrow{\varphi \mapsto \mathscr{F}(g)\varphi} & \mathrm{Mor}_{\mathscr{C}}(X, \mathscr{F}(Y)) \\ {\scriptstyle \eta^{-1}_{X,\mathscr{G}(X)}}\uparrow & & {\scriptstyle \eta^{-1}_{X,Y}}\uparrow \\ \mathrm{Mor}_{\mathscr{B}}(\mathscr{G}(X), \mathscr{G}(X)) & \xrightarrow{\psi \mapsto g\psi} & \mathrm{Mor}_{\mathscr{B}}(\mathscr{G}(X), Y) \end{array}$$

Запишем условие коммутативности, примененное к тождественному морфизму $\psi = \mathrm{id}_{\mathscr{G}(X)}$:

$$\eta^{-1}_{X,Y}(g) = \mathscr{F}(g)\eta^{-1}_{X,\mathscr{G}(X)}(\mathrm{id}_{\mathscr{G}(X)}).$$

Посмотрев на определение g видим, что в левой части стоит f. По определению f_X получаем

$$f = \mathscr{F}(g)f_X,$$

что и означает, что g является морфизмом $(\mathscr{G}(X), f_X) \to (Y, f)$ в категории $\mathscr{M}_X$.

Пусть g' другой морфизм $(\mathscr{G}(X), f_X) \to (Y, f)$ в категории $\mathscr{M}_X$. Это означает, что $f = \mathscr{F}(g')f_X$. Заменив на коммутативном квадрате g на g', получим

$$\eta^{-1}_{X,Y}(g') = \mathscr{F}(g')\eta^{-1}_{X,\mathscr{G}(X)}(\mathrm{id}_{\mathscr{G}(X)}) = \mathscr{F}(g')f_X = f,$$

то есть $g' = \eta_{X,Y}(f) = g$.

Обратно, пусть $(\mathscr{G}(X), f_X)$ – инициальный объект в категории $\mathscr{M}_X$. Здесь $\mathscr{G}$ пока что просто отображение $\mathrm{Obj}\,\mathscr{C} \to \mathrm{Obj}\,\mathscr{B}$. Сейчас мы определим действие $\mathscr{G}$ на морфизмах и докажем, что он и есть функтор, сопряженный к $\mathscr{F}$. Если $\alpha \in \mathrm{Mor}(X, X')$, то $f_{X'}\alpha \in \mathrm{Mor}(X, \mathscr{F}(\mathscr{G}(X')))$. По универсальному свойству существует единственный морфизм $\gamma : \mathscr{G}(X) \to \mathscr{G}(X')$ такой, что $f_{X'}\alpha = \mathscr{F}(\gamma)f_X$. Положим $\mathscr{G}(\alpha) = \gamma$. Таким образом, по опрделению $\mathscr{G}$ имеем

$$f_{X'}\alpha = \mathscr{F} \circ \mathscr{G}(\alpha)f_X.$$

Нетрудно видеть, что $\mathscr{G}(\alpha\beta) = \mathscr{G}(\alpha)\mathscr{G}(\beta)$, а $\mathscr{G}(\mathrm{id}) = \mathrm{id}$. Таким образом $\mathscr{G}$ является функтором $\mathscr{C} \to \mathscr{B}$.

Определим функцию

$$\eta_{X,Y} : \mathrm{Mor}_{\mathscr{C}}\big(X, \mathscr{F}(Y)\big) \to \mathrm{Mor}_{\mathscr{B}}\big(\mathscr{G}(X), Y\big)$$

по правилу: для $f : X \to \mathscr{F}(Y)$ морфизм $\eta_{X,Y}(f)$ – это тот единственный морфизм, для которого $f = \mathscr{F}\big(\eta_{X,Y}(f)\big)f_X$. Обратное отображение строится очевидным образом: для $g \in \mathrm{Mor}_{\mathscr{B}}\big(\mathscr{G}(X), Y\big)$ положим $\eta^{-1}_{X,Y}(g) = \mathscr{F}(g)f_X$. Поэтому $\eta_{X,Y}$ – биекция (изоморфизм в категории множеств).

Докажем, что этот изоморфизм является естественным. Пусть $\alpha : X' \to X$ и $\beta : Y \to Y'$ – морфизмы в категориях $\mathscr{C}$ и $\mathscr{B}$ соответственно. Рассмотрим диаграмму

$$\begin{array}{ccc} \mathrm{Mor}_{\mathscr{B}}(\mathscr{G}(X), Y) & \xrightarrow{\psi \mapsto \beta\psi\mathscr{G}(\alpha)} & \mathrm{Mor}_{\mathscr{B}}(\mathscr{G}(X'), Y') \\ {\scriptstyle \eta^{-1}_{X,Y}}\uparrow & & {\scriptstyle \eta^{-1}_{X',Y'}}\uparrow \\ \mathrm{Mor}_{\mathscr{C}}(X, \mathscr{F}(Y)) & \xrightarrow{\varphi \mapsto \mathscr{F}(\beta)\varphi\alpha} & \mathrm{Mor}_{\mathscr{C}}(X', \mathscr{F}(Y')) \end{array}$$

Морфизм $\psi : \mathscr{G}(X) \to Y$ переходит направо в $\beta\psi\mathscr{G}(\alpha)$ и далее, вниз, в

$$\mathscr{F}\big(\beta\psi\mathscr{G}(\alpha)\big)f'_X = \mathscr{F}(\beta)\mathscr{F}(\psi)\mathscr{F} \circ \mathscr{G}(\alpha)f'_X.$$

Этот же морфизм вниз переходит в $\mathscr{F}(\psi)f_X$ и далее, направо, в

$$\mathscr{F}(\beta)\mathscr{F}(\psi)f_X\alpha.$$

Из определения действия $\mathscr{G}$ на морфизмах следует, что диаграмма коммутативна, что завершает доказательство. □

Упражнение 5.3. Сформулируйте аналогичное утверждение для правых сопряженных.

Теперь, зная, что для нахождения левого (правого) сопряженного нам надо найти соответствующий универсальный морфизм, продолжим приводить примеры.

5.4. Примеры.

(3) Топологические примеры. Левый сопряженный к забывающему функтору **Top** $\to$ **Set** – дискретная топология на данном множеством. Правый сопряженный к тому же функтору – антидискретная топология.

(4) Свободные объекты конкретных категорий – левые сопряженные к забывающим. Пусть X – множество, а R – коммутативное кольцо с 1. Пусть $\mathscr{C} =$ **Mon**, **Grp**, R-**Mod** или R-**Alg** (в принципе можно рассмотреть любую конкретную категорию, просто конструкции будут разные, а постановка задачи одинаковая). Рассмотрим категорию $(X \downarrow \mathscr{F}$, где $\mathscr{F}$ – забывающий функтор из $\mathscr{C}$ в категорию множеств. Инициальный объект такой категории называется *свободным объектом категории* $\mathscr{C}$*, порожденным* X. По теореме 5.2 функтор, сопоставляющий множеству X свободный объект, является левым сопряженым к забывающему.

Конструкции свободной группы и свободного R-модуля мы изучали, сейчас мы (не вникая в детали) построим свободный моноид M_X и свободную R-алгебру, порожденные X. Алгебра обычно обозначается $R\langle X\rangle$ и называется кольцом некоммутирующих многочленов от X.

Свободный моноид – это просто множество слов в алфавите X, включая пустое слово, с операцией конкатенации. Ясно, что как только мы знаем образы элементов множества X, то мы знаем и образы всех слов в алфавите X под действием гомоморфизма моноидов. Так строится единственный гомоморфизм из свободного моноида в произвольный моноид M по заданной функции $X \to M$.

Свободная R-алгебра $R\langle X\rangle$ – это свободный R-модуль с базисом M_X, умножение на котором задано очевидным образом. Действительно, на базисе умножение уже задано, а дальше оно единственным образом продолжается по билинейности.

(5) Еще один пример левого сопряженного к забывающему. Пусть $\mathscr{F} :$ **Cat** $\to$ **Grph** – функтор, который забывает про композицию морфизмов в категории. Найти левый сопряженный к нему – значит построить универсальную категорию, в которую (точнее в $\mathscr{F}$ от которой) вкладывается граф. Каждой вершине должен соответствовать объект категории, каждой стрелке – морфизм, каждому пути – композиция морфизмов. Следовательно, эта категория – категория путей графа, которая называется еще свободной категорией графа.

(6) Левые сопряженные к функторам вложения. Первый пример – вложение **Ab** $\to$ **Grp**. По гомоморфизму из группы G в абелеву группу A надо построить универсальный гомоморфизм $G^{\mathrm{ab}} \to A$, где G^{ab} какая-то абелева группа. При любом гомоморфизме из G в абелеву группу $[x, y]$ отображается в 1, следовательно, коммутант группы G лежит в ядре, и исходный гомоморфизм $G \to A$ пропускается через $G/[G, G]$. Заметим, что последняя группа абелева. Таким образом, $G^{\mathrm{ab}} = G/[G, G]$, а каноническая проекция $G \to G^{\mathrm{ab}}$ является единицей сопряжения.

(7) Второй пример – вложение **Grp** $\to$ **Mon**. Аналогично предыдущему примеру строим универсальную группу, в которую отображается моноид M. Можно сделать это совсем быстро, хотя при такой конструкции мы не много сможем сказать о реальном строении этой группы. Возьмем свободную группу F_M и профакторизуем по нормальному замыканию множества $\{xyz^{-1} \mid x, y, z \in M,\ xy = z\}$. Такая группа называется группой Гротендика моноида M и обозначается $K(M)$, хотя на самом деле Гротендик придумал эту конструкцию для абелевых групп и коммутативных моноидов.

(8) Правый сопряженный к функтору вложения. Рассмотрим то же вложение, что и в предыдущем примере. Коединица – это универсальное отображение группы в данный моноид. При гомоморфизме моноидов обратимые элементы переходят в обратимые. Поэтому любая группа перейдет в множество обратимых элементов моноида M. Таким образом, достаточно взять $\mathscr{G}(M) = M^*$.

(9) Пусть $\mathscr{C}$ – категория с конечными произведениями, а $\mathscr{F} : \mathscr{C} \times \mathscr{C} \to \mathscr{C}$ – функтор, отображающий (A, B) в $A \times B$ (универсальное свойство произведение индуцирует действие $\mathscr{F}$ на морфизмах). Левый сопряженный к нему будет диагональный функтор Δ, посылающий A в (A, A). Действительно, элемент $\mathrm{Mor}\big(\Delta(C), (A, B)\big)$ – это пара морфизмов $C \to A$ и $C \to B$, которая однозначно соответствует морфизму $C \to \mathscr{F}(A, B) = A \times B$ по определению произведения. Единица этого сопряжения – канонический морфизм $C \to C \times C$, соответствующий диаграмме $C \xleftarrow{\mathrm{id}} C \xrightarrow{\mathrm{id}} C$. Коединица – пара проекций $A \times B \to A$ и $A \times B \to B$.

(10) Левые сопряженные к функтору Mor. Пусть $\mathscr{C} = \mathbf{Set}$ или R-$\mathbf{Mod}$, а $A \in \mathscr{C}$. Рассмотрим функтор $\mathrm{Mor}(A, \_) : \mathscr{C} \to \mathscr{C}$ (множество гомоморфизмов R-модулей естественным образом превращается в R-модуль). Мы хотим найти функтор $\mathscr{F}_A$ такой, что

$$\mathrm{Mor}\big(X, \mathrm{Mor}(A, Y)\big) \cong \mathrm{Mor}(\mathscr{F}_A(X), Y).$$

Для множеств из экспоненциального закона мы знаем, что можно взять $\mathscr{F}_A(X) = X \times A$.

Для модулей $\mathrm{Mor}\big(X, \mathrm{Mor}(A, Y)\big)$ естественно изоморфно множеству билинейных отображений из $X \times A \to Y$. К счастью, определение тензорного произведения модулей говорит нам, что билинейные отображения из $X \times A$ можно заменить на линейные отображения из $X \otimes A$. Таким образом, в категории R-модулей

$$\mathrm{Mor}\big(X, \mathrm{Mor}(A, Y)\big) \cong \mathrm{Mor}(X \otimes A, Y),$$

т. е. левым сопряженным к функтору $\mathrm{Mor}(A, \_)$ является функтор $\_ \otimes A$.

Единицей этого сопряжения для множеств является функция $X \to \mathrm{Mor}(A, X \times A)$, которая элементу $x \in X$ сопоставляет функцию $a \mapsto (x, a)$. Коединица – это функция $\mathrm{Mor}(A, Y) \times A \to Y$, которая паре (f, a) сопоставляет $f(a)$.

(11) Пусть $\varphi : R \to A$ – гомоморфизм коммутативных колец. Он задает на A структуру R-алгебры, в частности, R-модуля. Пусть M – R-модуль. Тогда на $M \otimes_R A$ легко задать структуру A-модуля: $(x \otimes a) \cdot b := x \otimes (ab)$ $\forall a, b \in A$, $x \in M$, и далее доопределяем по линейности. Это отображение легко превращается в функтор $\varphi_\# : R$-$\mathbf{Mod} \to A$-$\mathbf{Mod}$, который называется *расширением скаляров*. Терминология пришла из теории векторных пространств, где R и A поля, следовательно, φ инъективно. В частности, для перехода к векторному пространству над алгебраическим замыканием $\overline{F}$ базового поля F не обязательно выбирать базис, а потом вкладывать пространство столбцов над F в пространство столбцов над $\overline{F}$. Вместо этого можно рассмотреть тензорное произведение пространства на $\overline{F}$.

Гомоморфизм φ задает также и функтор *сужения скаляров* $\varphi^\# : A$-$\mathbf{Mod} \to R$-$\mathbf{Mod}$. А именно, любой A-модуль M является R-модулем с умножением $x \cdot r := x\varphi(r)$, где $r \in R$ и $x \in M$. Нетрудно видеть, что $\varphi_\#$ является левым сопряженным к $\varphi^\#$. Единицей сопряжения является R-линейное отображение $M \to M \otimes_R A$, $x \mapsto x \otimes 1$. Действительно, для R-модуля M, A-модуля N и R-линейного отображения $\alpha : M \to N$ отображение $M \times A \to N$, $(x, a) \mapsto \alpha(x)a$ является R-билинейным. Поэтому оно единственным образом продолжается до R-линейного отображения $M \otimes A \to N$. Проверка того, что это отображение A-линейно и является единицей сопряжения, не представляет труда.

(12) Правого сопряженного к функтора вложения $\mathbf{Ab} \to \mathbf{Grp}$ не существует, потому что в произвольной группе невозможно универсальным образом выделить абелеву подгруппу. Строгое доказательство этого утверждения оставляется читателю в качестве упражнения.

6. Универсальные алгебраические конструкции

6.1. Универсальное свойство колец многочленов. Пусть R – коммутативное кольцо с 1. Для начала обсудим категорию R-алгебр. Напомним, что R-алгебра – это R-модуль с R-билинейным умножением. Рассмотрим категорию ассоциативных R-алгебр с 1, обозначаемую R-**Alg**. Морфизм в этой категории – это гомоморфизм R-алгебр, т. е. R-линейный гомоморфизм колец с 1. Так как 1 переходит в 1, то для любой R-алгебры A существует единственный гомоморфизм $R \to A$, $r \mapsto r \cdot 1_A$. Так что R является инициальным объектом категории R-**Alg**. В частности, инициальным объектом категории $\mathbb{Z}$-**Alg** = **Ring** является кольцо целых чисел. Заметим, что финальный объект в категории R-**Alg** – это нулевая алгебра.

Более того, так как $a(rb) = r \cdot 1_A \cdot (ab)$ для любых $r \in R$ и $a, b \in A$, то при $b = 1$ получаем $a(r \cdot 1_A) = (r \cdot 1_A)a$, т. е. образ R лежит в центре алгебры A.[4] Поэтому категория R-**Alg** изоморфна (даже не просто эквивалентна, а именно изоморфна) категории морфизмов $R \to A$ в категории **Ring**, таких, что образ R содержится в центре A.

Пусть теперь R-$\mathbf{Alg}_n$ обозначает категорию R-алгебр с n отмеченными точками. Точнее, объекты этой категории – это наборы $(A, a_1, \ldots, a_n)$, где A – R-алгебра, а $a_1, \ldots, a_n$ – коммутирующие между собой элементы алгебры A, а морфизмы $(A, a_1, \ldots, a_n) \to (B, b_1, \ldots, b_n)$ – это гомоморфизмы R-алгебр $\varphi : A \to B$ такие, что $\varphi(a_i) = b_i$ при всех i. Инициальным объектом в этой категории является алгебра многочленов $(R[t_1, \ldots, t_n], t_1, \ldots, t_n)$. Это и есть универсальное свойство кольца многочленов. Единственный морфизм $(R[t_1, \ldots, t_n], t_1, \ldots, t_n) \to (A, a_1, \ldots, a_n)$ – это гомоморфизм подстановки: $p \mapsto p(a_1, \ldots, a_n)$.

Другими словами, функтор **Set** $\to R$-**Alg**, посылающий множество X в кольцо многочленов $R[X]$, является левым сопряженным к забывающему.

6.2. Локализация. Пусть R – коммутативное кольцо с 1, а S – мультипликативное подмножество в R. Рассмотрим категорию $\mathscr{C}$ морфизмов $R \to A$ в категории **CRing** таких, что образы всех элементов из S обратимы в A. По определению инициальным объектом этой категории является гомоморфизм локализации $\lambda_S : R \to S^{-1}R$.

То, что S замкнуто относительно умножения и содержит 1, не играет никакой роли при таком определении. Локализацией в произвольном подмножестве $X \subseteq R$ будет локализация в мультипликативном подмножестве, порожденном X, т. е. в наименьшем мультипликативном подмоноиде R, содержащим X. Условие, что S мультипликативно, используется только для конструкции локализации.

Можно также отбросить и условие коммутативности. В этом случае локализация все еще будет существовать, но конструкция ее будет намного сложнее. Она является частным случаем конструкции локализации категории.

Легко видеть, что локализация является функтором. Более строго: пусть $\mathscr{A}$ – категория с объектами (R, S), где R – коммутативное кольцо с 1, а S – подмножество в R. Морфизм $(R, S) \to (R', S')$ в $\mathscr{A}$ – это гомоморфизм $R \to R'$, отображающий S в S'. Тогда сопоставление $(R, S) \mapsto S^{-1}R$ определяет функтор $\mathscr{A} \to$ **CRing**.

Обозначим через $\mathscr{B}$ – полную подкатегорию в $\mathscr{A}$, состоящую из тех пар, где $S \subseteq R^*$. Тогда отображение $(R, S) \mapsto (S^{-1}R, \lambda_S(S))$ определяет функтор из $\mathscr{A}$ в $\mathscr{B}$. Легко видеть, что этот функтор является левым сопряженным к функтору вложения $\mathscr{B} \to \mathscr{A}$.

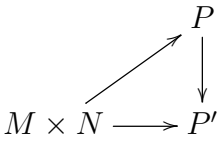
6.3. Локализация – эпиморфизм в категории CRing. Действительно, рассмотрим коммутативную диаграмму

$$R \xrightarrow{\lambda} S^{-1}R \underset{\psi}{\overset{\varphi}{\rightrightarrows}} X.$$

[4] В частности, алгебра кватернионов не является алгеброй над $\mathbb{C}$.

Так как при гомоморфизме колец с 1 обратимые элементы переходят в обратимые, то $\varphi\lambda = \psi\lambda$ – морфизм в категории $\mathscr{C}$. Так как λ является инициальным объектом в этой категории, то существует единственный морфизм $S^{-1}R \to X$, делающий нашу диаграмму коммутативной, т. е. $\varphi = \psi$. Нетрудно провести доказательство, пользуясь конструкцией локализации, а не ее универсальным свойством, но такое доказательство не будет работать для некоммутативных локализаций, а наше будет.

6.4. Тензорное произведение модулей. Пусть R – коммутативное кольцо с 1, а M и N – R-модули. Рассмотрим категорию $\mathscr{C}$, объектами которой являются билинейные отображения $M \times N \to P$, где P также является R-модулем, а морфизмами – R-линейные отображения $P \to P'$, для которых диаграмма



коммутативна (как диаграмма отображений множеств). Инициальный объект в такой категории называется тензорным произведением модулей M и N над R и обозначается через $M \otimes_R N$.

Формально, конструкция тензорного произведения чрезвычайно проста. Сначала мы забываем про билинейность отображений $M \times N \to P$ и строим универсальный объект без учета билинейности. Такую задачу мы только что решили: это свободный R-модуль с базисом $M \times N$. Для того, чтобы превратить каноническое отображение $M \times N \to \langle M \times N\rangle_R$ в билинейное, в свободном модуле должны быть выполнены условия

$$(m + m', n) = (m, n) + (m', n), \quad (m, n + n') = (m, n) + (m, n'), \quad r(m, n) = (rm, n) = (m, rn).$$

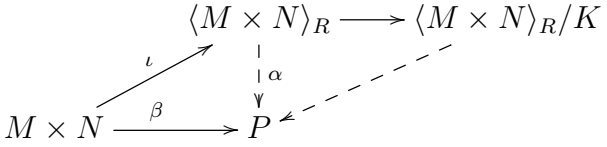
Универсальный модуль, в котором выполнены эти условия – фактормодуль

$$M \otimes_R N = \frac{\langle M \times N\rangle_R}{K},$$

где K – подмодуль в $\langle M \times N\rangle_R$, порожденный элементами $(m, n) + (m', n) - (m + m', n)$, $(m, n) + (m, n') - (m, n + n')$, $r(m, n) - (rm, n)$, и $r(m, n) - (m, rn)$ по всем $m, m' \in M$, $n, n' \in N$, и $r \in R$.

Предложение 6.1. *Модуль $M \otimes_R N$, определенный выше, действительно является тензорным произведением модулей M и N над R.*

Доказательство. Пусть $\beta : M \times N \to P$ – билинейное отображение. По универсальному свойству свободного модуля существует единственный гомоморфизм R-модулей $\alpha : \langle M \times N\rangle_R \to P$ такой, что $\beta = \alpha\iota$, где ι – каноническое вложение $M \times N$ в $\langle M \times N\rangle_R$. Билинейность β равносильна 4 равенствам, первое из которых $\beta(m + m', n) = \beta(m, n) + \beta(m', n)$. Из этих равенств следуют такие же равенства с заменой β на α, так как ι – вложение. Следовательно, K лежит в ядре α. По универсальному свойству факторгруппы (или фактормодуля) α единственным образом пропускается через каноническую проекцию $\langle M \times N\rangle_R \to \langle M \times N\rangle_R/K$, что доказывает существование. Единственность вытекает из единственности в определениях свободного модуля и фактормодуля.



□

Если кольцо R известно из контекста, то часто пишут $M \otimes N$ вместо $M \otimes_R N$. Канонический образ элемента $(m, n) \in M \times N$ в $M \otimes N$ обозначается $m \otimes n$ и называется разложимым тензором.

Если M и N свободные модули с базисами B и B' соответственно, то $M \otimes N$ – свободный модуль с базисом $\{b \otimes b' \mid b \in B,\ b' \in B'\}$. Антиинтуитивный пример: $\mathbb{Q} \otimes_{\mathbb{Z}} \mathbb{Z}/p\mathbb{Z} = 0$. Действительно,

$$a \otimes b = \frac{a}{p} \cdot p \otimes b = \frac{a}{p} \otimes pb = \frac{a}{p} \otimes 0 = 0,$$

а тензорное произведение порождено разложимыми тензорами.

Как и любой универсальный объект, тензорное произведение определяет функтор. В данном случае это функтор $\otimes : R\text{-}\mathbf{Mod} \times R\text{-}\mathbf{Mod} \to R\text{-}\mathbf{Mod}$, $(M, N) \mapsto M \otimes_R N$, который действует на морфизмах следующим образом. Пусть $\alpha : M \to M'$ и $\beta : N \to N'$ – R-линейные отображения. Они индуцируют билинейное отображение $M \times N \to M' \otimes_R N'$, $(m, n) \mapsto \alpha(m) \otimes \beta(n)$. Это билинейное отображение индуцирует линейное отображение $M \otimes_R N \to M' \otimes_R N'$, которое и называется $\alpha \otimes \beta$.

Более подробно тензорное произведение модулей будет изучаться в главе 10

6.5. Алгебраическое замыкание. Пусть F – поле, а $\mathscr{C}$ – категория гомоморфизмов $F \to K$, где K – алгебраически замкнутое поле. По аналогии с определением локализации хочется сказать, что алгебраическим замыканием поля F называется инициальный объект категории $\mathscr{C}$. Однако это не так. К сожалению, единственность не может быть выполнена, что показывает пример алгебраического замыкания поля $\mathbb{R}$: существует 2 автоморфизма поля $\mathbb{C}$, тождественных на $\mathbb{R}$, тождественный и сопряжение. Таким образом, алгебраическое замыкание F – это такое поле $\overline{F}$, вместе с вложением $\iota : F \to \overline{F}$, что любое вложение F в замкнутое поле пропукается через ι (мы говорим “вложение” вместо “гомоморфизм”, потому что любой гомоморфизм полей инъективен). То, что алгебраическое замыкание единственно с точностью до изоморфизма, уже не доказывается средствами теории категорий, хотя сам факт верен.

В такой ситуации иногда говорят “версальный объект” (универсальный без “уни”, т. е. без единственности). То есть можно сказать, что алгебраическое замыкание – это версальный объект категории $\mathscr{C}$.

6.6. Образ ретракции под действием функтора. Ретракцией называется морфизм, обратимый справа. Другими словами, если композиция морфизмов $X \to Y \to X$ равна тождественному, то правая стрелка называется ретракцией, а левая – сечением этой ретракции (терминология пришла из топологии). В параграфе 7 главы 8 мы доказали, что ретракция в категории групп – это проекция полупрямого произведения на один из сомножителей. Следовательно, в категории абелевых групп ретракция – это проекция на сомножитель прямого произведения. Пример ретракции в категории **CRing**: гомоморфизм ε подстановки значения из кольца многочленов над R в R.

Так как любой функтор сохраняет тождественные морфизмы, то он сохраняет и ретракции. Таким образом, если у вас есть произвольный функтор $\mathscr{F} : \mathbf{CRing} \to \mathbf{Ab}$, то $\mathscr{F}(R[t]) \cong \mathscr{F}(R) \times K$ для некоторой абелевой группы K. Действительно, применяя $\mathscr{F}$ к морфизмам $R \to R[t] \to R$, где правая стрелка – любой гомоморфизм подстановки, получаем гомоморфизмы $\mathscr{F}(R) \to \mathscr{F}(R[t]) \to \mathscr{F}(R)$ с тождественной композицией. А это означает, что $\mathscr{F}(R[t]) \cong \mathscr{F}(R) \oplus K$, где $K = \operatorname{Ker}\big(\mathscr{F}(\varepsilon)\big)$.

7. Пределы

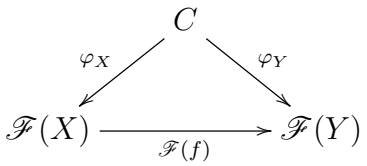
Обобщим понятие диаграммы в категории $\mathscr{C}$ на случай, когда индексирующим множеством является не граф, а категория.

Определение 7.1. *Диаграммой* в категории $\mathscr{C}$ называется ковариантный функтор $\mathscr{F} : \mathscr{J} \to \mathscr{C}$. Категория $\mathscr{J}$ называется *индексирующей*.

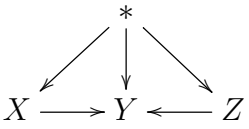
В частности, если $\mathscr{J}$ – дискретная категория (нет никаких морфизмов кроме кроме тождественных), то это означает, что мы индексируем класс $\operatorname{Obj}\mathscr{C}$ другим классом $\operatorname{Obj}\mathscr{J}$.

Определение 7.2. *Конусом* над диаграммой $\mathscr{F}\colon \mathscr{J} \to \mathscr{C}$ называется объект $C \in \operatorname{Obj}\mathscr{C}$ вместе с семейством морфизмов $\varphi_X : C \to \mathscr{F}(X)$ для всех $X \in \mathscr{J}$ таким, что для всякого морфизма $f \in \operatorname{Mor}_{\mathscr{J}}(X, Y)$ следующая диаграмма в $\mathscr{C}$ коммутативна:

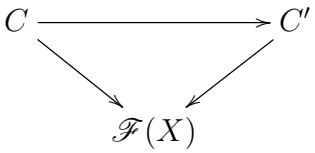
(34)



Конус часто обозначается $C \to \mathscr{F}$. Удобно представлять себе конус, как диаграмму, индексированную категорией $\mathscr{J} \cup \{*\}$. Последняя категория – это $\mathscr{J}$ с присоединенным инициальным объектом $\{*\}$ (при этом инициальный объект категории $\mathscr{J}$, если он существовал, перестает быть инициальным, так как из него нет морфизмов в $*$). Скажем, если $\mathscr{J} = (X \to Y \leftarrow Z)$, то $\mathscr{J} \cup \{*\}$ имеет вид



(так как $*$ – инициальный объект, то композиции стрелок по любому пути из $*$ в Y равны). Определим морфизм между двумя конусами $C \to \mathscr{F}$ и $C' \to \mathscr{F}$ над $\mathscr{F}$, как морфизм $C \to C'$, для которого коммутативен любой треугольник



Таким образом, получаем категорию конусов над $\mathscr{F}$.

Определение 7.3. *Пределом* диаграммы $\mathscr{F}\colon \mathscr{J} \to \mathscr{C}$ называется финальный объект в категории конусов. Он обозначается через $\varprojlim \mathscr{F}$.

Если индексирующая категория $\mathscr{J}$ конечная (счетная, малая),[5] то и предел называют *конечным* (*счетным*, *малым*).

Понятие конуса и категории конусов можно сформулировать немного другими словами. Определим категорию диаграмм $\operatorname{Funct}(\mathscr{J}, \mathscr{C})$, индексированных категорией $\mathscr{J}$. Объектами в ней являются диаграммы $\mathscr{J} \to \mathscr{C}$, а морфизмами – естественные преобразования диаграмм.[6] Заметим, что $\varprojlim$ является фунтором из категории $\operatorname{Funct}(\mathscr{J}, \mathscr{C})$ в категорию $\mathscr{C}$.

Для $C \in \mathscr{C}$ обозначим через $\Delta C : \mathscr{J} \to \mathscr{C}$ постоянный функтор, отображающий любой объект категории $\mathscr{J}$ в C, а любой морфизм – в тождественный Тогда конус – это морфизм $\Delta C \to \mathscr{F}$ в категории диаграмм (заметим, что естественность преобразования в точности означает коммутативность диаграммы (34)). Так как естественное преобразование $\Delta C \to \Delta C'$ однозначно определяется морфизмом $C \to C'$, то морфизм конусов на $\mathscr{F}$ – это морфизм в категории $\operatorname{Mor}_{\operatorname{Funct}(\mathscr{J},\mathscr{C})}$. Таким образом, категория конусов становится подкатегорией в $\operatorname{Mor}_{\operatorname{Funct}(\mathscr{J},\mathscr{C})}$. В этих терминах понятие предела можно сформулировать следующим образом.

Предложение 7.4. *Объект $L \in \mathscr{C}$ является пределом диаграммы $\mathscr{F} : \mathscr{J} \to \mathscr{C}$ тогда и только тогда, когда существует изоморфизм $\operatorname{Mor}_{\operatorname{Funct}(\mathscr{J},\mathscr{C})}(\Delta C, \mathscr{F}) \cong \operatorname{Mor}_{\mathscr{C}}(C, L)$ естественный по C.*

Другими словами, функтор $\varprojlim$ является правым сопряженным к функтору Δ.

[5] Учитывается мощность множества всех морфизмов в этой категории.

[6] Здесь тоже есть теоретико-множественная проблема, состоящая в том, что класс естественных преобразований между двумя фиксированными функторами не обязан быть множеством. Но мы не будем обращать на это внимание.

Доказательство. По определению предела естественное преобразование $\Delta \circ \varprojlim \to \mathrm{id}$ является коединицей сопряжения.

Обратно, если существует указанный естественный изоморфизм, то морфизм $\Delta C \to \mathscr{F}$, соответствующий id_C, удовлетворяет свойству предела, см. доказатество теоремы 5.2 с развернутыми стрелками. □

Примеры.

(1) Если $\mathscr{J}$ – категория без объектов, то для всякой категории $\mathscr{C}$ существует единственная *пустая диаграмма* $\mathscr{J} \to \mathscr{C}$. Конус над такой диаграммой – это просто объект $X \in \mathrm{Obj}(\mathscr{C})$. Предел пустой диаграммы – это финальный объект.

(2) Если $\mathscr{J}$ – дискретная категория (все стрелки – тождественные 1_X), то диаграмма $\mathscr{J} \to \mathscr{C}$ – это набор (не обязательно всех) объектов в категории $\mathscr{C}$, индексируемых $\mathscr{J}$. Предел такой диаграммы называется *произведением* этого набора объектов (ясно, что от индексации ничего не зависит). В частности, если $\mathscr{J}$ состоит из двух объектов, то получается определенное ранее произведение пары объектов из $\mathscr{C}$. Нетрудно доказать, что существует естественный изоморфизм

$$A_1 \times A_2 \times \cdots \times A_{n-1} \times A_n \cong \big(\ldots (A_1 \times A_2) \times \cdots \times A_{n-1}\big) \times A_n.$$

(3) Возьмем теперь $\mathscr{J} = (X \to Y \leftarrow Z)$. Тогда диаграмма – эта пара морфизмов $A \to B \leftarrow C$ в $\mathscr{C}$. Пределом этой диаграммы будет пулбэк $A \times_B C$.

(4) Если $\mathscr{J} = (\ X \rightrightarrows Y\ )$, то диаграмма – это пара морфизмов $\varphi, \psi \in \mathrm{Mor}_{\mathscr{C}}(A, B)$ а ее предел – эквалайзер этих морфизмов.

(5) Пусть $\mathscr{C}_{\mathbb{N}}$ – категория, связанная с упорядоченным множеством $\mathbb{N}$. Тогда диаграмма типа $\mathscr{C}_{\mathbb{N}}^{op}$ – это счетная последовательность морфизмов $\cdots \to A_n \to A_{n-1} \to \cdots \to A_1$. Предел такой диаграммы называется обратным пределом последовательности морфизмов и обозначается $\varprojlim\limits_{n\to\infty} A_n$. Более общий случай прямых и обратных пределов мы рассмотрим чуть ниже, а сейчас приведем 2 важных алгебраических конструкции, являющихся обратными пределами.

(6) Пусть R – коммутативное кольцо с 1. Рассмотрим диаграмму

$$\cdots \to R[t]/(t^n) \to R[t]/(t^{n-1}) \to \cdots \to R[t]/(t)$$

с очевидными морфизмами. Обратный предел этой диаграммы – это кольцо формальных степенных рядов $R[[t]]$. Если R – область целостности, то и $R[[t]]$ обладает этим свойством. Поле частных кольца $R[[t]]$ называют полем формальных степенных рядов и обозначают $R((t))$ (термин обычно испльзуется когда R – поле).

Заметим, что кольцо формальных степенных рядов над полем R является локальным кольцом с единственным максимальным идеалом, порожденным независимой переменной, любой ряд с ненулевым свободным членом обратим. Поэтому общий вид элемента поля формальных степенных рядов – формальный ряд Лорана $\sum\limits_{n=-k}^{\infty} r_n t^n$, где $k \in \mathbb{N}$.

(7) Пусть $p \in \mathbb{Z}$. Рассмотрим диаграмму

$$\cdots \to \mathbb{Z}/p^n\mathbb{Z} \to \mathbb{Z}/p^{n-1}\mathbb{Z} \to \cdots \to \mathbb{Z}/p\mathbb{Z}.$$

Обратный предел этой диаграммы называется кольцом целых p-адических чисел и обозначается $\mathbb{Z}_p$ (именно поэтому в научных статьях не используется обозначение $\mathbb{Z}_p$ для кольца вычетов по модулю p; мы тоже начиная с этого места будем писать только $\mathbb{Z}/p\mathbb{Z}$ или $\mathbb{Z}/p$ для кольца вычетов). С помощью китайской теоремы об остатках легко доказать, что для взаимно простых $a, b \in \mathbb{Z}$ имеем $\mathbb{Z}_{ab} \cong \mathbb{Z}_a \times \mathbb{Z}_b$. С другой стороны, если выкинуть часть диаграммы, оставив бесконечно много морфизмов, то предел не измениться (это несложное упражнение), так что $\mathbb{Z}_{p^k} = \mathbb{Z}_p$. Поэтому обычно считают, что p – простое число.

Целое p-адическое число – это последовательность чисел a_n, $0 \leqslant a_n \leqslant p^n - 1$, $n \in \mathbb{N}$ такая, что $a_{n+1} \equiv a_n \mod p^n$. Сложение определено следующим образом: $(a_n) + (b_n) = (a_n + b_n \mod p^n)$, аналогично определено умножение. Из этого описания ясно, что при простом p кольцо $\mathbb{Z}_p$ является областью целостности, а его поле частных обозначется $\mathbb{Q}_p$ и называется полем p-адических чисел.

(8) Последние 2 примера можно обобщить, взяв произвольный идеал I коммутативного кольца R и рассмотрев обратный предел диаграммы

$$\cdots \to R/I^n \to R/I^{n-1} \to \cdots \to R/I$$

(k-й степенью идеала называется наименьший идеал, содержащий всевозможные произведения k элементов идеала). Это кольцо называется пополнением кольца R в I-адической топологии и обозначается $\hat{R}_I$. Чаще $I = sR$ – главный идеал, тогда говорят про s-адическую топологию.

Если теперь развернуть стрелки, то получатся *ко*конусы и *ко*пределы.

Предел или копредел произвольного функтора – вещь достаточно экзотическая. Чаще всего рассматриваются пределы диаграмм, индексированных направленными множествами.

Определение 7.5. *Направленным множеством* называется непустое частично упорядоченное множество $(J, \leqslant)$, в котором для любых $i, j \in J$ найдется $k \in J$, такой что $i, j \leqslant k$.

Если в качестве индексирующей категории $\mathscr{J}$ выступает направленное множество $(J, \leqslant)$, то копредел диаграммы $F\colon \mathscr{J} \to \mathscr{C}$ называется *прямым или инъективным пределом* и обозначается $\varinjlim$. Конструкцию прямого предела можно пересказать чуть подробнее. Пусть имеется направленное множество $(J, \leqslant)$ и задано семейство $\{X_i\}_{i \in J}$ объектов категории $\mathscr{C}$, так что выполняются следующие свойства:

(1) Для всех $i \leqslant j$ задан морфизм $X_i \xrightarrow{f_i^j} X_j$.
(2) Для всех $i \leqslant j \leqslant k$ выполнено $f_j^k f_i^j = f_i^k$ и $f_i^i = 1_{X_i}$.

Тогда говорят, что морфизмы f_i^j образуют *направленнное семейство*.

Если взять копредел по этому набору морфизмов, то получится объект

$$\varinjlim_{i \in J} X_i,$$

который называется *прямым пределом* соответствующего семейства объектов и морфизмов.

Простой пример: пусть X_i – множества, частично упорядоченные по включению, и пусть они образуют прямое семейство. Тогда $\varinjlim X_i = \bigcup_i X_i$. Прямые пределы обычно существуют в конкретных категориях и, если морфизмы инъективны, то равны объединению.

Вот еще один два полезных примера прямых пределов в категории коммутативных колец с 1.

(1) Для кольца R рассмотрим диаграмму всех его конечнопорожденных $\mathbb{Z}$-подалгебр, отображения – включение меньших подалгебр в большие. Это направленное семейство морфизмов, потому что наименьшая подалгебра, содержащая две конечнопорожденные подалгебры, конечно порождена. Так как любой элемент содержится в какой-нибудь конечнопорожденной подалгебре, прямой предел этого набора морфизмов равен R. Это позволяет сводить некоторые вопросы о произвольных кольцах к конечнопорожденным, которые по теореме Гильберта о базисе являются нетеровыми.
(2) Пусть S – мультипликативное подмножество кольца R. Главная локализация R_s имеет канонический гомоморфизм в R_{st}, где $s, t \in S$. Множество главных локализаций в элементах из S с каноническими гомоморфизмами является направленным множеством. Действительно, для $s, t \in S$ имеем гомоморфизмы $R_s \to R_{st}$ и $R_t \to R_{st}$. Прямой предел этого направленного множества – локализация $S^{-1}R$. Это наблюдение позволяет сводить некоторые вопросы о произвольных локализциях к главным.

Двойственно определяется обратный (проективный) предел (ко)направленного семейства морфизмов. Два примера проективных пределов мы уже рассмотрели выше.

Определение 7.6. Категория называется полной (кополной), если в ней существуют все малые пределы (соотв. копределы).

Без доказательства сформулируем теорему о существовании конечных (ко)пределов.

Теорема 7.7. *В категории существуют все конечные (ко)пределы, если в ней существуют (ко)произведения любых двух объектов, (ко)эквалайзеры и финальный (инициальный) объект. Равносильно: в ней существует финальный (инициальный) объект и пулбэки (пушауты).*

Заметим, что даже в категориях, в которых конечные произведения совпадают с конечными копроизведениями, бесконечные произведения и копроизведения обычно отличаются. Например, в категории R-модулей $\prod_{i=1}^{\infty} M_i$ – это множество всех последовательностей $(a_i \in M_i \mid i \in \mathbb{N})$, а $\bigoplus_{i=1}^{\infty} M_i := \coprod_{i=1}^{\infty} M_i$ – множество *финитных* последовательностей. В соответствии с этим и в других конкретных категориях декартовым (или прямым) произведением любого количества объектов называется множество всех последовательностей, а прямой суммой – множество финитных последовательностей. Заметим, что в категории колец с 1 бесконечных прямых сумм не существует, потому что в множестве финитных последовательностей нет 1 (последовательности, состоящей из 1).

Одним из важных свойств функтора является сохранение (ко)пределов. Оказывается, что для сопряженного функтора одно из этих свойств выполнено автоматически.

Теорема 7.8. *Левый сопряженный функтор сохраняет копределы, а правый – пределы.*

Для доказательства этого факта нам понадобится следующая лемма.

Лемма 7.9. *Пусть $\mathscr{F} : \mathscr{B} \to \mathscr{C}$ – функтор, а $\mathscr{G}$ – левый сопряженный к нему. Для диаграмм $H : \mathscr{J} \to \mathscr{C}$ и $\mathscr{K} : \mathscr{J} \to \mathscr{B}$ имеет место биекция $\mathrm{Mor}(\mathscr{G} \circ \mathscr{H}, \mathscr{K}) \cong \mathrm{Mor}(\mathscr{H}, \mathscr{F} \circ \mathscr{K})$ естественная по $\mathscr{H}$ и $\mathscr{K}$ (здесь Mor обозначает множество естественных преобразований одного функтора в другой).*

Доказательство. Пусть $\eta : \mathscr{G} \circ \mathscr{H} \to \mathscr{K}$ – естественное преобразование. Для каждого $X \in \mathscr{J}$ обозначим через θ_X образ морфизма $\eta_X \in \mathrm{Mor}_{\mathscr{B}}\Big(\mathscr{G}\big(\mathscr{H}(X)\big), \mathscr{K}(X)\Big)$ в $\mathrm{Mor}_{\mathscr{C}}\Big(\mathscr{H}(X), \mathscr{F}\big(\mathscr{K}(X)\big)\Big)$. Для морфизма $\varphi \in \mathrm{Mor}_{\mathscr{J}}(X, Y)$ имеем диаграмму

$$\begin{array}{ccccc} \mathscr{H}(X) & \xrightarrow{f_{\mathscr{H}(X)}} & \mathscr{F}\mathscr{G}\mathscr{H}(X) & \xrightarrow{\mathscr{F}(\eta_X)} & \mathscr{F}\mathscr{K}(X) \\ \downarrow & & \downarrow & & \downarrow \\ \mathscr{H}(Y) & \xrightarrow{f_{\mathscr{H}(Y)}} & \mathscr{F}\mathscr{G}\mathscr{H}(Y) & \xrightarrow{\mathscr{F}(\eta_Y)} & \mathscr{F}\mathscr{K}(Y) \end{array}$$

на которой вертикальные стрелки индуцированы морфизмом φ, а f – единица сопряжения. Из доказательства теоремы 5.2 видно, что композиции горизонтальных стрелок равны θ_X и θ_Y соответственно. Так как f и η являются естественными пребразованиями, то левый и правый квадраты на диаграмме коммутативны. Следовательно, коммутативна и вся диаграмма, откуда θ также естественно.

Мы получили отображение $\mathrm{Mor}(\mathscr{G} \circ \mathscr{H}, \mathscr{K}) \to \mathrm{Mor}(\mathscr{H}, \mathscr{F} \circ \mathscr{K})$. Двойственным образом построим отображение в обратную сторону. Проверка того, что эти отображения естественны по $\mathscr{H}$ и $\mathscr{K}$ и взаимно обратны является рутиной. □

Доказательство теоремы. Пусть $\mathscr{F}$, $\mathscr{G}$ и $\mathscr{H}$ – как в лемме. Для $C \in \mathscr{C}$ имеем $\mathscr{G} \circ \Delta C = \Delta\mathscr{G}(C)$. Тогда

$$\mathrm{Mor}(\Delta C, \mathscr{F} \circ \mathscr{H}) \cong \mathrm{Mor}(\Delta\mathscr{G}(C), \mathscr{H}) \cong \mathrm{Mor}_{\mathscr{B}}(\mathscr{G}(C), \varprojlim \mathscr{H}) \cong \mathrm{Mor}_{\mathscr{C}}(C, \mathscr{F}(\varprojlim \mathscr{H})),$$

откуда $\mathscr{F}(\varprojlim \mathscr{H}) = \varprojlim \mathscr{F} \circ \mathscr{H}$. □

8. Представимые функторы и лемма Йонеды

Определение 8.1. Функтор $\mathscr{C} \to \mathbf{Set}$ называется представимым, если он естественно изоморфен функтору $\mathrm{Mor}(A, \_)$ для некоторого $A \in \mathscr{C}$.

В случае, когда $\mathscr{C}$ – категория коммутативных R-алгебр, представимый функтор называется *аффинной схемой* над R. Если $A = R[t_1, \dots, t_n]$, то по универсальному свойству кольца многочленов гомоморфизмы $A \to B$ находятся в биективном соответствии с наборами из n элементов алгебры B. Таким образом, функтор $B \mapsto B^n \cong \mathrm{Mor}(A, B)$ является аффинной схемой над R. Эта схема называется аффинным пространством и обозначается через $\mathbb{A}^n_R$.

Если A – конечнопорожденная алгебра над R, то она изоморфна алгебре $R[t_1, \dots, t_n]/I$ для некоторого идеала I кольца многочленов $R[t_1, \dots, t_n]$. Если R нетерово, то по теореме Гильберта о базисе I конечно порожден, скажем, многочленами $f_1, \dots, f_m$. Тогда множество $\mathrm{Mor}(A, B)$ состоит из точек аффинного пространства $b = (b_1, \dots, b_n) \in B^n$, для которых $f_1(b) = \dots = f_m(b) = 0$. Действительно, каждому гомоморфизму $A \to B$ соответствует композиция $R[t_1, \dots, t_n] \to A \to B$ и так определенное отображение $\mathrm{Mor}(A, B) \to \mathrm{Mor}(R[t_1, \dots, t_n], B) \cong B^n$ инъективно. По универсальному свойству факторкольца гомоморфизм $\varepsilon_b : R[t_1, \dots, t_n] \to B$ пропускается через A тогда и только тогда, когда $\varepsilon_b(I) = 0 \iff \varepsilon_b(f_i) = f_i(b) = 0 \forall i$.

Такие схемы (если A – факторкольцо кольца многочленов от конечного числа переменных по конечнопорожденному идеалу) называются схемами конечного типа над R. Множество общих корней конечного набора многочленов называется алгебраическим множеством в аффинном пространстве. Таким образом, каждому алгебраическому множеству соответствует аффинная схема конечного типа, это соответствие сюръективно, но разным множествам могут соответствовать изоморфные схемы.

Общим элементом представимого функтора $\mathscr{G} \cong \mathrm{Mor}(A, \_)$ называется элемент $g \in \mathscr{G}(A)$, соответствующий тождественному морфизму id_A. Он обладает замечательным свойством:

$$\forall B\, \forall x \in \mathscr{G}(B)\, \exists!\, \varphi_x \in \mathrm{Mor}(A, B) :\ \mathscr{G}(\varphi_x)(g) = x.$$

Например, аффинной схемой над $\mathbb{Z}$ является полная линейная группа $\mathrm{GL}_n(B) \cong \mathrm{Mor}(A, B)$, где $A = \mathbb{Z}[t, g_{ij} \mid 1 \leqslant i, j \leqslant n]/(t \det g = 1)$ (главная локализация кольца многочленов $\mathbb{Z}[g_{ij}]$ в элементе $\det g$), а g матрица с элементами g_{ij}. Матрица $b \in \mathrm{GL}_n(B)$ отождествляется с гомоморфизмом φ_b, который отображает g_{ij} в b_{ij}, а t в $(\det b)^{-1}$. Ясно, что общим элементом является матрица $g \in \mathrm{GL}_n(A)$.

Теорема 8.2 (лемма Йонеды). *Пусть $\mathscr{C}$ – малая категория, а $\mathscr{F} : \mathscr{C} \to \mathbf{Set}$ – функтор. Тогда существует биекция*

$$\mathrm{Nat}\big(\mathrm{Mor}(A, \_), \mathscr{F}\big) \to \mathscr{F}(A), \qquad \varphi \mapsto \varphi_A(\mathrm{id}_A)$$

где Nat *обозначает множество естественных преобразований. Более того, эта биекция естественна по A и F.*

Доказательство. Зададим отображение из $\mathscr{F}(A)$ в $\mathrm{Nat}\big(\mathrm{Mor}(A, \_), \mathscr{F}\big)$ по следующему правилу: элементу $a \in \mathscr{F}(A)$ сопоставим естественное преобразование $\eta : \mathrm{Mor}(A, \_) \to \mathscr{F}$ по формуле $\eta^{(a)}_B(f) = \mathscr{F}(f)(a)$. Так как любой функтор сохраняет тождественные морфизмы, то $\eta^{(a)}_A(\mathrm{id}_A) = a$. Обратно, пусть $\eta_A(\mathrm{id}_A) = a$, а $f \in \mathrm{Mor}(A, B)$. Так как η – естественное преобразование, то диаграмма

$$\begin{array}{ccc} \mathrm{Mor}(A, A) & \xrightarrow{\alpha \mapsto f\alpha} & \mathrm{Mor}(A, B) \\ {\scriptstyle \eta_A}\big\downarrow & & \big\downarrow{\scriptstyle \eta_B} \\ \mathscr{F}(A) & \xrightarrow{\mathscr{F}(f)} & \mathscr{F}(B) \end{array}$$

коммутативна. Применив к id_A верхний и правый морфизм, получим $\eta_B(f)$, а нижний и левый – $\mathscr{F}(f)(a)$. Таким образом, $\eta = \eta^{(a)}$. Следовательно, заданные нами отображения действительно взаимно обратны.

Пусть $\varphi \in \mathrm{Mor}(A,B)$, а $\theta : \mathscr{F} \to \mathscr{G}$ – естественное преобразование. Для доказательства естественности биекции из условия теоремы необходимо проверить коммутативность диаграммы

$$\begin{array}{ccccc}
\mathrm{Nat}\big(\mathrm{Mor}(A,_),\mathscr{F}\big) & \xrightarrow{\eta\mapsto\theta\eta} & \mathrm{Nat}\big(\mathrm{Mor}(A,_),\mathscr{G}\big) & \longrightarrow & \mathrm{Nat}\big(\mathrm{Mor}(B,_),\mathscr{G}\big) \\
\downarrow{\scriptstyle \eta\mapsto\eta_A(\mathrm{id}_A)} & & \downarrow{\scriptstyle \eta\mapsto\eta_A(\mathrm{id}_A)} & & \downarrow{\scriptstyle \eta\mapsto\eta_B(\mathrm{id}_B)} \\
\mathscr{F}(A) & \xrightarrow{\theta_A} & \mathscr{G}(A) & \xrightarrow{\mathscr{G}(\varphi)} & \mathscr{G}(B)
\end{array}$$

Любое естественное преобразование из левого верхнего угла диаграммы равно $\eta^{(a)}$ для некоторого $a \in \mathscr{F}(A)$. Отображая $\eta^{(a)}$ вниз, а затем налево, получим $\theta_A(a)$. Образ $\eta^{(a)}$ под действием верхней левой стрелки равен $\theta\eta^{(a)}$. Отображая это естественное преобразование вниз получаем $\theta_A\eta^{(a)}_A(\mathrm{id}_A) = \theta_A(a)$. Таким образом, левый квадрат коммутативен.

Для доказательства коммутативности правого квадрата надо выяснить, как устроена верхняя правая стрелка. Морфизм φ индуцирует естественное преобразование $\lambda^{(\varphi)} : \mathrm{Mor}(B,\_) \to \mathrm{Mor}(A,\_)$, заданное формулой $\lambda^{(\varphi)}_C(\alpha) = \alpha\varphi$, где $\alpha \in \mathrm{Mor}(B,C)$. Следовательно, верхняя правая стрелка посылает естественное преобразование η в $\eta\lambda^{(\varphi)}$. Пусть $c \in \mathscr{G}(A)$ и $\eta^{(c)} \in \mathrm{Nat}\big(\mathrm{Mor}(A,\_),\mathscr{G}\big)$. Отображая $\eta^{(c)}$ направо и вниз получаем $\eta^{(c)}_B\lambda^{(\varphi)}_B(\mathrm{id}_B) = \eta^{(c)}_B(\varphi)$. Посылая то же естественное преобразование сначала вниз, а потом направо, получим $\mathscr{G}(\varphi)(c)$. Теперь правый квадрат коммутативен по определению $\eta^{(c)}_B$ из первого предложения доказательства. □

Следствие 8.3. *Если A, B – объекты малой категории $\mathscr{C}$, то*

$$\mathrm{Nat}\big(\mathrm{Mor}(A,\_),\mathrm{Mor}(B,\_)\big) \cong \mathrm{Mor}(B,A).$$

*Следовательно, категория представимых функторов $\mathscr{C} \to$ **Set** (морфизмы – естественные преобразования) антиэквивалентна категории $\mathscr{C}$. Например, категория аффинных схем над кольцом R антиэквивалентна категории R-**Alg**.*

9. Абелевы категории

Определение 9.1. Категория называется *предаддитивной*, если на каждом множестве $\mathrm{Mor}(A,B)$ задана структура аддитивной абелевой группы, и это сложение дистрибутивно относительно композиции морфизмов.

Функтор из одной предаддитивной категории в другую называется аддитивным, если он переводит сумму морфизмов в сумму.

Лемма 9.2. *В предаддитивной категории $\mathscr{C}$ следующие условия на объект C эквивалентны.*

(1) *C – инициальный объект.*
(2) *C – финальный объект.*
(3) $\mathrm{Mor}(C,C) = 0$.
(4) $\mathrm{id}_C = 0$.

Определение 9.3. Диаграмма

$$A \underset{p_1}{\overset{i_1}{\leftrightarrows}} C \underset{p_2}{\overset{i_2}{\leftrightarrows}} B$$

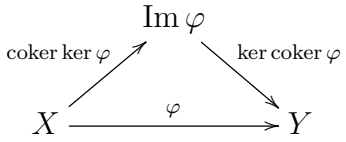
называется бипроизведением, если $p_1i_1 = \mathrm{id}$, $p_2i_2 = \mathrm{id}$ и $i_1p_1 + i_2p_2 = \mathrm{id}$.

Предаддитивная категория называется *аддитивной*, если в ней существуют все конечные произведения, в том числе произведение пустого множества объектов – нулевой объект.

Аддитивная категория называется *абелевой*, если в ней существуют все ядра и коядра, а все мономорфизмы и эпиморфизмы (ко)нормальны.

Теорема 9.4 (теорема Митчела о вложении). *Для любой малой абелевой категории существует полный, строгий и точный функтор из нее в категорию R-модулей для некоторого кольца R.*

В абелевой категории определяются понятия комплексов и точных последовательностей и может быть развита теория (ко)гомологий. Из определения следует, что любой морфизм в абелевой категории раскладывается в композицию эпиморфизма и мономорфизма. Объект, стоящий в середине, называется образом морфизма. Он равен (канонически изоморфен) ядру коядра и коядру ядра.



Комплекс – это последовательность морфизмов

$$\dots \to A_{k-1} \xrightarrow{\delta_{k-1}} A_k \xrightarrow{\delta_k} A_{k+1} \to \dots$$

с нулевой композицией. Легко доказать, что композиция $\operatorname{Im}\delta_{k-1} \to A_k \to A_{k+1}$ нулевая, следовательно существует единственный морфизм $\operatorname{Im}\delta_{k-1} \to \operatorname{Ker}\delta_k$. Коядро этого морфизма (фактор ядра следующего по образу предыдущего) называется гомологиями комплекса в k-м члене. Комплекс с нулевыми гомологиями называется точной последовательностью. Аддитивный функтор называется точным, если он сохраняет точные последовательности. Точная последовательность

$$0 \longrightarrow A \xrightarrow{\alpha} B \xrightarrow{\beta} C \longrightarrow 0$$

называется короткой точной последовательностью. В категории модулей это означает, что $C \cong B/A$ (формально, фактормодуль не по самому A, а по его образу в B).

Глава 10

Полилинейная алгебра

1. Простейшие свойства тензорного произведение

В этом параграфе мы работаем в категории R-**Mod**. Множество морфизмов из M в N обычно обозначают $\mathrm{Hom}_R(M,N)$ вместо категорного $\mathrm{Mor}_{R\text{-}\mathbf{Mod}}(M,N)$. В разделе 6.4 предыдущей главы мы определили тензорное произведение модулей. Сейчас докажем несколько простейших свойств.

Предложение 1.1. *Существуют следующие естественные изоморфизмы:*

(1) $X \otimes Y \cong Y \otimes X$;
(2) $(X \otimes Y) \otimes Z \cong X \otimes (Y \otimes Z)$;
(3) $(X \oplus Y) \otimes Z \cong (X \otimes Z) \oplus (Y \otimes Z)$;
(4) $R \otimes_R X \cong X$.
(5) $R^n \otimes_R X \cong X^n$.
(6) $R^m \otimes R^n \cong R^{mn}$

Доказательство. Все эти изоморфизмы легко строятся и проверяются непосредственно. Укажем для каждого пункта образы разложимых тензоров, оставив необходимые проверки читателю (везде в этом доказательстве $x \in X$, $y \in Y$, $z \in Z$, а $r \in R$).

(1) $x \otimes y \mapsto y \otimes x$;
(2) $(x \otimes y) \otimes z \mapsto x \otimes (y \otimes z)$;
(3) $(x, y) \otimes z \mapsto (x \otimes z, y \otimes z)$;
(4) $r \otimes x = 1 \otimes rx \mapsto rx$;
(5) следует из 3 и 4, учитывая, что $X^n = \underbrace{X \oplus \cdots \oplus X}_{n \text{ раз}}$;
(6) является частным случаем предыдущего пункта.

□

Заметим, что первые 4 свойства говорят о том, что класс R-модулей является (большим) коммутативным полукольцом с 1 относительно операций прямой суммы и тензорного произведения. Коммутативное полукольцо с 1 – это множество с двумя операциями, относительно каждой из которых оно является коммутативным моноидом, и которые удовлетворяют дистрибутивности. Слово "большим" относится к тому, что вместо множеств рассматриваются классы. Тогда последние 2 свойства являются чисто формальными свойствами, выполненными в любом полукольце.

Из последнего пункта следует, что тензорное произведение свободных модулей свободно. Нетрудно видеть, что базисом будет набор из тензорных произведений базисных векторов.

Из конструкции тензорного произведения легко видеть, что функтор $M \otimes \_$ сохраняет эпиморфизмы. С другой стороны, он не обязан сохранять мономорфизмы даже для $\mathbb{Z}$-модулей, т. е. абелевых групп: $\mathbb{Z} \hookrightarrow \mathbb{Q}$, но $\mathbb{Z}/p\mathbb{Z} \cong \mathbb{Z} \otimes \mathbb{Z}/p\mathbb{Z} \to \mathbb{Q} \otimes \mathbb{Z}/pZ = 0$ совсем не инъективно. Если для модуля M функтор $M \otimes \_$ сохраняет мономорфизмы (и, следовательно, является точным), то модуль M называется *плоским*. Из предыдущего предложения легко следует, что свободные модули являются плоскими.

Упражнение 1.2. Пусть L, L' – линейные операторы на конечномерных векторных пространствах V и V' соответственно. По функториальности тензорного произведения они

индуцируют оператор $L \otimes L'$ на пространстве $V \otimes V'$. Найдите след и определитель оператора $L \otimes L'$ (выразите их через следы и определители операторов L и L' и размерности пространств).

Следующие естественные отображения, связывающие между собой функторы Hom_R и тензорного произведения, менее очевидны, и не все из них являются изоморфизмами в общем случае. Однако для классического случая конечномерных векторных пространств они – изоморфизмы, что будет содержательно использоваться в дальнейшем.

Лемма 1.3. *В категории модулей над (коммутативным) кольцом R с 1 существуют следующие естественные преобразования функторов.*

(1) $\mathrm{Hom}(R, X) \cong X$.
(2) $\mathrm{Hom}\big(X, \mathrm{Hom}(Y, Z)\big) \cong \mathrm{Bil}\big(X \times Y, Z\big) \cong \mathrm{Hom}\big(X \otimes Y, Z\big)$.
(3) $M \to M^{**}$.
(4) $\mathrm{Hom}(X, Y \otimes Z) \to \mathrm{Hom}(X \otimes Y^*, Z)$.

Доказательство. Первая эквивалентность очевидна: $f \mapsto f(1)$, обратно $x \mapsto f(r) = rx$. Следующие две были построены в примере 10 параграфа 5 и примере 6 параграфа 3 предыдущей главы.

Пусть $f : X \to Y \otimes Z$ – линейное отображение, $x \in X$, $\varphi \in Y^*$, а $g \in \mathrm{Hom}(X \otimes Y^*, Z)$ – образ f, который мы хотим построить. Положим

$$g(x \otimes \varphi) = \sum_{i=1}^{k} \varphi(y_i) z_i, \text{ где } f(x) = \sum_{i=1}^{k} y_i \otimes z_i,$$

и продоложим по линейности.[1] Так как разложение $f(x)$ в сумму разложимых тензоров неоднозначно, надо еще проверить, что $g(x \otimes \varphi)$ не зависит от выбора этого разложения. Так как тензорное произведение – фактормодуль свободного по подмодулю, порожденному 4 типами элементов, то достаточно проверить, что g не изменится при добавлении элемента одного из этих четырех типов. Все эти проверки очевидны. Например, если к $f(x)$ добавится $(a + b) \otimes z - a \otimes z - b \otimes z$, то к $g(x \otimes \varphi)$ добавится $\varphi(a + b)z - \varphi(a)z - \varphi(b)z = 0$.

Естественность построенного отображения по каждому из аргументов X, Y и Z проверяется непосредственно (надо только не забыть, что функтор Hom контравариантен по первому аргументу). □

Теорема 1.4. *Все естественные преобразования из леммы 1.3 являются естественными изоморфизмами на категории конечномерных векторных пространств.*

Доказательство. Учитывая, что $\dim(U \otimes V) = (\dim U) \cdot (\dim V) = \dim \mathrm{Hom}(U, V)$, легко заметить, что размерности левой и правой части каждого естественного преобразования совпадают. Таким образом достаточно доказать, что указанные отображения инъективны. Инъективность отображения 3 очевидна, что уже отмечалось в предыдущей главе.

Осталось доказать инъективность отображения 4. Пусть $(y_1, \ldots, y_n)$ – базис пространства Y. Запишем $f(x) = \sum_{i=1}^{n} y_i \otimes z_i$ для некоторых $z_i \in Z$. Так как $f \neq 0$, то найдется $x \in X$ и индекс m такие, что $z_m \neq 0$. Зададим $\varphi \in Y^*$ действием на базисных элементах: $\varphi(y_i) = \delta_{im}$. Тогда

$$g(x \otimes \varphi) = \sum_{i=1}^{n} \varphi(y_i) z_i = z_m \neq 0.$$

Таким образом, ненулевой гомоморфизм f не может отобразиться в 0, т. е. ядро нашего естественного отображения нулевое. □

[1] На самом деле, когда говорят, "зададим отображение g на разложимых тензорах, а потом продолжим по линейности", имеют ввиду следующую формальную процедуру. Зададим отображение $G : X \times Y^* \to Z$ формулой $G(x, \varphi) = \sum_{i=1}^{k} \varphi(y_i) z_i$, проверим, что оно билинейно, и определим g как единственное отображение $X \otimes Y^* \to Z$, соответствующее G по универсальному свойству тензорного произведения.

Следующие естественные биекции являются ключевыми для понимания того, что два разных подхода к понятию тензора эквивалентны.

Следствие 1.5. *Пусть $U_1, \dots, U_m, V_1, \dots, V_n$ – конечномерные векторные пространства над полем F. Множество полилинейных отображений*

$$U_1 \times \dots \times U_m \to V_1 \otimes \dots \otimes V_n$$

естественно изоморфно множеству полилинейных отображений

$$U_1 \times \dots \times U_m \times V_1^* \times \dots \times V_n^* \to F,$$

а также пространству

$$U_1^* \otimes \dots \otimes U_m^* \otimes V_1 \otimes \dots \otimes V_n.$$

Следующая группа естественных преобразований менее важна, чем предыдущая, поэтому их доказательство оставляется читателю в качестве упражнения.

Упражнение 1.6. Постройте следующие естественные отображения и докажите, что в категории конечномерных векторных пространств они являются изоморфизмами.

(1) $\mathrm{Hom}(X, Y) \to \mathrm{Hom}(Y^*, X^*)$.
(2) $\mathrm{Hom}(X, Y) \to (Y^* \otimes X)^*$.
(3) $Y \otimes X^* \to \mathrm{Hom}(X, Y)$.
(4) $X^* \otimes Y^* \to (X \otimes Y)^*$.

2. Тензоры

Пусть V – конечномерное векторное пространство над F.

Определение 2.1. Тензором называется полилинейное отображение

$$T : \underbrace{V \times \dots \times V}_{p \text{ раз}} \times \underbrace{V^* \times \dots \times V^*}_{q \text{ раз}} \to F.$$

Другими словами, тензор – это линейное отображение

$$T : \underbrace{V \otimes \dots \otimes V}_{p \text{ раз}} \otimes \underbrace{V^* \otimes \dots \otimes V^*}_{q \text{ раз}} \to F.$$

Такой тензор называется p раз ковариантный и q раз контравариантный или, короче, тензором типа (p, q).

Билинейная форма – тензор типа $(2, 0)$, линейный оператор отождествляется с тензором типа $(1, 1)$, так как $\mathrm{Hom}(V, V)$ естественно изоморфно $\mathrm{Hom}(V \otimes V^*, F)$. Билинейная бинарная операция $V \times V \to V$ задается тензором типа $(2, 1)$.

Тензорное произведение тензора T типа (p, q) и тензора T' типа (p', q') – это тензорное произведение отображений $T \otimes T'$, определенное в примере 6.4 предыдущей главы. $T \otimes T'$ является тензором типа $(p + p', q + q')$.

Ясно, что полилинейное отображение однозначно определяется своими значениями на наборах базисных элементов. Эти числа будут называться координаты тензора. Но прежде чем их определить, нам надо выбрать базис в двойственном пространстве.

Определение 2.2. Пусть X – векторное пространство с базисом $(e_1, \dots, e_n)$. Зададим функционалы $e^k \in X^*$ равенствами $e^k(e_i) = \delta_{ki}$ для всех $k = 1, \dots, n$. Тогда $e^* = \{e^1, \dots, e^n\}$ называется двойственным (по отношению к e) базисом пространства X^*.

Лемма 2.3. *Для $x \in X$ число $e^k(x)$ – это k-я координата вектора x в базисе e.*
Множество e^* действительно является базисом пространства X^*.

Доказательство. Первое утверждение сразу следует из линейности e_k^* и определения его значений на базисных векторах. Для любого функционала $f \in X^*$ имеем

$$f\big(x\big) = f\big(\sum_{i=1}^{n} e_i e^i(x)\big) = \sum_{i=1}^{n} f(e_i)e^i(x),$$

т. е. $f = \sum_{i=1}^{n} f(e_i)e^i$, откуда e^* – система образующих. Пусть $\sum_{i=1}^{n} \alpha_i e^i = 0$. Подставляя в это равенство базисный элемент e_k имеем $\alpha_k = 0$, что доказывает линейную независимость. □

Замечание 2.4. Если e бесконечный базис, то набор $\{e^i\}$ не является системой образуюших. Для конкретного $x \in X$ вынесенная формула все еще верна, и в ней конечное число слагаемых, потому что только для конечного числа индексов $e^i(x)$ отлично от нуля. Но убрать x из этой формулы уже не получится, потому что все коэффициенты $f(e_i)$ могут быть не равны 0.

В дальнейшем мы будем использовать обозначения для координат векторов и ковекторов, принятые в полилинейной алгебре (напомним, что ковекторы – это элементы V^*, в анализе они называются линейными функционалами, еще их можно называть линейными формами). Координаты вектора x в базисе e обозначаются через $x^1, \dots, x^n$, а координаты ковектора f в базисе e^* – через $f_1, \dots, f_n$. Элементы набора векторов, также как и элементы базиса V, нумеруются нижними индексами, а ковекторов – верхними. Все эти соглашения нужны для того, чтобы суммирование всегда происходило по тем индексам, которые встречаются и сверху и снизу. Элементы матриц в этой системе обозначений надо было бы писать в виде a^i_j, где верхний индекс – номер строки, но мы не будем пользоваться этим обозначением.

Определение 2.5. Координатами тензора в базисе $e = (e_1, \dots, e_n)$ пространства V называется $p+q$-мерный массив, состоящий из элементов поля

$$T^{j_1\dots j_q}_{i_1\dots i_p} = T(e)^{j_1\dots j_q}_{i_1\dots i_p} = T(e_{i_1}, \dots, e_{i_p}, e^{j_1}, \dots, e^{j_q}), \quad i_1, \dots, i_p, j_1, \dots, j_q \in \{1, \dots, n\}$$

Ясно, что координаты тензора полностью определяют этот тензор:

$$T(v_1, \dots, v_p, f^1, \dots, f^q) = \sum T^{j_1\dots j_q}_{i_1\dots i_p} v_1^{i_1} \dots v_p^{i_p} f^1_{j_1} \dots f^q_{j_q},$$

где сумма берется по всем $i_1, \dots, i_p, j_1, \dots, j_q$, независимо друг от друга пробегающих множество $\{1, \dots, n\}$.

Для того, чтобы увидеть, как координаты тензора меняются при замене базиса, посмотрим, как связаны между собой матрицы перехода $C_{e\to g}$ и $C_{e^*\to g^*}$. Для этого определим "умножение" $V^* \times V \to F$, по формуле $f \cdot v = f(v)$. Ясно, что это умножение билинейно, поэтому можно пользоваться формализмом, определенном в параграфе 2 главы 2. В этих обозначениях определение двойственного базиса можно записать в виде $(e^*)^\intercal e = E$.

Лемма 2.6. $C_{e^*\to g^*} = (C^{-1}_{e\to g})^\intercal$.

Доказательство. По определению матрицы перехода $g = eC_{e\to g}$ и $g^* = e^* C_{e^*\to g^*}$. По определению двойственного базиса

$$E = (g^*)^\intercal g = C^\intercal_{e^*\to g^*}(e^*)^\intercal e C_{e\to g} = C^\intercal_{e^*\to g^*} C_{e\to g},$$

откуда получается требуемое равенство. □

Теорема 2.7. *Пусть e и g – базисы пространства V, а T – тензор на V типа (p, q).*

$$T(g)^{m_1\dots m_q}_{k_1\dots k_p} = \sum (c')^{m_1\dots m_q}_{j_1\dots j_q} T(e)^{j_1\dots j_q}_{i_1\dots i_p} c^{i_1\dots i_p}_{k_1\dots k_p}, \text{ где}$$

$$c^{i_1\dots i_p}_{k_1\dots k_p} = \prod_{l=1}^{p} (C_{e\to g})_{i_l k_l}, \qquad (c')^{m_1\dots m_q}_{j_1\dots j_q} = \prod_{r=1}^{q} (C_{g\to e})_{m_r j_r}.$$

Доказательство.

$$T(g_{k_1},\dots,g_{k_p},g^{m_1},\dots,g^{m_q}) = \\ T\left(\sum_{i_1=1}^{n} e_{i_1}(C_{e\to g})_{i_1k_1},\dots,\sum_{i_p=1}^{n} e_{i_p}(C_{e\to g})_{i_pk_p},\sum_{j_1=1}^{n} e^{j_1}(C_{g\to e})_{m_1j_1},\dots,\sum_{j_p=1}^{n} e^{j_p}(C_{g\to e})_{m_pj_p}\right).$$

Пользуясь полилинейностью T получаем требуемое равенство. □

3. Тензорная алгебра модуля

Пусть R – коммутативное кольцо с 1, а $\mathscr{F} : R\text{-}\mathbf{Alg} \to R\text{-}\mathbf{Mod}$ – забывающий функтор. Сейчас мы построим функтор T, сопряженный к $\mathscr{F}$. Для R-модуля M алгебра $T(M)$ называется тензорной алгеброй модуля M.

Напомним, что $\bigoplus_{k=0}^{\infty} M_k$ – это прямая сумма или копроизведение R-модулей M_k, т. е. множество всех финитных последовательностей $(m_1, m_2, \dots)$, где $m_k \in M_k$.

Обозначим через $M^{\otimes k}$ тензорное произведение k экземпляров модуля M, при этом положим $M^{\otimes 0} = R$, так как именно R является "нейтральным элементом" по отношению к тензорному произведению. Тогда формула $T(M) = \bigoplus_{k=0}^{\infty} M^{\otimes k}$ задает структуру R-модуля на $T(M)$. Умножение достаточно задать на разложимых тензорах после чего распространить это определение по линейности. Итак

$$(x_1 \otimes \dots \otimes x_k) \cdot (y_1 \otimes \dots \otimes y_n) = x_1 \otimes \dots \otimes x_k \otimes y_1 \otimes \dots \otimes y_n$$

(таким образом, произведение любого элемента из $M^{\otimes k}$ на элемент из $M^{\otimes n}$ лежит в $M^{\otimes(k+n)}$). Действие функтора T на морфизмах очевидна.

Теорема 3.1. *Функтор T, построенный выше, является сопряженным к забывающему функтору $\mathscr{F} : R\text{-}\mathbf{Alg} \to R\text{-}\mathbf{Mod}$. Вложение $f : M = M^{\otimes 1} \hookrightarrow T(M)$ является единицей сопряжения.*

Доказательство. Достаточно доказать, что f – единица сопряжения, т. е. для любого R-линейного отображения $g : M \to A$ модуля M в алгебру A существует единственный гомоморфизм R-алгебр $g' : T(M) \to A$ такой, что $g = g' \circ f$. Ясно, что отображение

$$G_k : M^k \to A, \qquad (m_1,\dots,m_k) \mapsto g(m_1)\cdot\ldots\cdot g(m_k)$$

является полилинейным. По универсальному свойству тензорного произведения оно пропускается через $M^{\otimes k}$ единственным образом. Далее, по универсальному свойству копроизведения модулей все отображения G_k единственным образом пропускаются через $T(M) = \bigoplus_k M^{\otimes k}$. Легко проверить, что полученное отображение $T(M) \to A$ сохраняет операцию умножения, т. е. является гомоморфизмом алгебр. □

Для свободного модуля с базисом X тензорная алгебра – это свободная алгебра на множестве X. Действительно, забывающий функтор $R\text{-}\mathbf{Alg} \to \mathfrak{Set}$ является композицией забывающих функторов $R\text{-}\mathbf{Alg} \to R\text{-}\mathbf{Mod} \to \mathfrak{Set}$. Следовательно, и сопряженный к нему является композицией сопряженных, т. е. $R\langle X\rangle = T(\langle X\rangle_R)$.

4. Градуированные алгебры

Прежде чем рассматривать факторалгебры тензорной алгебры, предназначенные для изучению полилинейных [анти]симметричных отображений, мы докажем пару несложных свойств градуированных алгебр. Пусть R – коммутативное кольцо с 1, N – коммутативный моноид в аддитивной записи, а A – не обязательно коммутативная R-алгебра с 1.

Определение 4.1. Предположим, что $A = \bigoplus\limits_{n\in N} A_n$, где A_n являются R-модулями, а умножение в A устроено так, что $A_kA_n \subseteq A_{k+n}$. Тогда A называется N-градуированной алгеброй.

Элемент любого модуля A_n называется однородным элементом. По определению любой $a \in A$ представляется единственным образом в виде суммы однородных элементов $a = \sum\limits_{n\in N} a_n$, где $a_n \in A_n$ и почти все a_n равны нулю. В этом случае ненулевые a_n называются однородными компонентами элемента a.

Простейшим примером $\mathbb{N}_0$-градуированной алгебры является кольца многочленов от некоторого числа переменных. В этом случае A_n – это множество однородных многочленов (полной) степени n. По определению тензорной алгебры модуля она тоже является $\mathbb{N}_0$-градуированной. Кольцо многочленов от m переменных можно рассматривать и как $\mathbb{N}_0^m$-градуированной алгебры. В этом случае $A_{(n_1,\dots,n_m)}$ является свободным R-модулем ранга 1, порожденным одночленом $t_1^{n_1}\dots t_m^{n_m}$.

Гомоморфизмом N-градуированных алгебр $A \to A'$ называется гомоморфизм алгебр, сохраняющий градуировку, т. е. отображающий A_n в A'_n для любого $n \in N$. В этом параграфе мы хотим выяснить, в каком случае факторизация является гомоморфизмов градуированных алгебр.

Лемма 4.2. *Пусть A – N-градуированная алгебра, а I – идеал в A. Следующие условия эквивалентны.*

(1) *I порожден однородными элементами.*
(2) *Однородные компоненты любого элемента из I лежат в I.*
(3) $I = \bigoplus\limits_{n\in N} A_n \cap I$.
(4) *Канонический гомоморфизм $\rho : A \to A/I$ является гомоморфизмом N-градуированных алгебр.*

Доказательство. (1) $\Longrightarrow$ (3). Любой элемент из I имеет вид $a = \sum\limits_{n\in N} b_nx_nc_n$, где $x_n \in I \cap A_n$ – однородные элементы из системы образующих идеала I, а $b_n, c_n \in A$. Записав $b_n = \sum\limits_{k\in N} b_{nk}$ и $c_n = \sum\limits_{k\in N} c_{nk}$, где $b_{nk}, c_{nk} \in A_k$, получим

$$a = \sum_{n,k,j\in N} b_{nk}x_nc_{nj} = \sum_{h\in N}\Big(\sum_{n+k+j=h} b_{nk}x_nc_{nj}\Big)$$

(во всех суммах почти все слагаемые равны нулю). Так как $b_{nk}x_nc_{nj} \in A_{n+k+j} \cap I$, то h-ая внутренняя скобка лежит в $A_h \cap I$.

(3) $\Longrightarrow$ (4). Зададим гомоморфизм модулей $A \to \bigoplus\limits_{n\in N} \frac{A_n}{A_n\cap I}$ по правилу $\sum\limits_{n\in N} a_n \mapsto (a_n + A_n \cap I)_{n\in N}$. Ясно, что это отображение сюръективно. Ядро же его равно $\sum\limits_{n\in N} A_n \cap I = I$. По теореме о гомоморфизме получаем изоморфизм модулей. Условие $(A_n \cap I)\cdot(A_k \cap I) \subseteq A_{k+n} \cap I$ очевидно. Таким образом, $A/I \cong \bigoplus\limits_{n\in N} \frac{A_n}{A_n\cap I}$ является N-градуированной алгеброй, а гомоморфизм $A \to A/I$ – гомоморфизмом N-градуированных алгебр.

(4) $\Longrightarrow$ (2). Пусть $a = \sum\limits_{n\in N} a_n \in I$, где $a_n \in A_n$. Тогда $0 = \rho(a) = \sum\limits_{n\in N} \rho(a_n)$. Так как ρ – гомоморфизм градуированных алгебр, то $\rho(a_n) \in (A/I)_n$, а так как сумма однородных компонент прямая, то $\rho(a_n) = 0$ для любого $n \in N$. А это и значит, что все однородные компоненты a_n элемента a лежат в I.

(2) $\Longrightarrow$ (1). Берем произвольную систему образующих идеала I и заменяем каждую образующую на набор ее однородных компонент. Ясно, что получившийся набор однородных

элементов порождает не меньше, чем I. Но по (2) все однородные компоненты лежат в I, поэтому они порождают в точности I. □

Идеал, удовлетворяющий условиям предыдущей леммы называется однородным идеалом.

Пусть $\varphi : N \to N'$ – гомоморфизм моноидов, а A – N-градуированная алгебра. Тогда можно рассматривать и N'-градуировку на A по формуле: $A_k = \bigoplus\limits_{n\in\varphi^{-1}(k)} A_n$. Рассмотрение N'-градуировки полезно, если хочется сделать какой-то идеал однородным, а относительно N-градуировки он таковым не является. Например, пусть $p \in F[t_1, \dots, t_m]$ – однородный многочлен степени n. Тогда главный идеал, порожденный $p-1$ не является $\mathbb{N}_0$-градуированным, но является $\mathbb{Z}/n\mathbb{Z}$-градуированным. Следовательно, алгебра $F[t_1, \dots, t_m]/(p-1)$ является $\mathbb{Z}/n\mathbb{Z}$-градуированной – лучше, чем ничего.

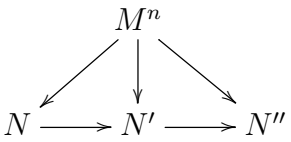
Пусть $\varepsilon : N \to \mathbb{Z}/2\mathbb{Z}$ – фиксированный гомоморфизм моноидов. N-градуированная алгебра A называется антикоммутативной (относительно ε), если $xy = (-1)^{\varepsilon(n)\varepsilon(k)}yx$ для любых $x \in A_n$ и $y \in A_k$. Важнейшим примером антикоммутативной $\mathbb{N}_0$-градуированной алгебры является алгебра Грассмана, изучение которой начинается в следующем параграфе. В англоязычной литературе кроме термина anticommutative используются также skew-commutative, supercommutative, graded-commutative, а иногда даже просто commutative, будьте осторожны имея дело с коммутативностью градуированных алгебр!

5. Симметрическая и внешняя алгебры

Напомним, что полилинейное отображение называется симметричным, если его значение не меняется при перестановке аргументов, и антисимметричным, если его значение равно нулю, как только какие-либо 2 аргумента равны (если $1/2 \in R$, то это равносильно тому, что оно меняет знак при транспозиции аргументов). Сначала мы построим универсальное n-линейное [анти]симметричное отображение из фиксированного модуля $M^n = M \times \dots \times M$.

Пусть M – R-модуль, а $n \in \mathbb{N}$. Рассмотрим категории $\mathscr{S}^{\pm}_{M,n}$:

- $\operatorname{Obj}\mathscr{S}^{\pm} =$
  $\{(N, f) \mid N \in R\text{-}\mathbf{Mod}$, f – [анти]симметричное полилинейное отображение $M^n \to N\}$;
  (симметричное для категории $\mathscr{S}^{+}_{M,n}$ и антисимметричное – для $\mathscr{S}^{-}_{M,n}$).
- $\operatorname{Mor}_{\mathscr{S}^{\pm}_{M,n}}\big((N, f), (N', f')\big) = \{g \in \operatorname{Mor}_{R\text{-}\mathbf{Mod}}(N, N') \mid g \circ f = f'\}$;
- композиция – это композиция отображений:



Инициальный объект в категории $\mathscr{S}^{\pm}_{M,n}$ называется n-ой симметрической (соотв. внешней) степенью модуля M и обозначаются через $S^n(M)$ и $\Lambda^n(M)$ соответственно.

Теорема 5.1. $S^n(M) = M^{\otimes n}/\langle x \otimes a \otimes b \otimes y - x \otimes b \otimes a \otimes y\rangle$;
$\Lambda^n(M) = M^{\otimes n}/\langle x \otimes a \otimes a \otimes y\rangle$;
где $a \in M$, а x, y – разложимые тензоры.

Доказательство. Любое полилинейное отображение из $M^n \to N$ единственным образом пропускается через $M^{\otimes n}$. Так как отображение [анти]симметрично, то указанные в формулировке элементы лежат в его ядре. По универсальному свойству фактормодуля наше отображение единственным образом пропускается через фактормодуль $S^n(M)$ или $\Lambda^n(M)$.

Так как $(a+b)\otimes(a+b) - a\otimes a - b\otimes b = a\otimes b + b\otimes a$, то подмодуль в $M^{\otimes n}$, порожденный элементами $\dots \otimes a \otimes a \otimes \dots$ содержит все элементы вида $\dots \otimes a \otimes b \otimes \dots + \dots \otimes b \otimes a \otimes \dots$, что позволяет в элементах $\Lambda^n(M)$ переставлять сомножители со сменой знака. Следовательно, любой элемент вида $\dots \otimes a \otimes \dots \otimes a \otimes \dots$ также равен нулю в $\Lambda^n(M)$. Таким образом, канонические отображения $M^n \to S^n(M)$ и $M^n \to \Lambda^n(M)$ являются [анти]симметричными. □

Обозначим канонический образ элемента $m_1 \otimes \cdots \otimes m_n$ в $S^n(M)$ через $m_1 \dots m_n$, а в $\Lambda^n(M)$ – через $m_1 \wedge \cdots \wedge m_n$ (символ "$\wedge$" в этом контексте называется символом внешнего произведения).

Рассмотрим теперь градуированные алгебры

$$S(M) := \bigoplus_{n=0}^{\infty} S^n(M),$$
$$\Lambda(M) := \bigoplus_{n=0}^{\infty} \Lambda^n(M),$$

умножение в которых индуцировано умножением в тензорной алгебре модуля M. Они называются *симметрической* и *внешней* алгеброй модуля M соответственно. Внешняя алгеброй называется также *алгеброй Грассмана*. По определению канонические отображения $T(M) \to S(M)$ и $T(M) \to \Lambda(M)$ будут гомоморфизмами градуированных алгебр.

Предложение 5.2. *Ядро отображения $T(M) \to S(M)$ порождено множеством $\{a \otimes b - b \otimes a \mid a, b \in M\}$.*

Ядро отображения $T(M) \to \Lambda(M)$ порождено множеством $\{a \otimes a \mid a \in M\}$.

Доказательство. Оба идеала порождены однородными элементами. По лемме 4.2 Однородные компоненты факторалгебры являются факторами однородных компонент $T(M)$ по своему пересечению с идеалом. Легко видеть, что эти пересечения совпадают с подмодулями из определения $S^n(M)$ и $\Lambda^n(M)$. □

Следствие 5.3. $S(M) \cong T(M)/\langle a \otimes b - b \otimes a \mid a, b \in M\rangle$;
$\Lambda(M) \cong T(M)/\langle a \otimes a \mid a \in M\rangle$.

Заметим, что конструкции симметрической и внешней алгебр, очевидно, функториальна, т. е. гомоморфизм модулей $\varphi : M \to M'$ индуцирует гомоморфизмы симметрических и внешних алгебр $S(\varphi) : S(M) \to S(M')$ и $\Lambda(\varphi) : \Lambda(M) \to \Lambda(M')$. При этом следующие диаграммы коммутативны.

$$\begin{array}{ccc} M & \longrightarrow & S(M) \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle S(\varphi)} \\ M' & \longrightarrow & S(M') \end{array} \qquad \begin{array}{ccc} M & \longrightarrow & \Lambda(M) \\ {\scriptstyle\varphi}\downarrow & & \downarrow{\scriptstyle \Lambda(\varphi)} \\ M' & \longrightarrow & \Lambda(M') \end{array}$$

Другими словами, вложения $M \to S(M)$ и $M \to \Lambda(M)$ являются естественными преобразованиями тождественного функтора в композицию S (соответственно Λ) с забывающим из алгебр в модули.

Теорема 5.4. *Построенные выше функторы S и Λ является левыми сопряженными к забывающим функторам из категории коммутативных R-алгебр (соответственно градуированных антикоммутативных R-алгебр) в категорию R-модулей.*

Если M – свободный модуль с базисом X, то $S(M) \cong R[X]$ – кольцо многочленов от X.

Доказательство. Указанные выше естественные преобразования являются единицами сопряжений. Доказательство этого факта абсолютно аналогично доказательству универсальных свойств симметрической и внешней степеней из теоремы 5.1.

Забывающий функтор из R-**Alg** в **Set** является композицией забывающих функторов из категории коммутативных R-алгебр с 1 R-**Calg** в R-**Mod** и из R-**Mod** в **Set**. Легко видеть, что композиция левых сопряженных функторов является левым сопряженным функтором к композиции. Свободный модуль – левый сопряженный к забывающему из R-**Mod** в **Set**, а S – левый сопряженный к забывающему из R-**Calg** в R-**Mod**. Их композиция будет левым сопряженным к забывающему из R-**Calg** в **Set**, т. е. кольцом многочленов, что следует из универсального свойства кольца многочленов. □

В заключении этого параграфа найдем базис модуля $\Lambda^n(R^k)$. Для доказательства следующего утверждения удобно будет ввести обозначения:

- $[n] = \{1, \dots, n\}$;
- для набора индексов $I = \{i_1, \dots, i_k\} \subseteq [n]$, $i_1 < \dots < i_k$, и кортежа элементов $(x_1, \dots, x_n)$ из M положим $x_I = x_{i_1} \wedge \dots \wedge x_{i_k}$.

Предложение 5.5. *Если $M \cong R^n$ – конечнопорожденный свободный модуль над R с базисом $e = (e_1, \dots, e_n)$, то базисом $S^k(M)$ является набор $\{e_{i_1} \cdots e_{i_k} \mid 1 \leqslant i_1 \leqslant \dots \leqslant i_k \leqslant n\}$, а базисом $\Lambda^k(M)$ – набор $\{e_{i_1} \wedge \dots \wedge e_{i_k} \mid 1 \leqslant i_1 < \dots < i_k \leqslant n\}$. В частности, ранг модуля $\Lambda^k(M)$ равен C_n^k.*

Доказательство. Ясно, что указанные наборы порождают модули $S^k(M)$ и $\Lambda^k(M)$. Утверждение про симметрическую степень следует из определения многочлена от нескольких переменных (линейная комбинация различных одночленов равна 0 $\iff$ все коэффициенты равны 0). Пусть

$$\sum_{I \subseteq [n],\, |I| = k} \alpha_I e_I = 0$$

Для каждого k-элементного подмножества $J \subseteq [n]$ домножим это равенство на $e_{[n] \smallsetminus J}$ в алгебре $\Lambda(M)$. Ясно, что в сумме останется только одно слагаемое $\pm \alpha_J e_{[n]}$. Для того чтобы доказать, что $e_{[n]} \neq 0$ в $\Lambda^n(M)$ зададим антисимметричное полилинейное отображение $\varphi : M^n \to R$ формулой $\varphi(m_1, \dots, m_n) = \det\big((m_1)_e, \dots, (m_n)_e\big)$. По определению $\Lambda^k(M)$ оно единственным образом пропусается через линейное отображение $\varphi' : \Lambda^n(M) \to R$. При этом $\varphi'(e_{[n]}) = \varphi(e_1, \dots, e_n) = \det E = 1$. Следовательно, $e_{[n]} \neq 0$, откуда $\alpha_J = 0$. Таким образом, все коэффициенты вынесенной линейной комбинации равны 0, а это и означает, что наш набор элементов линейно независим. □

Замечание 5.6. Альтернативное (даже более простое) доказательство последнего предложения: строим свободный модуль с базисом из предложения, определяем очевидное (анти)симметричное отображение из M в $\Lambda^k(M)$ (соотв. $S^k(M)$) и проверяем универсальное свойство.

6. Теорема Бине–Коши

В этом параграфе V – конечномерное векторное пространство над F, а все вычисления происходят в алгебре Грассмана $\Lambda(V)$.

Для k-элементного подмножества $I = \{i_1, \dots, i_k\} \subseteq [n]$ и матрицы $C \in \mathrm{M}_{n,k}(F)$ обозначим через C^I подматрицу матрицы C, состоящую из строк с номерами $i_1, \dots, i_k$. Аналогично, для $B \in \mathrm{M}_{k,n}(F)$ через B_I обозначается матрица, составленная из столбцов с номерами $i_1, \dots, i_k$.

Лемма 6.1. *Пусть $u = (u_1, \dots, u_n)$ и $v = (v_1, \dots, v_k)$ два набора элементов пространства V, а $C \in \mathrm{M}_{n,k}(F)$. Если $v = uC$, то*

$$v_{[k]} = \sum_{I \subseteq [n],\ |I| = k} u_I \cdot \det C^I,$$

В частности, если $k = n$, то $v_{[n]} = u_{[n]} \cdot \det C$.

Доказательство. Докажем сначала частный случай $k = n$ при условии, что u – базис V. Тогда любой элемент пространства $\Lambda^n(V)$ однозначно записывается в виде $u_{[n]}\alpha$ для некоторого $\alpha \in F$. Обозначим $v_{[n]} = u_{[n]} \cdot \varphi(C)$. Легко видеть, что $\varphi : \mathrm{M}_n(F) \to F$ – антисимметричная полилинейная форма столбцов матрицы C, а $\varphi(E) = 1$. Поэтому $\varphi(C) = \det C$.

Пусть теперь $k \leqslant n$, а u по-прежнему является базисом. По предложению 5.5 множество $\{u_I \mid I \subseteq [n],\ |I| = k\}$ является базисом пространства $\Lambda^k(V)$. Поэтому

$$v_{[k]} = \sum_{I \subseteq [n],\ |I| = k} u_I \cdot \alpha_I,$$

для некоторых $\alpha_I \in F$. Зафиксируем k-элементное подмножество $J \subseteq [n]$. Пусть V_J факторпространство пространства V по подпространству, порожденному u_i по всем $i \notin J$. Обозначим через $\bar{x}$ канонический образ элемента $x \in V$ в V_J. Аналогичное обозначение будем использовать для набора элементов пространства V. Тогда

$$\bar{v} = (\bar{u}_{j_1}, \ldots, \bar{u}_{j_k})C^J,$$

где $J = \{j_1, \ldots, j_k\}$ и $j_1 < \cdots < j_k$. По первой части доказательства $\bar{v}_{[k]} = \bar{u}_J \cdot \det C^J$. С другой стороны ясно, что $\bar{u}_I = 0$ для любого $I \neq J$. Поэтому $\bar{v}_{[k]} = \bar{u}_J \cdot \alpha_J$. Так как $(\bar{u}_{j_1}, \ldots, \bar{u}_{j_k})$ является базисом пространства V_J, то $\bar{u}_J \neq 0$, следовательно $\alpha_I = \det C^I$.

Рассмотрим теперь общий случай. Пусть $(e_1, \ldots, e_n)$ – стандартный базис пространства F^n. По доказанному в предыдущем параграфе

$$c_{*1} \wedge \cdots \wedge c_{*k} = \sum_{I \subseteq [n],\ |I|=k} e_I \cdot \det C^I,$$

Рассмотрим гомоморфизм $\varphi : F^n \to V$, заданный равенством $\varphi(a) = ua$. Тогда $\varphi(e_i) = u_i$ и $\varphi(c_{*i}) = v_i$. Так как Λ является функтором, то φ индуцирует гомоморфизм $\Lambda(\varphi) : \Lambda(F^n) \to \Lambda(V)$. Применяя $\Lambda(\varphi)$ к последней вынесенной формуле, получаем результат. □

Теорема 6.2 (Бине–Коши). *Пусть $A = BC \in \mathrm{M}_k(F)$, где $B \in \mathrm{M}_{k,n}(F)$, а $C \in \mathrm{M}_{n,k}(F)$. Тогда*

$$\det A = \sum_{I \subseteq [n],\ |I|=k} \det B_I \det C^I.$$

Доказательство. Пусть $u_i = b_{*i}$, $v_i = a_{*i}$, а e – стандартный базис F^k. Тогда $u = (u_1, \ldots, u_n) = eB$, а $v = (v_1, \ldots, v_k) = eA = uC$. По лемме 6.1 имеем

$$e_{[k]} \det A = v_{[k]} = \sum_{I \subseteq [n],\ |I|=k} u_I \det C^I = e_{[k]} \sum_{I \subseteq [n],\ |I|=k} \det B_I \det C^I.$$

□

7. Разложение определителя по группе столбцов

Пусть $I = \{i_1, \ldots, i_k\} \subseteq [n]$, $[n] \smallsetminus I = \{j_1, \ldots, j_{n-k}\}$, $i_1 < \cdots < i_k$ и $j_1 < \cdots < j_{n-k}$. Для доказательства следующего утверждения нам необходимо вычислить четность перестановки

$$\sigma_I = \begin{pmatrix} 1 & \ldots & k & k+1 & \ldots & n \\ i_1 & \ldots & i_k & j_1 & \ldots & j_{n-k} \end{pmatrix}$$

Лемма 7.1. $\varepsilon(\sigma_I) = \left(\sum_{i \in I} i + \frac{k(k+1)}{2}\right) \mod 2$.

Доказательство. Посчитаем количество инверсий в перестановке σ_I, которое, очевидно, равно количеству пар (i_l, j_h) таких, что $i_l > j_h$. При фиксированном l имеем $I \cap [i_l] = \{i_1, \ldots, i_l\}$. Поэтому $|[i_l] \smallsetminus I| = i_l - l$, а это и есть то, что надо посчитать. Суммируя $i_l - l$ по $l = 1, \ldots, k$, получаем результат. □

Обозначим через ε_I четность перестановки σ_I.

Лемма 7.2. *Пусть V – векторное пространство с базисом $e = (e_1, \ldots, e_n)$. Пространство $\Lambda^{n-k}(V)$ отождествляется с пространством $\Lambda^k(V)^*$ посредством отображения $\varphi : \Lambda^{n-k}(V) \to \Lambda^k(V)^*$, заданного формулой $e_{[n]}\varphi(x)(y) = x \wedge y$, где $x \in \Lambda^{n-k}(V)$, а $y \in \Lambda^k(V)$ ($x \wedge y$ всегда равно произведению некоторой константы на $e_{[n]}$; эта константа и есть то число, в которое $\varphi(x)$ отображает y).*

Положим $e^I = e_{[n]\smallsetminus I} \cdot (-1)^{\varepsilon_{[n]\smallsetminus I}}$. Тогда базисы $\{e_I \mid |I| = k\}$ и $\{e^I \mid |I| = k\}$ являются двойственными, другими словами, $e^I \wedge e_I = e_{[n]}$ и $e^I \wedge e_J = 0$ для любого $J \neq I$, $|J| = k$.

Теорема 7.3. *Пусть $A = (B\,C)$ – матрица, разбитая на блоки $B \in \mathrm{M}_{n,k}(F)$ и $C \in \mathrm{M}_{n,n-k}(F)$. Тогда*

$$\det A = \sum_{I\subseteq[n],\ |I|=k} (-1)^{\varepsilon_I} \det B^I \cdot \det C^{[n]\smallsetminus I}.$$

Доказательство. Пусть $V = F^n$ – пространство со стандартным базисом e. По лемме 6.1 во внешней алгебре этого пространства имеет место равенство

$$e_{[n]} \cdot \det A = a_{*1} \wedge \cdots \wedge a_{*n} = (b_{*1} \wedge \cdots \wedge b_{*k}) \wedge (c_{*1} \wedge \cdots \wedge c_{*\,n-k}).$$

По лемме 6.1 сомножители правой части равны

$$\sum_{I\subseteq[n],\ |I|=k} e_I \cdot \det B^I \text{ и } \sum_{J\subseteq[n],\ |J|=n-k} e_J \cdot \det C^J, \text{ соответственно.}$$

По лемме 7.2 во внешнем произведении этих сумм остаются только слагаемые с $J = [n] \smallsetminus I$. Таким образом

$$e_{[n]} \cdot \det A = \sum_{I\subseteq[n],\ |I|=k} e_I \cdot \det B^I \wedge e_{[n]\smallsetminus I} \cdot \det C^{[n]\smallsetminus I} = \sum_{I\subseteq[n],\ |I|=k} e_{[n]} (-1)^{\varepsilon_I} \det B^I \det C^{[n]\smallsetminus I}.$$

□

8. Миноры обратной матрицы

В следующей теореме мы выразим миноры матрицы A^{-1} через миноры матрицы A. Для этого вначале создадим матрицу, состоящую из миноров матрицы A. Пусть $I, J \subseteq [n]$, а $A \in \mathrm{M}_n(F)$. Обозначим через A_I^J подматрицу в A, стоящую на пересечении строк с номерами из J и столбцов с номерами из I.

Пусть $L : V \to V$ – линейный оператор. Обозначим через $\Lambda^k L : \Lambda^k(V) \to \Lambda^k(V)$ оператор, заданный на разложимых тензорах формулой $\Lambda^k L(x_1 \wedge \cdots \wedge x_k) = L(x_1) \wedge \cdots \wedge L(x_k)$ и продолженный по линейности (таким образом, Λ^k превращается в функтор из категории векторных пространств в себя; аналогично можно сделать и для категории модулей над коммутативным кольцом). Если e – базис пространства V, то через $\Lambda^k e$ обозначим базис пространства $\Lambda^k(V)$ из предложения 5.5. Нам понадобится писать матрицу оператора в этом базисе, но упорядочивать его неудобно. Поэтому строки и столбцы матриц будут нумероваться k-элементными подмножествами в $[n]$, где $n = \dim V$, также как и базисные элементы $e_I = e_{i_1} \wedge \cdots \wedge e_{i_k}$.

Лемма 8.1. *Пусть V – n-мерное векторное пространство, а $I, J \subseteq [n]$ – подмножества порядка k. Если $L_e = A$, то элемент матрицы оператора $\Lambda^k L$ в базисе $\Lambda^k e$ в позиции (J, I) равны $\det A_I^J$*

Доказательство. По определению матрицы линейного оператора $L(e) = eA$, откуда $\big(L(e_{i_1}), \ldots, L(e_{i_k})\big) = eA_I^{[n]}$, где $I = \{i_1, \ldots, i_k\}$, $i_1 < \cdots < i_k$. По лемме 6.1

$$L(e_I) = L(e_{i_1}) \wedge \cdots \wedge L(e_{i_k}) = \sum_{|J|=k} e_J \det A_I^J,$$

что и требовалось. □

Так как Λ^k – функтор, то $\Lambda^k L^{-1} = (\Lambda^k L)^{-1}$. Мы хотим вычислить элементы матрицы этого оператора. Мы могли бы написать, что они равны алгебраическим дополнениям матрицы оператора $\Lambda^k L$, поделенным на ее определитель. Однако эта формула слишком громоздка. Мы вычислим обратную к матрице оператора $\Lambda^k L$, пользуясь двойственным базисом к $\Lambda^k e$, указанном в лемме 7.2.

Теорема 8.2. *Пусть $A \in \mathrm{GL}_n(F)$. Тогда $\det(A^{-1})_J^I = \frac{1}{\det A}(-1)^{s_{IJ}} \det A_{[n]\smallsetminus I}^{[n]\smallsetminus J}$, где $s_{IJ} = \sum_{l\in I\triangle J} l$.*

Доказательство. Пусть $L : F^n \to F^n$ – оператор умножения на матрицу A, так что $L_e = A$, где e – стандартный базис в F^n. Положим $B = \Lambda^k L_{\Lambda^k e}$. По предыдущей лемме

$$B^{-1} = \big(\det(A^{-1})^J_I\big)_{|I|=|J|=k} \text{ и } \Lambda^{n-k} L_{\Lambda^{n-k}e} = \Big(\det A^{[n]\smallsetminus J}_{[n]\smallsetminus I}\Big)_{|I|=|J|=k}.$$

Пусть f – двойственный к $\Lambda^k e$ базис пространства $\Lambda^{n-k}(V)$, найденный в лемме 7.2. Так как знаки элементов базиса f отличаются от знаков элементов базиса $\Lambda^{n-k}e$, то предыдущая формула превращается в

$$D := \Lambda^{n-k} L_f = \Big((-1)^{\varepsilon_{[n]\smallsetminus I}+\varepsilon_{[n]\smallsetminus J}} \det A^{[n]\smallsetminus J}_{[n]\smallsetminus I}\Big)_{|I|=|J|=k}.$$

Заметим, что

$$(\varepsilon_{[n]\smallsetminus I} + \varepsilon_{[n]\smallsetminus J}) \mod 2 = \Big(\sum_{l\notin I} l + \frac{(n-k)(n-k+1)}{2} + \sum_{l\notin J} l + \frac{(n-k)(n-k+1)}{2}\Big) \mod 2 =$$
$$\Big(\sum_{l=1}^{n} l - \sum_{l\in I} l + \sum_{l=1}^{n} l - \sum_{l\in J} l\Big) \mod 2 = \Big(\sum_{l\in I\triangle J} l\Big) \mod 2 = s_{IJ}.$$

По лемме 7.2 $(\Lambda^k e)^\intercal f = e_{[n]} \cdot E$ (произведение матриц с элементами из алгебры $\Lambda(F^n)$). Так как Λ – функтор, то L индуцирует эндоморфизм $\Lambda(L)$ алгебры $\Lambda(F^n)$, а также эндоморфизмы $\mathrm{M}_{r,s}\big(\Lambda(L)\big)$ модулей $\mathrm{M}_{r,s}\big(\Lambda(F^n)\big)$, сохраняющие произведения матриц. Допуская вольность речи, все эти отображения будут обозначаться через L.

Применяя L к последнему равенству, получаем

$$L(\Lambda^k e)^\intercal L(f) = L(e_{[n]})\cdot E \iff (\Lambda^k eB)^\intercal f D = e_{[n]} \det A \cdot E \iff B^\intercal D = \det A \cdot E \iff B^{-1} = \frac{1}{\det A} D^\intercal.$$

Сравнивая 2 различных выражения для элементов матрицы B^{-1}, получаем требуемые формулы. □

9. Грассманиан и соотношения Плюккера

Рассмотрим множество k-мерных подпространств в n-мерном векторном пространстве V. Оно обозначается через $\mathrm{Gr}_k(V)$ или Gr^n_k. В частности, $\mathrm{Gr}^n_1 = \mathbb{P}(V) = \mathbb{P}^{n-1}$ – проективное пространство.

Выберем базис $u = (u_1, \dots, u_k)$ подпространства $U \leqslant V$ и рассмотрим внешнее произведение $u_{[k]} = u_1 \wedge \dots \wedge u_k \in \Lambda^k(V)$. Если $w = uC$ – другой базис U, где $C \in \mathrm{GL}_k(F)$, то по лемме 6.1 $w_{[k]} = u_{[k]} \det C$. Получили корректно заданное отображение

$$\gamma : \mathrm{Gr}_k(V) \to \mathbb{P}\big(\Lambda^k(V)\big).$$

Лемма 9.1. *Отображение γ инъективно.*

Доказательство. Пусть $u = (u_1, \dots, u_k)$ базис подпространства U, $w = (w_1, \dots, w_k)$ – базис подпространства W, а $\gamma(W) = \gamma(U)$, т. е. $w_{[k]} = u_{[k]}c$ для некоторого $c \in F^*$. Дополним u до базиса $u' = (u_1, \dots, u_n)$ пространства V и выразим w через u': $w = u'A$, где $A \in \mathrm{M}_{nk}(F)$. По лемме 6.1 $u_{[k]}c = w_{[k]} = \sum\limits_{I\subseteq[n],\ |I|=k} u'_I \cdot \det A^I$. Так как набор u'_I линейно независим, то $A^I = 0$ при всех $I \neq [k]$, а так как w – базис, то $w_{[k]} \neq 0$, откуда матрица $A^{[k]}$ обратима. Домножая A на матрицу, обратную к $A^{[k]}$ получаем матрицу вида $\left(\begin{smallmatrix} E \\ B \end{smallmatrix}\right)$, у которой по прежнему все миноры, кроме верхнего, равны нулю. Легко видеть, что последнее условие влечет $B = 0$. Следловательно, и в матрице A все строки, начиная с $(k+1)$-й, нулевые. Из этого следует, что все w_i лежат в линейной оболочке набора u, откуда $W = U$. □

Таким образом, грассманиан (т. е. многообразие Грассмана) вкладывается в проективное пространство, что дает возможность определить на нем структуру алгебраического многобразия, или, в случае $F = \mathbb{R}$ или $F = \mathbb{C}$, аналитического многообразия. Это многообразие является замкнутым в проективном пространстве. Точнее, оно является множеством корней некоторого набора однородных многочленов. Полиномиальные уравнения, которым удовлетворяют точки образа отображения γ, называются соотношениями Плюккера.

Пусть $e = (e_1, \dots, e_n)$ – базис пространства V, а $\Lambda^k e$ – базис $\Lambda^k V$. Подпространству U с базисом $u = eA$, где A – матрица размера $n \times k$ ранга k, отображение γ ставит в соответствие прямую в $\Lambda^k V$, порожденную вектором $u[k] = \sum\limits_{I \subseteq [n],\ |I|=k} e_I \cdot \det A^I$. Координаты этого вектора в базисе $\Lambda^k e$ равны $(A^I \mid I \subseteq [n],\ |I| = k)$ и являются однородными координатами точки $\gamma(U) \in \mathbb{P}\big(\Lambda^k(V)\big)$. Они называются плюккеровыми координатами подпространства U. Соотношения Плюккера – это набор условий на эти координаты.

Для удобства записи соотношений Плюккера введем следующие обозначения. Пусть $x \in \mathbb{P}\big(\Lambda^k(V)\big)$ – точка с однородными координатами $(x_I \mid I \subseteq [n],\ |I| = k)$, а $J \in [n]$ – $(k-1)$-элементное подмножество. Положим $x_{J,l} = 0$, если $l \in J$. В противном случае $x_{J,l} = (-1)^m x_{J \cup \{l\}}$, где m – количество элементов J, больших l.

Теорема 9.2. *Точка $x \in \mathbb{P}\big(\Lambda^k(V)\big)$ принадлежит образу отображения γ тогда и только тогда, когда ее плюккеровы координаты удовлетворяют следующим уравнениям:*
для любых $J, L \subseteq [n]$, $|J| = k-1$, $|L| = k+1$

$$\sum_{l \in L} x_{J,l} x_{L \smallsetminus \{l\}} = 0.$$

Глава 11

Многочлены

Все кольца в этой главе предполагаются коммутативными с 1

Кольца многочленов над коммутативным кольцом R являются свободными объектами в категории коммутативных R-алгебр с 1, т. е. образами множеств под действием левого сопряженного к забывающему функтору в множества. Соответственно, любое кольцо является факторкольцом какого-то кольца многочленов. Довольно часто вопросы о произвольных коммутативных кольцах можно свести к конечнопорожденным кольцам. Поэтому важно изучать кольца многочленов над R от конечного числа переменных.

В настоящей главе мы докажем, что такие кольца являются нетеровыми, как только R обладает этим свойством. Кроме того, мы докажем теорему Гильберта о нулях, которая является началом классической алгебраической геометрии, изучим симметрические многочлены, играющие важнейшую роль в теории Галуа, а также узнаем, что такое базисы Грёбнера – инструмент позволяющий алгоритмически определять, является ли полиномиальное тождество следствием нескольких данных или нет. Стоит отметить, что базисы Грёбнера (система образующих идеала кольца многочленов, удовлетворяющая определенным свойствам) являются важнейшей частью современных систем компьютерной алгебры.

1. Определения кольца многочленов

Мы уже знаем несколько определений кольца многочленов. Давайте их перечислим. Пусть R – коммутативное кольцо с 1 (различные обобщения на некоммутативные кольца коэффициентов оставляются читателю в качестве упражнения). Введем обозначения: $T = \{t_1, \dots, t_n\}$, $R[T] = R[t_1, \dots, t_n]$ – кольцо многочленов над R от переменных $t_1, \dots, t_n$ – то, что мы сейчас определим. Если $a = (a_1, \dots, a_n) \in \mathbb{N}_0^n$ – мультииндекс, то положим $T^a := t_1^{a_1} \cdot \ldots \cdot t_n^{a_n}$. Сложение мультииндексов производится покомпонентно.

Мы пишем все определения для конечного множества T, однако конечность существенна только для рекурсивного определения.

(1) Школьное определение. $R[T] = \{ \sum\limits_{a \in \mathbb{N}_0^n} \alpha(a) T^a \mid \alpha : \mathbb{N}_0^n \to R$ – финитная функция$\}$. Вместо $\alpha(a)$ пишут α_a. Операции определены следующим образом:

$$\sum_{a \in \mathbb{N}_0^n} \alpha_a T^a + \sum_{a \in \mathbb{N}_0^n} \beta_a T^a = \sum_{a \in \mathbb{N}_0^n} (\alpha_a + \beta_a) T^a,$$

$$\Big(\sum_{a \in \mathbb{N}_0^n} \alpha_a T^a \Big) \cdot \Big(\sum_{a \in \mathbb{N}_0^n} \beta_a T^a \Big) = \sum_{a \in \mathbb{N}_0^n} \left(\sum_{b,c \in \mathbb{N}_0^n,\, b+c=a} (\alpha_b \beta_c) \right) T^a.$$

(2) То же самое с использованием понятия свободного модуля. $R[T]$ – это R-алгебра, которая является свободным R-модулем с множеством образующих $\{T^a \mid a \in \mathbb{N}_0^n\}$, а умножение на базисных элементах задается формулой $T^b T^c = T^{b+c}$. Так как умножение в R-алгебре R-билинейно, то на базисе оно может быть задано произвольно, а затем продолжено единственным образом. Для линейного отображения это просто определение свободного модуля, для билинейного – применение определения 2 раза.

(3) Рекурсивное определение: $R[t_1, \dots, t_n] = R[t_1, \dots, t_{n-1}][t_n]$. При этом кольцо многочленов от одной переменной определяем одним из вышеуказанных способов. Это

определение полезно для доказательств по индукции. Доказательство того, что оно эквивалентно предыдущим – рутина.

(4) Если $\mathscr{P}$ – левый сопряженный к забывающему функтору из категории коммутативных R-алгебр в категорию множеств, то $R[T] = \mathscr{P}(T)$. Другими словами, $R[T]$ – это R-алгебра, содержащая множество T, такая что любая функция из T в R-алгебру A однозначно продолжается до гомоморфизма R-алгебр $R[T] \to A$. Это универсальное свойство уже отмечалось в примерах 6.1 и 5.1 главы 9. Доказательство аналогично доказательству предложения 16.2 главы 4.

(5) Кольцо многочленов $R[T]$ – это симметрическая алгебра свободного R-модуля с базисом T. Эквивалентность этого определения предыдущему обсуждалась в теореме 5.4 главы 10.

На множестве одночленов $\{\alpha T^a \mid \alpha \in F, a \in \mathbb{N}_0^n\}$ можно различными способами ввести понятие степени. Наиболее употребительное – полная степень:

$$\deg(\alpha T^a) := \sum_{i=1}^{n} a_i.$$

При использовании рекурсивного определения кольца многочленов полезно пользоваться степенью по фиксированной переменной:

$$\deg_{t_k}(\alpha T^a) := a_k.$$

Кроме этого есть понятие мультистепени[1]:

$$\mathrm{mdeg}(\alpha T^a) := a.$$

Степень многочлена – это наибольшая из степеней его одночленов. При этом для полной степени и степени по фиксированной переменной это определение не требует пояснений, а вот для мультистепени неясно, что значит “наибольшая”. Это станет понятно в параграфе 3, когда мы превратим $\mathbb{N}_0^n$ в упорядоченный моноид. Для того чтобы не делать оговорок, будем считать, что степень нулевого многочлена равна символу $-\infty$, который меньше любой другой степени, а в сумме с чем угодно дает снова $-\infty$.

2. Нетеровость кольца многочленов

Настоящий параграф посвящен доказательству следующего результата.

Теорема 2.1 (теорема Гильберта о базисе)**.** *Кольцо многочленов от конечного числа переменных над нетеровым кольцом нетерово.*

Пусть R – кольцо, а $R[t]$ – кольцо многочленов над R. Обозначим через $R[t]_m$ подмодуль в $R[t]$, состоящий из многочленов, степени не выше m. Ясно, что это модуль (свободно) порожден элементами t^k, $k = 0, \ldots, m$. В доказательстве теоремы мы сведем вопрос о конечнопорожденности идеала к конечнопорожденности подмодуля в $R[t]_m$.

Доказательство теоремы. Применяя индукцию по количеству переменных и рекурсивное определение кольца многочленов, легко свести задачу к случаю одной переменной.

Для подмножества $X \subseteq R[t]$ обозначим через $L(X)$ множество старших коэффициентов многочленов из X. Легко видеть, что если X – идеал в $R[t]$, то $L(X)$ – идеал в R. Действительно, если $\alpha, \beta \in L(X)$, то существуют многочлены $p = \alpha t^m + \ldots$ и $q = \beta t^k + \ldots$ из X. Тогда $\gamma p = \gamma\alpha t^m + \cdots \in X$, откуда $\gamma\alpha \in L(X)$ для любого $\gamma \in R$. Считая для определенности, что $m \geqslant k$ имеем $t^{m-k}q + p = (\alpha + \beta)t^m + \cdots \in X$, откуда $\alpha + \beta \in L(X)$.

Пусть I – идеал в $R[t]$. Так как R нетерово, то идеал $L(I) \leqslant R$ конечнопорожден, скажем, элементами $\alpha_1, \ldots, \alpha_n$. Это значит, что существуют многочлены $p_i = \alpha_i t^{m_i} + \cdots \in I$. Пусть

[1] Термин не является общеупотребительным.

$q = \alpha t^k + \dots \in I$, где $k \geqslant m := \max m_i$. Так как $\alpha \in L(I)$, то существуют $\beta_i \in R$ такие, что $\alpha = \sum \alpha_i \beta_i$. Тогда $\deg(q - \sum \beta_i t^{k-m_i} p_i) < k$. Таким образом, вычитая линейную комбинацию многочленов $p_1, \dots, p_n$ мы можем уменьшить степень любого многочлена из I до $m-1$ (процесс аналогичен делению многочленов "в столбик"). Следовательно, $I = \langle p_1, \dots, p_n \rangle_{R[t]} + I \cap R[t]_{m-1}$. По предложению 11.1 главы 4 второе слагаемое правой части конечно порождено над R, следовательно, I конечно порожденный идеал, что завершает доказательство. □

3. Порядок на множестве мономов

В праграфах 3–6, посвященных базисам Гребнера, мы изучаем кольцо многочленов $F[T] = F[t_1, \dots, t_n]$ над произвольным полем F.

Задача: по данным многочленам $f, f_1, \dots, f_m \in F[t_1, \dots, t_n]$ определить, лежит ли f в идеале, порожденном f_i-ми.

Решение: научиться делить с остатком f на $f_1, \dots, f_m$.

Для этого надо определить, что такое остаток, а для этого – определить линейный порядок на множестве мономов.

Мы продожаем использовать обозначения из начала главы: $T = (t_1, \dots, t_n)$ и $T^a = t_1^{a_1} \dots t_n^{a_n}$, где $a = (a_1, \dots, a_n) \in \mathbb{N}_0^n$.

Определение 3.1. Пусть N – моноид в аддитивной записи, на котором задан линейный порядок $\preccurlyeq$, удовлетворяющий условиям

(1) $0 \prec a$ для любого $a \in N \setminus \{0\}$, и
(2) $a \preccurlyeq b \implies a + c \preccurlyeq b + c$ для любых $a, b, c \in N$.

Линейный порядок на $\mathbb{N}_0^n$, превращающей его в упорядоченный моноид, называется *допустимым*.

Везде далее $\preccurlyeq$ – допустимый линейный порядок на $\mathbb{N}_0^n$ или на абстрактном частично упорядоченном множестве.

Лемма 3.2. *Если T^a делится на T^b, то $b \preccurlyeq a$.*

В случае $b \preccurlyeq a$ говорят, что одночлен βT^b *младше* одночлена αT^a или что αT^a *старше* βT^b.

После того как мы ввели порядок на множестве мультииндексов, исчезло препятствие к тому, чтобы определить мультистепень многочлена. А именно, мультистепень многочлена $f = \sum\limits_{a \in \mathbb{N}_0^n} \beta_a T^a$ – это наибольшее (по отношению к заданому линейному порядку) a такое, что $\beta_a \neq 0$. Так как почти все β_a равны 0, то наибольший существует. Мультистепень многочлена f будем обозначать через $\operatorname{mdeg} f$. Обозначим через $LM(f)$ старший моном многочлена f, так что $\operatorname{mdeg} f = \operatorname{mdeg} LM(f)$.

Следующие свойства мультистепени почти очевидны.

Предложение 3.3. *Пусть $f, g \in R[T]$.*

(1) $LM(fg) = LM(f) \cdot LM(g)$.
(2) $\operatorname{mdeg}(fg) = \operatorname{mdeg} f + \operatorname{mdeg} g$.
(3) $\operatorname{mdeg}(f+g) \preccurlyeq \max(\operatorname{mdeg} f, \operatorname{mdeg} g)$, *причем неравенство возможно, только если* $LM(f) = -LM(g)$.

Тремя основными примерами допустимых упорядочений являются:

`Lex:` – лексикографический порядок;
`InvLex:` – обратный лексикографический порядок (т. е. сначала учитывается степень старшей буквы); и
`DegLex:` – сначала сравниваются полные степени, а многочлены одинаковой степени упорядочиваются лексикографически.

Разумеется, каждый из этих порядков зависит от выбора линейного порядка на множестве переменных. Обычно подразумевается, что $t_1 \prec t_2 \ldots \prec t_n$.

Для обратного лексикографического порядка одночлены располагаются точно так же, как слова в словаре. Под чисто лексикографическим порядком в большинстве работ по компьютерной алгебре подразумевается именно `InvLex`.

Однако чисто лексикографический порядок не всегда удобен. Дело в том, что сложность проводимых с помощью `Lex` вычислений часто драматическим образом зависит от порядка переменных. Для вычислений обычно значительно удобнее пользоваться `DegLex`.

Для проведения индукции по мультистепени многочлена необходимо использовать условие обрыва убывающих цепей (descending chain condition, сокращенно DCC). Будем говорить, что частично упорядоченное множество X удовлетворяет DCC, если не существует бесконечных строго убывающих цепочек. Другими словами, для любой последовательности $x_k \in X$, $x_k \succcurlyeq x_{k+1}$ для любого $k \in \mathbb{N}$, найдется такое $n \in \mathbb{N}$, что $x_i = x_n$ при всех $i \geqslant n$. Из леммы Цорна вытекает следующее утверждение.

Лемма 3.4. *Для частично упорядоченного множества X DCC равносильно тому, что любое подмножество в X содержит минимальный элемент.*

Если X линейно упорядочено, то этот минимальный элемент является наименьшим.

Лемма 3.5. *Пусть $N = N_1 \oplus N_2$ – линейно упорядоченный моноид с отношением $\preccurlyeq$. Если N_1 и N_2 удовлетворяют DCC, то и N обладает этим свойством.*

Доказательство. Обозначим через π отображение проекции N на N_1. Рассмотрим множество $X = X_0 \subseteq N$ и найдем в нем наименьший элемент. Пусть a_0 – наименьший элемент множества $\pi(X)$, а b_0 – наименьший элемент множества $\{b \in N_2 \mid (a_0, b) \in X\}$ (такие существуют по DCC для N_1 и N_2). Пусть далее $X_1 = \{x \in X \mid x \prec (a_0, b_0)\}$. Если $X_1 = \varnothing$, то (a_0, b_0) искомый наименьший элемент в X. В противном случае продоложим процесс, построив элементы $(a_i, b_i) \in X_i$. Заметим, что $(a_i, b_i) \succ (a_{i+1}, b_{i+1})$, но $a_i \prec a_{i+1}$. Отсюда следует, что $b_i \succ b_{i+1}$ (именно здесь используются условия из определения 3.1), и по DCC для N_2 цепочка $b_0 \succ b_1 \succ \ldots$ конечна. Это означает, что для некоторого $m \in \mathbb{N}$ множество X_m пусто, откуда (a_{m-1}, b_{m-1}) – наименьший элемент множества X. □

Следующее утверждение выводится из предыдущего индукцией по n, учитывая, что $\mathbb{N}_0$ удовлетворяет DCC.

Следствие 3.6. *Допустимый порядок на $\mathbb{N}_0^n$ удовлетворяет DCC.*

Другое доказательство следствия[2] также проводится индукцией по n. Пусть $X = \{a^i \in \mathbb{N}_0^n \mid i \in \mathbb{N}_0\}$, где $a^i \succ a^{i+1}$ при любом i, – убывающая цепочка. Для любых $i = 1, \ldots, n$ и $k \in \mathbb{N}_0$ множество элементов из X, у которых на i-м месте стоит k, образуют подцепочку в X, которая обрывается по DCC для $\mathbb{N}_0^{n-1}$. Поэтому множества $X_i^k = \{b \in \mathbb{N}_0^{n-1} \mid (b_1, \ldots, b_{i-1}, k, b_i, \ldots, b_{n-1}) \in X\}$ конечны. Следовательно, $\widetilde{X} = \bigcup\limits_{i=1}^{n} \bigcup\limits_{k=0}^{a_i^0} X_i^k$ конечно и содержит a^0. Наименьший элемент d множества $\widetilde{X}$ существует и не превосходит a^0. Осталось заметить, что компоненты любого элемента $c \in X \setminus \widetilde{X}$ не меньше, чем соответсвующие компоненты a^0, поэтому $c \succcurlyeq a^0$. Таким образом, d является наименьшим элементом X.

4. Деление с остатком и базисы Гребнера

Мы умеем делить с остатком многочлены от одной переменной. Принципиальное отличие деления в общем случае состоит в следующем. Для одной переменной всегда можно убрать именно *старший* член делимого. В общем случае может оказаться, что старший моном $LM(f)$

[2] которое в 2022 было придумано и озвучено на лекции Федором Ушаковым

многочлена f не делится ни на один из старших мономов $LM(f_1), \dots, LM(f_m)$ но, тем не менее, какие-то другие члены f на них делятся. Это заставляет нас модифицировать процедуру деления с остатком следующим образом.

Рассмотрим старший среди тех мономов многочлена f, которые делятся на старший моном какого-то из многочленов $f_1, \dots, f_m$. Пусть, например, αT^a – член f, делящийся на $LM(f_i) = \beta T^b$. Тогда в многочлене $f - \frac{\alpha}{\beta} T^{a-b} f_i$ коэффициент при T^k равен нулю. А так как $\mathrm{mdeg}(\frac{\alpha}{\beta} T^{a-b} f_i) = a - b + \mathrm{mdeg}(f_i) = a$, то никаких членов старше, чем T^a при этом не появляется.

Таким образом, при каждой такой операции старший из тех членов многочлена f, которые делятся на какой-то из $LM(f_i)$, *строго убывает*. Из DCC следует, что продолжая действовать таким образом, проведя *конечное число* элементарных редукций мы придем к многочлену r, ни один из членов которого не делится на какой-либо из $LM(f_i)$. В этом случае говорят, что f *редуцируется* к r посредством $f_1, \dots, f_m$.

Таким образом мы доказали следующую теорему.

Теорема 4.1. *Для любых $f, f_1, \dots, f_m \in F[T]$ таких, что не все f_i равны 0, найдутся многочлены $q_1, \dots, q_m, r \in F[T]$ со следующими свойствами.*

(1) $f = q_1 f_1 + \dots + q_m f_m + r$.
(2) *Ни один из членов r не делится ни на один из одночленов $LM(f_1), \dots, LM(f_m)$.*
(3) $\mathrm{mdeg}(q_i f_i) \preccurlyeq \mathrm{mdeg}\, f$ *при всех* $i = 1, \dots, m$.

Любой многочлен r, удовлетворяющий условиям теоремы, будет называться **остатком** от деления f на $f_1, \dots, f_m$. Проблема, однако, состоит в том, что остаток от деления f на $f_1, \dots, f_m$ определен неоднозначно. Неоднозначность может возникать из-за того, что αT^a может делиться на старшие мономы *нескольких* многочленов $f_1, \dots, f_m$.

Пусть $f(x, y) = x^2 y$, $f_1(x, y) = x^2$, $f_2(x, y) = xy - 1$. Тогда $f = f_1 y + 0 = f_2 x + x$, и оба равенства удовлетворяют определению деления с остатком f на f_1, f_2. Заметим, что в этом случае многочлен x принадлежит идеалу, порожденному f_1, f_2, но не делится на них без остатка, потому что неравенства $\mathrm{mdeg}(f_i q_i) \preccurlyeq \mathrm{mdeg}(x)$ могут выполняться только при $q_1 = q_2 = 0$.

Оказывается, что для любого набора многочленов $f_1, \dots, f_m \in F[T]$ найдутся $g_1, \dots, g_s \in F[T]$, порождающие тот же самый идеал I, для которых деление с остатком дает *единственный* остаток.

Определение 4.2. Пусть $f_1, \dots, f_m \in F[T]$ порождают идеал I. Говорят, что $f_1, \dots, f_m$ образуют *базис Гребнера* идеала I, если старший член любого многочлена $f \in I \setminus \{0\}$, делится на старший член какого-то из многочленов $f_1, \dots, f_m$.

Как видно, при добавлении чего угодно к базису Гребнера он не перестает быть базисом Гребнера. Поэтому логично было бы называть такой набор многочленов “системой образующих Гребнера”, а базисом Гребнера называть минимальный или даже редуцированный базис Гребнера, о которых пойдет речь ниже. К сожалению, терминология уже сложилась, и мы не вправе ее менять.

Обычно базис Гребнера нормируется таким образом, чтобы в каждом f_i коэффициент при старшем члене равнялся 1.

Теорема 4.3. *Следующие условия эквивалентны*

(1) *$f_1, \dots, f_m$ — базис Гребнера идеала I.*
(2) *Остаток от деления любого многочлена $f \in F[T]$ на $f_1, \dots, f_m$ определен однозначно.*
(3) *Для любого $f \in I$ остаток от деления f на $f_1, \dots, f_m$ единственен и равен 0.*
(4) *Идеал $L(I)$, порожденный $LM(f)$, $f \in I$, порождается $LM(f_1), \dots, LM(f_m)$.*

Доказательство. (1) $\Longrightarrow$ (2). Предположим, что $f_1, \dots, f_m$ – базис Гребнера идеала I, а $f = q_1 f_1 + \dots + q_m f_m + r_1 = p_1 f_1 + \dots + p_m f_m + r_2$, где r_1 и r_2 — два остатка при делении

f на $f_1, \dots, f_m$. Тогда $r_1 - r_2 = (p_1 - q_1)f_1 + \ldots + (p_m - q_m)f_m \in I$ и по определению базиса Гребнера $LM(r_1 - r_2)$ делится на старший член какого-то из многочленов $f_1, \dots, f_m$. Однако по определению остатка ни у r_1, ни у r_2 нет вообще *никаких* членов, делящихся на старшие члены $f_1, \dots, f_m$.

(2) $\Longrightarrow$ (3). Пусть $f = q_1f_1 + \ldots + q_mf_m \in I$ – выражение f через образующие идеала I. Предположим для определенности, что максимум $\operatorname{mdeg}(q_if_i)$ достигается при $i = 1$. Если $\operatorname{mdeg}(q_1f_1) \preccurlyeq \operatorname{mdeg}(f)$, то остаток от деления f на $f_1, \dots, f_m$ равен нулю.

В противном случае, $\operatorname{mdeg}(f - q_1f_1) = \operatorname{mdeg}(q_1f_1) \succcurlyeq \operatorname{mdeg}(q_if_i)$. Следовательно, равенство $f - q_1f_1 = q_2f_2 + \ldots + q_mf_m + 0$ является делением с остатком. Пусть теперь $f = p_1f_1 + \cdots + p_mf_m + r$ – деление с остатком f на $f_1, \dots, f_m$, т. е. $\operatorname{mdeg}(p_if_i) \preccurlyeq \operatorname{mdeg}(f)$. Тогда равенство $f - q_1f_1 = (p_1 - q_1)f_1 + p_2f_2 + \cdots + p_mf_m + r$ также будет делением с остатком. Действительно, мультистепени всех слагаемых правой части, кроме первого, не больше $\operatorname{mdeg}(f) \prec \operatorname{mdeg}(f - f_1q_1)$. Для первого же слагаемого

$$\operatorname{mdeg}\big((p_1 - q_1)f_1\big) \preccurlyeq \max\big(\operatorname{mdeg}(p_1f_1), \operatorname{mdeg}(q_1f_1)\big) = \operatorname{mdeg}(q_1f_1) = \operatorname{mdeg}(f - f_1q_1).$$

По единственности остатка в любом случае $r = 0$.

(3) $\Longrightarrow$ (4). По предположению для $f \in I$ остаток от деления f на $f_1, \dots, f_m$ равен 0 и, таким образом, $f = q_1f_1 + \ldots + q_mf_m$. По определению деления с остатком $\operatorname{mdeg}(q_if_i) \preccurlyeq \operatorname{mdeg}(f)$. Поэтому найдется $i = 1, \dots, m$ такой, что $LM(f)$ ассоциирован с $LM(q_if_i) = LM(q_i)LM(f_i)$.

(4) $\Longrightarrow$ (1). Последнее условие означает, что для любого $f \in I$ существуют $q_1, \dots, q_m \in F[T]$ такие, что $LM(f) = LM(f_1)q_1 + \cdots + LM(f_m)q_m$. Ясно, что все мономы p каждого из многочленов q_i, для которых $\operatorname{mdeg}(f_ip) \neq \operatorname{mdeg}(f)$, можно выкинуть. Какой-то из q_i останется после выкидывания ненулевым одночленом. Тогда $LM(f)$ делится на $LM(f_i)$. □

В частности, теорема означает, что базис Гребнера – это в точности такой базис, для которого каждый элемент $f \in I$ представляется в виде

$$f = q_1f_1 + \ldots + q_mf_m, \text{ где } \operatorname{mdeg}(q_if_i) \preccurlyeq \operatorname{mdeg}(f).$$

Теперь мы готовы доказать, что в любом идеале существует базис Гребнера. Фактически, это сразу вытекает из теоремы Гильберта о базисе.

Теорема 4.4 (Хиронака). *В любом идеале I кольца многочленов $F[T]$ существует базис Гребнера.*

Доказательство. Рассмотрим идеал $L(I)$ порожденный $LM(f)$ по всем $f \in I$. По теореме Гильберта кольцо многочленов нетерово, поэтому можно из любой системы образующих можно выбрать конечную подсистему образующих $LM(f_1), \dots, LM(f_j)$. Дополнив набор $f_1, \dots, f_j$ до системы образующих идеала I получим базис Гребнера этого идеала. □

5. Алгоритм Бухбергера

Предшествующие теоремы не являются конструктивными. При помощи следующего *алгоритма Бухбергера* можно узнать за конечное число шагов, будет ли набор многочленов $f_1, \dots, f_m$ базисом Гребнера порожденного ими идеала I.

Пусть $\alpha T^a = LM(f)$, $\beta T^b = LM(g)$, а $T^c = \gcd(T^a, T^b)$. Определим s-многочлен многочленов f и g посредством[3]

$$s(f, g) := \frac{\beta T^b}{T^c}f - \frac{\alpha T^a}{T^c}g = \frac{LM(g)f - LM(f)g}{\gcd\big(LM(f), LM(g)\big)},$$

где для определенности наибольший общий делитель выбирается унитальным. Он строится так, чтобы старшие члены f и g сокращались. Оказывается, равенство нулю остатка достаточно проверять только для конечного числа s-многочленов.

[3]Более стандартное определение отличается от нашего делением на $\alpha\beta$.

Теорема 5.1 (Критерий Бухбергера). *Многочлены $f_1, \dots, f_m$ в том и только том случае образуют базис Гребнера порожденного ими идеала I, когда все s-многочлены $s(f_i, f_j)$, $i \neq j$, дают при делении на $f_1, \dots, f_m$ остаток 0.*

В одну сторону утверждение следует из теоремы 4.3. Для доказательства обратной импликации нам потребуются следующие леммы.

Лемма 5.2. *Если g_1, g_2 – одночлены, то $s(f_1g_1, f_2g_2)$ делится на $s(f_1, f_2)$.*

Доказательство.

$$s(f_1g_1, f_2g_2) = \frac{LM(f_2g_2)f_1g_1 - LM(f_1g_1)f_2g_2}{\gcd\big(LM(f_1g_1), LM(f_2g_2)\big)} = \frac{g_1g_2(LM(f_2)f_1 - LM(f_1)f_2)}{\gcd\big(LM(f_1g_1), LM(f_2g_2)\big)} =$$
$$s(f_1, f_2)\frac{g_1g_2\gcd\big(LM(f_1), LM(f_2)\big)}{\gcd\big(LM(f_1g_1), LM(f_2g_2)\big)} = s(f_1, f_2)\frac{\gcd\big(g_1g_2LM(f_1), g_1g_2LM(f_2)\big)}{\gcd\big(g_1LM(f_1), g_2LM(f_2)\big)}$$

Последняя дробь является одночленом, потому что мономы под знаком gcd в числителе делятся на соответствующие мономы под знаком gcd в знаменателе. □

Лемма 5.3. *Пусть $f_1, \dots, f_m$ – многочлены одинаковой мультистепени a, а $f = \sum \lambda_i f_i$, где $\lambda_i \in F$. Если $\operatorname{mdeg} f \prec a$, то f является линейной комбинацией многочленов $s(f_i, f_j)$ с коэффициентами из F.*

Доказательство. Так как по условию $f_i = \alpha_i T^a + \dots$ для некоторых $\alpha_i \in F \setminus \{0\}$, то $s(f_i, f_j) = \alpha_j f_i - \alpha_i f_j$. С другой стороны, так как коэффициент f при T^a равен 0, то $\lambda_1\alpha_1 + \dots + \lambda_m\alpha_m = 0$. Таким образом:

$$\sum_{i=1}^{m-1} \frac{\lambda_i}{\alpha_m} s(f_i, f_m) = \sum_{i=1}^{m-1} \lambda_i f_i - \Big(\sum_{i=1}^{m-1} \lambda_i\alpha_i\Big)\frac{f_m}{\alpha_m} = \sum_{i=1}^{m} \lambda_i f_i - \Big(\sum_{i=1}^{m} \lambda_i\alpha_i\Big)\frac{f_m}{\alpha_m} = f$$

□

Доказательство теоремы. Пусть $g_1, \dots, g_m \in F[T]$, а $f = f_1g_1 + \dots + f_mg_m$ – произвольный элемент из I. Предположим, что $a := \max\big(\operatorname{mdeg}(f_1g_1), \dots, \operatorname{mdeg}(f_mg_m)\big)$ – наименьшая возможная мультистепень для всех таких представлений f. Мы хотим доказать, что f делится на $f_1, \dots, f_m$ без остатка, т. е. $a \preccurlyeq \operatorname{mdeg}(f)$. Переставляя при необходимости слагаемые, можно записать f в виде $f = \sum_{i=1}^{k} f_ih_i + \sum_{j=1}^{m} f_j\hat{g}_j$, где h_i – одночлены, а $\operatorname{mdeg}(f_ih_i) = a \succ \operatorname{mdeg}(f_j\hat{g}_j)$ при всех $i = 1, \dots, k$ и $j = 1, \dots, m$. Если $a \succ \operatorname{mdeg}(f)$, то сумма старших членов равна нулю. Тогда, по предыдущей лемме сумма $f_1h_1 + \dots + f_kh_k$ представляется в виде линейной комбинации s-многочленов $s(f_ih_i, f_jh_j)$ с коэффициентами из поля:

$$f_1h_1 + \dots + f_kh_k = \sum_{i<j} \mu_{ij} s(f_ih_i, f_jh_j).$$

Так как $s(f_i, f_j)$ делится на $f_1, \dots, f_m$ без остатка, а $s(f_ih_i, f_jh_j)$ делится на $s(f_i, f_j)$, то и $s(f_ih_i, f_jh_j)$ делится на $f_1, \dots, f_m$ без остатка. Следовательно,

$$f_1h_1 + \dots + f_kh_k = \sum_{i<j} \mu_{ij} s(f_ih_i, f_jh_j) = f_1\tilde{h}_1 + \dots + f_m\tilde{h}_m,$$

где $\operatorname{mdeg}\big(f_i\tilde{h}_i\big) \preccurlyeq \operatorname{mdeg}\big(\sum_{i<j} \mu_{ij} s(f_ih_i, f_jh_j)\big) \prec a$. Но это означает, что в исходном представлении f мультистепень слагаемых можно уменьшить, что противоречит предположению. □

Эта теорема показывает, как *построить* базис Гребнера идеала. Пусть $f_1, \dots, f_m$ – *какая-то* система образующих идеала I. Образуем s-многочлены $s(f_i, f_j)$ для всех $i < j$ и вычислим их остатки при делении на $f_1, \dots, f_m$. Добавляя к набору $f_1, \dots, f_m$ все получившиеся ненулевые остатки, составим набор $f_1, \dots, f_{m_1}$, где $m_1 \geqslant m_0 := m$. Продолжая действовать таким образом,

на i-ом шаге получим набор $f_1, \dots, f_{m_i}$. Равенство $m_i = m_{i-1}$ означает, что все s-многочлены делятся на $f_1, \dots, f_{m_{i-1}}$ без остатка. В этом случае по критерию Бухбергера $f_1, \dots, f_{m_{i-1}}$ уже базис Гребнера. В противном случае добавились многочлены, старшие члены которых, не выражаются через $LM(f_1), \dots, LM(f_{m_{i-1}})$ (это верно для любого ненулевого остатка от деления на $f_1, \dots, f_m$). Получим строго возрастающую цепочку идеалов, порожденных $LM(f_1), \dots, LM(f_{m_i})$, которая должна оборваться по теореме Гильберта о базисе.

Изложенная выше процедура является простейшим вариантом *алгоритма Бухбергера*. Разумеется, фактически настоящий алгоритм Бухбергера устроен чуть сложнее и работает гораздо быстрее, так как в нем на каждом шаге производится исключение лишних образующих и редукция каждой образующей по модулю остальных. Как это делается, мы сейчас увидим.

6. Минимальные и редуцированные базисы Гребнера

Базис Гребнера $f_1, \dots, f_m$ называется *минимальным*, если старшие мономы двух любых многочленов f_i и f_j, $i \neq j$, не делятся друг на друга.

Теорема 6.1. *В каждом идеале кольца многочленов существует минимальный базис Гребнера.*

Теорема сразу вытекает из следующего утверждения.

Лемма 6.2. *Если $f_1, \dots, f_m$ – базис Гребнера, а $LM(f_m)$ делится на $LM(f_i)$ для некоторого $i < m$, то $f_1, \dots, f_{m-1}$ – базис Гребнера того же идеала.*

Доказательство. Пусть I – идеал, порожденный $f_1, \dots, f_m$, а $r = f_m - \sum_{i=1}^{m-1} f_j g_j$ – остаток от деления f_m на $f_1, \dots, f_{m-1}$. Так как $r \in I$, то по определению базиса Гребнера его старший член делится на старший член какого-то f_k. По условию существует деление, при котором $LM(r) \prec LM(f_m)$, поэтому $k \neq m$. Но, если остаток ненулевой, то это невозможно по определению остатка. Таким образом, $f_m \in (f_1, \dots, f_{m-1})$ и $LM(f_m) \in \big(LM(f_1), \dots, LM(f_{m-1})\big)$. □

Оказывается, старшие мономы многочленов, входящих в минимальный базис Гребнера, определены однозначно, с точностью до перестановки.

Предложение 6.3. *Пусть $f_1, \dots, f_m$ и $g_1, \dots, g_k$ – два минимальных базиса Гребнера идеала I. Тогда $k = m$ и после перестановки* $\operatorname{mdeg}(f_i) = \operatorname{mdeg}(g_i)$.

Доказательство. Так как $g_1, \dots, g_k$ базис Гребнера, а $f_1 \in I$, то старший член f_1 делится на старший член какого-то g_i. Перенумеровывая, если нужно, g_i, можно считать, что $LM(f_1)$ делится на $LM(g_1)$. Поменяв в предыдущем рассуждении f и g, мы видим, что в свою очередь $LM(g_1)$ делится на старший член какого-то f_i. Если при этом $i \neq 1$, то $LM(f_1)$ делится на $LM(f_i)$, $i \neq 1$, что противоречит минимальности. Таким образом, старшие мономы f_1 и g_1 пропорциональны, т. е. $\operatorname{mdeg}(f_1) = \operatorname{mdeg}(g_1)$.

Продолжая действовать таким образом, мы видим, что существует перестановка, для которой $\operatorname{mdeg}(f_i) = \operatorname{mdeg}(g_i)$ при всех $i \leqslant \min(m, k)$. Если $m > k$, то, так как $g_1, \dots, g_k$ – базис Гребнера, то $LM(f_{k+1})$ делится на $LM(g_j) = LM(f_j)$ для некоторого $j \leqslant k$, что снова противоречит минимальности. Аналогично, $k > m$ невозможно, откуда $k = m$. □

Базис Гребнера $f_1, \dots, f_m$ называется *редуцированным*, если ни один из членов любого многочлена f_i не делится на старшие члены многочленов f_j, $j \neq i$. Другими словами, каждый элемент этого базиса редуцирован по модулю остальных. Иначе говоря, остаток при делении f_i на $\{f_j \mid j \neq i\}$, совпадает с f_i.

Теорема 6.4. *В каждом идеале кольца многочленов существует единственный с точностью до перестановки редуцированный базис Гребнера.*

Доказательство. Стартуем с минимального базис Гребнера $f_1, \dots, f_m$. Пусть g_1 – остаток от деления f_1 на $f_2, \dots, f_m$. В силу минимальности базиса Гребнера $\mathrm{mdeg}(g_1) = \mathrm{mdeg}(f_1)$. Поэтому $g_1, f_2, \dots, f_m$ – снова является базисом Гребнера того же идеала, причем ни один из мономов g_1 не делится на старшие мономы остальных многочленов. Продолжая процесс, получим редуцированный базис $g_1, \dots, g_m$.

Пусть теперь $f_1, \dots, f_m$ и $g_1, \dots, g_k$ – два редуцированных базиса Гребнера идеала I. Так как эти базисы минимальны, то из предыдущей теоремы вытекает, что $k = m$ и, после перестановки, $\mathrm{mdeg}(f_i) = \mathrm{mdeg}(g_i)$. Предположим, что $f_i \neq g_i$ для какого-то i. По определению базиса Гребнера $LM(f_i - g_i)$ делится на один из $LM(f_j) = LM(g_j)$. Так как $\mathrm{mdeg}(f_i - g_i) \prec \mathrm{mdeg}(f_i)$, то $j \neq i$. Но $LM(f_i - g_i)$ пропорционален одному из мономов многочлена f_i или g_i, которые не делятся на $LM(f_j) = LM(g_j)$ при $j \neq i$ по определению редуцированного базиса. Таким образом, $f_i = g_i$ для всех i. □

7. Симметрические многочлены

Пусть R – коммутативное кольцо. Многочлен $f \in R[T] := R[t_1, \dots, t_n]$ называется симметрическим, если он не меняется при перестановке аргументов. Более строго: определим правое действие симметрической группы S_n на $R[T]$ по правилу

$$f^\sigma(x_1, \dots, x_n) = f(x_{\sigma(1)}, \dots, x_{\sigma(n)}).$$

Таким образом, $\sigma \in S_n$ задает автоморфизм R-алгебры $R[T]$. Множество неподвижных точек под действием любого автоморфизма является подалгеброй. Неподвижный элемент под действием всей группы S_n называется симметрическим многочленом. Таким образом, множество всех симметрических многочленов является подалгеброй в A.

Рассмотрим многочлен

$$f(x) = (x - t_1) \dots (x - t_n) = x^n - s_1^{(n)} x^{n-1} + \dots + (-1)^n s_n^{(n)}$$

из кольца $R[T][x]$. Так как f не меняется при перестановке переменных $t_1, \dots, t_n$, то его коэффициенты (с точностью до знака) $s_1^{(n)}, \dots, s_n^{(n)}$ являются симметрическими многочленами из $R[T]$. Они называются *элементарными симметрическими многочленами*. Верхний индекс (n) обычно опускается. Легко видеть, что s_1 – сумма переменных, s_2 – сумма всех попарных произведений переменных и т. д (этот факт обычно называется теремой Виета).

Очевидно также, что $\deg s_k = k$, а $s_k^{(n)}(t_1, \dots, t_{n-1}, 0) = s_k^{(n-1)}(t_1, \dots, t_{n-1})$. Основная цель этого параграфа – доказать что любой симметрический многочлен единственным образом выражается, как многочлен от элементарных симметрических многочленов. Обозначим $S = (s_1^{(n)}, \dots, s_n^{(n)})$.

Теорема 7.1. *Алгебра симметрических многочленов порождена элементарными симметрическими многочленами, которые являются алгебраически независимыми.*

Другими словами, гомоморфизм R-алгебр $\varphi : R[z_1, \dots, z_n] \to R[T]$, посылающий z_k в s_k, является инъективным, а его образ равен алгебре симметрических многочленов.

В этом параграфе будем использовать алфавитный порядок на множестве мультииндесов $\mathbb{N}_0^n$: $a > b$, если первый ненулевой член последовательности $(a_1 - b_1, \dots, a_n - b_n)$ положителен.

Перед доказательством теоремы сформулируем несложную лемму.

Моном αT^a называется монотонным, если его мультистепень является монотонно убывающей последовательностью: $a_1 \geqslant a_2 \geqslant \dots \geqslant a_n$.

Лемма 7.2. *Старший моном симметрического многочлена является монотонным.*

Доказательство. Если моном αT^a входит в симметрический многочлен f, то и $\alpha t_1^{a_1} \dots t_i^{a_{i+1}} t_{i+1}^{a_i} \dots t_n^{a_n}$ также входит в f. Поэтому, если αT^a – старший моном, то $a_i \geqslant a_{i+1}$ для любого i. □

Вычисление образа φ. Предположим, что множество симметрических многочленов, не лежащих в $R[S] = \operatorname{Im}\varphi$, непусто. Пусть f – симметрический многочлен из этого множества наименьшей мультистепени (такой многочлен существует, потому что лексикографический порядок является допустимым и, следовательно, удовлетворяет DCC). Пусть $LM(f) = \alpha T^a$.

По пункту (1) предложения 3.3

$$LM(S^b) = LM(s_1^{b_1} s_2^{b_2} \dots s_n^{b_n}) = t_1^{b_1}(t_1t_2)^{b_2} \dots (t_1 \dots t_n)^{b_n} = \prod_{i=1}^{n} t_i^{\sum_{j=i}^n b_j}$$

(здесь все многочлены унитальные, поэтому не важно, что R не поле). Положив $b_i = a_i - a_{i+1}$, где $a_{n+1} := 0$, получим $LM(\alpha S^b) = \alpha T^a$ (по лемме 7.2 все b_i неотрицательны). Следовательно, $\operatorname{mdeg}(f - \alpha S^b) \prec \operatorname{mdeg} f$. Из минимальности a следует, что $f - \alpha S^\ell \in R[S]$, следовательно, и $f \in R[S]$ – противоречие. □

На самом деле, в процессе доказательства мы получили соотношение между степенями исходного симметрического многочлена и того многочлена, через который он выражается.

Следствие 7.3. *Пусть $f \in R[T]$ – симметрический многочлен, а $g \in R[z_1, \dots, z_n]$ таков, что $f\big(T\big) = g\big(s_1(T), \dots, s_n(T)\big)$. Тогда $\deg g = \deg_{t_1} f$ (напомним, что $\deg_{t_1} f$ обозначает степень f по переменной t_1).*

Доказательство. В обозначениях предыдущего доказательства степень g равна максимально возможной сумме $b_1 + \dots + b_n = a_1$. А максимальное a_1 как раз и есть степень f, как многочлена от x_1. Осталось заметить, что разным мультииндексам a соответствуют разные мультииндексы b, поэтому одночлен старшей степени не может сократиться. □

Вычисление ядра φ. Если $g(s_1, \dots, s_n) = 0$, то по предыдущему следствию $\deg g = -\infty$, т. е. $g = 0$. Это наблюдение завершает доказательство теоремы 7.1. □

8. Степенные суммы

Степенной суммой называется многочлен

$$p_k = p_k^{(n)} = t_1^k + \dots + t_n^k.$$

Так как степенные суммы – симметрические многочлены, то они выражаются через основные симметрические. Если же R является $\mathbb{Q}$-алгеброй, то и основные симметрические многочлены выражаются через первые n степенных сумм. Условие на R необходимо, для того чтобы можно было делить на натуральные числа.

Положим $s_0 = 1$ и $s_k^{(n)} = 0$ при $k > n$.

Теорема 8.1. *Основные симметрические многочлены и степенные суммы связаны следующими соотношениями, которые называются соотношения Ньютона:*

$$k s_k = \sum_{i=1}^{k} (-1)^{i+1} s_{k-i} p_i \qquad \text{при любом } k \geqslant 1.$$

Доказательство. Так как доказываемые равенства не содержат ничего, кроме целых чисел и независимых переменных, можно считать, что $R = \mathbb{Z}$, а потом спроектировать доказанное в кольцо многочленов над любым кольцом R.

Подставив $x = t_j$ в тождество

$$(x - t_1) \dots (x - t_n) = x^n - s_1^{(n)} x^{n-1} + \dots + (-1)^n s_n^{(n)}$$

получим

$$0 = (-1)^n s_n^{(n)} + \sum_{i=1}^{n} (-1)^{n-i} t_j^i s_{n-i}^{(n)}.$$

Суммируя по всем $j = 1, \ldots, n$ и сокращая на $(-1)^n$ получим

$$0 = ns_n^{(n)} + \sum_{i=1}^{n}(-1)^i p_i^{(n)} s_{n-i}^{(n)},$$

что совпадает с доказываемым равенством для $k = n$.

Напишем доказанное равенство для $k > n$ переменных:

$$0 = ks_k^{(k)} + \sum_{i=1}^{k}(-1)^i p_i^{(k)} s_{k-i}^{(k)},$$

и положим $t_{n+1} = \cdots = t_k = 0$. Тогда $s_i^{(k)}$ превратятся в $s_i^{(n)}$, степенные суммы $p_i^{(n+k)}$ – в $p_i^{(n)}$, а последняя формула – в требуемую формулу для $k > n$.

Случай $k \leqslant n$ будем доказывать индукцией по $n - k$. База индукции, $n - k = 0$, уже доказана. При $k < n$ рассмотрим могочлен $f = ks_k^{(n)} - \sum\limits_{i=1}^{k}(-1)^{i+1} s_{k-i}^{(n)} p_i^{(n)}$. Подставив в это равенство $t_j = 0$ получим

$$f(t_1, \ldots, t_{j-1}, 0, t_{j+1}, \ldots, t_n) = ks_k^{(n-1)}(t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_n) - \sum_{i=1}^{k}(-1)^{i+1} s_{k-i}^{(n-1)}(t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_n) p_i^{(n-1)}(t_1, \ldots, t_{j-1}, t_{j+1}, \ldots, t_n).$$

По индукционному предположению правая часть равенства равна нулю. Следовательно, f делится на t_j для любого j. Так как t_j неприводимы, а $\mathbb{Z}[T]$ факториально, то f делится на $t_1 \ldots t_n$. Но $\deg f = k < n$, поэтому $f = 0$, что и требовалось. □

9. Результант

Рассмотрим многочлены $f(t) = \sum_{i=0}^{n} \alpha_i t^i$ и $g(t) = \sum_{j=0}^{m} \beta_j t^j$, где α_i, β_j лежат в некотором поле F, а $\alpha_n, \beta_m \neq 0$. В терминах коэффициентов α_i, β_j мы хотим сформулировать утверждение о том, что многочлены f и g взаимно просты.

Лемма 9.1. *Многочлены $f, g \in F[t]$ не являются взаимно простыми тогда и только тогда, когда существуют ненулевые многочлены $u, v \in F[t]$ такие, что $uf = vg$, $\deg u < m$, а $\deg v < n$.*

Доказательство. Если $d(t) := \gcd(f, g) \in F[t]$ имеет ненулевую степень, то можно взять $u = g/d, v = f/d \in F[t]$.

Обратно, пусть f и g взаимно просты, а $uf = vg$ для некоторых ненулевых $u, v \in F[t]$. Существуют $p, q \in F[t]$ такие, что $fp + gq = 1$. Тогда $u = ufp + ugq = g(vp + uq)$, откуда $\deg u \geqslant m$. □

Равенство $uf = vg$ можно записать в виде системы линейных уравнений относительно коэффициентов многочленов $u = u_{m-1}t^{m-1} + \cdots + u_0$ и $v = v_{n-1}t^{n-1} + \cdots + v_0$:

$$\begin{gathered}
\alpha_n u_{m-1} = \beta_m v_{n-1} \\
\alpha_{n-1} u_{m-1} + \alpha_n u_{m-2} = \beta_{m-1} v_{n-1} + \beta_m v_{n-2} \\
\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
\alpha_0 u_0 = \beta_0 v_0
\end{gathered}$$

Многочлены f и g имеют необратимый общий делитель тогда и только тогда, когда эта система имеет ненулевое решение, т. е. когда определитель матрицы этой системы равен нулю:

$$\begin{vmatrix} \alpha_n & 0 & \dots & 0 & -\beta_m & 0 & \dots & 0 \\ \alpha_{n-1} & \alpha_n & \dots & 0 & -\beta_{m-1} & -\beta_m & \dots & 0 \\ \alpha_{n-2} & \alpha_{n-1} & \ddots & 0 & -\beta_{m-2} & -\beta_{m-1} & \ddots & 0 \\ \vdots & \vdots & \ddots & \alpha_n & \vdots & \vdots & \ddots & -\beta_m \\ \alpha_0 & \alpha_1 & \dots & \vdots & -\beta_0 & -\beta_1 & \dots & \vdots \\ 0 & \alpha_0 & \ddots & \dots & 0 & -\beta_0 & \dots & \vdots \\ \vdots & \vdots & \ddots & \alpha_1 & \vdots & \dots & \ddots & -\beta_1 \\ 0 & \dots & 0 & \alpha_0 & 0 & \dots & \dots & -\beta_0 \end{vmatrix} = 0$$

Убирая минусы перед всеми β_j и транспонируя, получаем определитель матрицы

$$S(f,g) := \begin{pmatrix} \alpha_n & \dots & \dots & \dots & \alpha_0 & 0 & \dots & 0 \\ 0 & \alpha_n & \dots & \dots & \dots & \alpha_0 & 0 & \dots \\ \dots & \ddots & \ddots & \dots & \dots & \dots & \ddots & \ddots \\ 0 & \dots & 0 & \alpha_n & \dots & \dots & \dots & \alpha_0 \\ \beta_m & \dots & \dots & \dots & \beta_0 & 0 & \dots & 0 \\ 0 & \beta_m & \dots & \dots & \dots & \beta_0 & \dots & 0 \\ \dots & \ddots & \ddots & \dots & \dots & \dots & \ddots & \dots \\ 0 & \dots & 0 & \beta_m & \dots & \dots & \dots & \beta_0 \end{pmatrix}$$

который совпадает с предыдущим определителем с точностью до знака.

Матрицу $S(f,g)$ называют матрицей Сильвестра многочленов f и g. Определитель матрицы $S(f,g)$ называют *результантом* многочленов f и g. Мы будем обозначать его через $r(f,g)$. Если f и g – многочлены от нескольких переменных $t, t_1, \dots, t_n$, а мы хотим вычислить их результант как многочленов от t над кольцом многочленов от остальных переменных, то такой результант будет обозначаться через $r_t(f,g)$. Ясно, что $r(f,g)$ – однородный многочлен степени m по переменным α_i, и степени n по переменным β_j.

Предложение 9.2. *Многочлены f и g имеют необратимый общий делитель в $F[t]$ тогда и только тогда, когда $r(f,g) = 0$.*

Результант имеет много разных приложений. Например, можно исключать переменные из системы полиномиальных уравнений.

Предложение 9.3. *Пусть $f, g \in F[x, y_1, \dots, y_n]$. Рассмотрим f и g, как многочлены от x и вычислим их результант $r = r_x(f,g) \in F[y_1, \dots, y_n]$. Если $f(\lambda, \mu_1, \dots, \mu_n) = g(\lambda, \mu_1, \dots, \mu_n) = 0$ для некоторых $\lambda, \mu_1, \dots, \mu_n \in F$, то $r(\mu_1, \dots, \mu_n) = 0$.*

Доказательство. Многочлены $f(x, \mu_1, \dots, \mu_n)$ и $g(x, \mu_1, \dots, \mu_n) \in F[x]$ имеют общий корень $x = \lambda$. Поэтому $r(\mu_1, \dots, \mu_n) = 0$. □

Несмотря на то, что мотивировка определения результанта требует рассмотрения многочленов над полем, сам результант определен над любым коммутативным кольцом R с 1. Сейчас мы хотим представить результант $r(f,g)$ в виде линейной комбинации $r(f,g) = pf + qg$. Над полем это очевидно: если f, g взаимно просты, то можно представить любой многочлен, а если нет, то $r(f,g) = 0$ и подойдут $p = q = 0$. Но мы хотим написать "естественное" представление, то есть представление, для которого коэффициенты многочленов p и q полиномиально зависят от коэффициентов f и g.

Теорема 9.4. *Пусть R – произвольное коммутативное кольцо с 1, $f = \sum_{i=0}^{n} \alpha_i t^i$ и $g = \sum_{j=0}^{m} \beta_j t^j$ – многочлены над R. Тогда существуют $p, q \in R[t]$ такие, что $r(f,g) = pf + qg$.*

Доказательство. Пусть сначала $\widetilde{R} = \mathbb{Z}[a_0, \dots, a_n, b_0, \dots, b_m]$, $\widetilde{f} = \sum_{i=0}^{n} a_i t^i$, $\widetilde{g} = \sum_{j=0}^{m} b_j t^j$, а F – поле частных кольца $\widetilde{R}[t]$. Положим $z = (t^{m+n-1}, t^{m+n-2}, \dots, 1)^\intercal$ и $w = (t^{m-1}\widetilde{f}(t), t^{m-2}\widetilde{f}(t), \dots, \widetilde{f}(t), t^{n-1}\widetilde{g}(t), t^{n-2}\widetilde{g}(t), \dots, \widetilde{g}(t))^\intercal$. Вычисление показывает, что $S(\widetilde{f}, \widetilde{g})z = w$. Рассматривая последнее равенство как систему линейных уравнений над F, напишем формулу Крамера для последнего элемента столбца z:

$$r(\widetilde{f}, \widetilde{g}) = 1 \cdot \det S(\widetilde{f}, \widetilde{g}) = \det(s_1, \dots, s_{m+n-1}, w),$$

где s_i – столбцы матрицы $S(\widetilde{f}, \widetilde{g})$. Раскладывая определелитель в правой части по последнему столбцу получим $r(\widetilde{f}, \widetilde{g}) = \widetilde{p}\widetilde{f} + \widetilde{q}\widetilde{g}$ для некоторых $\widetilde{p}, \widetilde{q} \in \widetilde{R}[t]$.

Подставляя α_i вместо всех a_i и β_j вместо всех b_j получаем требуемое равенство.[4] □

Для того чтобы вывести следующие свойства результанта, заметим, что деление с остатком многочленов от одной переменной возможно над любым кольцом при условии, что старший коэффициент делителя обратим. Следовательно, и теорема Безу верна не только над полем, но и над любым кольцом.

Лемма 9.5. *Пусть R – кольцо, $h \in R[t]$, а $a \in R$. Если $h(a) = 0$, то h делится на $t - a$ в $R[t]$.*

Теорема 9.6. *Пусть R – кольцо. Предположим, что многочлены $f, g \in R[t]$ раскладываются на линейные множители:*

$$f(t) = \alpha_n \prod_{i=1}^{n}(t - \xi_i) \text{ и } g(t) = \beta_m \prod_{j=1}^{m}(t - \omega_j),$$

где $\alpha_n, \beta_m, \xi_i, \omega_j \in R$. Тогда

$$r := r(f, g) = \alpha_n^m \beta_m^n \prod_{i,j}(\xi_i - \omega_j) = \alpha_n^m \prod_{i=1}^{n} g(\xi_i) = (-1)^{mn} \beta_m^n \prod_{j=1}^{m} f(\omega_j).$$

Доказательство. Заметим, что второе и третье равенства выполнены по определению многочленов f и g. Требуемое равенство полиномиально относительно $\alpha_n, \beta_m, \xi_1, \dots, \xi_n, \omega_1, \dots, \omega_m$, поэтому его можно будет спроектировать из “generic case” в любое кольцо. В “generic case” давайте считать, что x_i и y_j – независимые переменные. Положим

$$\widetilde{f}(t) = \sum_{i=0}^{n} a_i t^i = a_n \prod_{i=1}^{n}(t - x_i) \text{ и } \widetilde{g}(t) = \sum_{j=0}^{m} b_j t^j = b_m \prod_{j=1}^{m}(t - y_i).$$

Таким образом, в первой части доказательства:

- $X = (x_1, \dots, x_n)$, $Y = (y_1, \dots, y_m)$,
- $\widetilde{R} = \mathbb{Z}[a_n, b_m, X, Y]$ – кольцо многочленов,
- $\widetilde{f}, \widetilde{g} \in \widetilde{R}[t]$,
- $a_0, \dots, a_n, b_0, \dots, b_m \in \widetilde{R}$,

и мы хотим доказать равенство из условия теремы с заменой соответствующих греческих букв на соответствующие латинские, а R на $\widetilde{R}$.

[4] Формально, мы рассматриваем гомоморфизм $\mathbb{Z}[t]$-алгебр $\varepsilon : \widetilde{R}[t] \to R[t]$, посылающий a_i в α_i, а b_j в β_j, который существует и единственный по универсальному свойству кольца многочленов, и применяем его к равенству $r(\widetilde{f}, \widetilde{g}) = \widetilde{p}\widetilde{f} + \widetilde{q}\widetilde{g}$.

Каждая строка матрицы Сильвестра $S(\widetilde{f},\widetilde{g})$ делится либо на a_n, либо на b_m. Вынося эти общие множители за скобку, получим $r(\widetilde{f},\widetilde{g}) = a_n^m b_m^n \cdot \widetilde{r}$, где

$$\widetilde{r} = \begin{vmatrix} 1 & -s_1(X) & \dots & \dots & (-1)^n s_n(X) & 0 & \dots & 0 \\ 0 & 1 & \dots & \dots & \dots & (-1)^n s_n(X) & 0 & \dots \\ \dots & \ddots & \ddots & \dots & \dots & \dots & \ddots & \ddots \\ 0 & \dots & 0 & 1 & \dots & \dots & \dots & (-1)^n s_n(X) \\ 1 & \dots & \dots & \dots & (-1)^m s_m(Y) & 0 & \dots & 0 \\ 0 & 1 & -s_1(Y) & \dots & \dots & (-1)^m s_m(Y) & \dots & 0 \\ \dots & \ddots & \ddots & \dots & \dots & \dots & \ddots & \dots \\ 0 & \dots & 0 & 1 & \dots & \dots & \dots & (-1)^m s_m(Y) \end{vmatrix},$$

$X = (x_1, \dots, x_n)$, а $Y = (y_1, \dots, y_m)$. Таким образом, нам надо доказать, что

$$\widetilde{r} = p := \prod_{i,j} (x_i - y_j).$$

Другими словами, можно считать, что $a_n = b_m = 1$.

Если в многочлен $\widetilde{f}$ вместо одного из x_i подставить y_j, полученный многочлен будет иметь общий множитель с $\widetilde{g}$. По предложению 9.2 $\widetilde{r}|_{x_i=y_j} = 0$. По предыдущей лемме $\widetilde{r}$ делится на $x_i - y_j$. Так как кольцо $\widetilde{R}$ факториально, а элементы $x_i - y_j$ неприводимы, то $\widetilde{r}$ делится на p. Рассмотрим $\widetilde{r}$ и p, как многочлены от Y над кольцом $\mathbb{Z}[X]$. Ясно, что $\deg_Y \widetilde{r} = \deg_Y p = mn$. Одночлен старшей степени по Y в p равен $(-1)^{mn} y_1^n \dots y_m^n$, а в $\widetilde{r}$ он получается из произведения диагональных элементов матрицы и равен $\big((-1)^m s_m(Y)\big)^n = (-1)^{mn} s_m(Y)^n$. Так как старшие одночлены равны, а $\widetilde{r}$ делится на p, то $\widetilde{r} = p$.

По универсальному свойству кольца многочленов существует (единственный) гомоморфизм $\varphi : \widetilde{R} \to R$ такой, что $\varphi(x_i) = \xi_i$, $\varphi(y_i) = \omega_i$, $\varphi(a_n) = \alpha_n$, $\varphi(b_m) = \beta_m$. Так как φ сохраняет операции сложения и умножения, то доказанное равенство переходит под действием φ в равенство из условия теоремы. □

Следствие 9.7. *Если* $f = gh + q$, *то* $r(g,f) = b_m^{n-\deg q} r(g,q)$, *где* b_m – *старший коэффициент многочлена* g.

Доказательство. Требуемое равенство полиномиально зависит от коэффициентов многочленов g, h и q. Поэтому, также как и в предыдущей теореме, его можно спроектировать из соответствующего утверждения над кольцом многочленов над $\mathbb{Z}$. Это кольцо многочленов является областью целостности и, потому, лежит в некотором алгебраически замкнутом поле F (скажем, в замыкании своего поля частных). Следовательно, достаточно доказать равенство для многочленов из $F[t]$.

Пусть y_i – корни многочлена g в поле F. Тогда $q(y_i) = f(y_i)$. Остается воспользоваться формулой из теоремы 9.6. □

Следствие 9.8. $r(f,gh) = r(f,g)r(f,h)$.

Доказательство. Также как и в предыдущем доказательстве, достаточно рассмотреть случай, когда $f, g, h \in F[t]$, а F – алгебраически замкнутое поле. Пусть ξ_i корни многочлена f в F, а α_n – его старший коэффициент. Тогда

$$r(f,gh) = \alpha_n^{\deg g + \deg h} \prod_{i=1}^n g(\xi_i)h(\xi_i) = \alpha_n^{\deg g} \prod_{i=1}^n g(\xi_i) \cdot \alpha_n^{\deg h} \prod_{i=1}^n h(\xi_i) = r(f,g)r(f,h).$$

□

10. Дискриминант

Пусть $x_1, \dots, x_n$ – корни многочлена $f(t) = a_n t^n + \dots + a_0$, причем $a_n \neq 0$.[5] Величину

$$d(f) = a_n^{2n-2} \prod_{i<j} (x_i - x_j)^2$$

называют *дискриминантом* многочлена f. Смысл дискриминанта над замкнутым полем состоит в том, что он равен нулю тогда и только тогда, когда многочлен имеет кратные корни. Над $\mathbb{R}$ с помощью знака дискриминанта можно сформулировать необходимое условие отсутствия вещественных корней. Очевидно, что произведение в определении дискриминанта является симметрическим многочленом от корней многочлена f, поэтому оно является многочленом (назовем его g) от a_i/a_n. Степень g равна степени произведения по x_1, см. следствие 7.3, которая равна $2(n-1)$. Таким образом,

$$d(f) = a_n^{2n-2} \prod_{i<j} (x_i - x_j)^2 = a_n^{2n-2} g(a_0/a_n, \dots, a_{n-1}/a_n),$$

является многочленом от $a_0, \dots, a_n$.

Теорема 10.1. *Пусть* $\deg f = n$, *а* $\deg f' = m$. *Тогда* $r(f, f') = (-1)^{n(n-1)/2} a_n^{m+2-n} d(f)$, *где* a_n *– старший коэффициент многочлена* f *(если* a_n *не обратимо в базовом кольце, а* $m + 2 < n$, *то надо формально домножить равенство на* a_n^{n-m-2}*).*

В частности, если R *– область целостности, характеристика которой не делит* n, *то* $r(f, f') = (-1)^{n(n-1)/2} a_n d(f)$.

Доказательство. По теореме 9.6 $r(f, f') = a_n^m \prod_i f'(x_i)$, где $x_1, \dots, x_n$ – корни многочлена f. Легко проверить, что $f'(x_i) = a_n \prod_{j \neq i} (x_i - x_j)$. Поэтому

$$r(f, f') = a_n^{m+n} \prod_{i \neq j} (x_i - x_j) = a_n^{m+n} (-1)^{n(n-1)/2} \prod_{i<j} (x_i - x_j)^2 = (-1)^{n(n-1)/2} a_n^{m+2-n} d(f).$$

□

Из теоремы снова следует что дискриминант является многочленом с целыми коэффициентами от коэффициентов многочлена f (даже в случае $m = n - 1$ можно сократить на a_n, которое является общим множителем первого столбца матрицы Сильвестра).

Теорема 10.2. *Пусть* f, g, h *– многочлены со старшим коэффициентом 1. Тогда*

$$d(fg) = d(f) d(g) r(f, g)^2,$$
$$d(fgh) = d(f) d(g) d(h) r(f, g)^2 r(g, h)^2 r(h, f)^2.$$

Доказательство. Пусть $x_1, \dots, x_n$ – корни многочлена f, а $y_1, \dots, y_m$ – корни многочлена g. Тогда

$$d(fg) = \prod_{i<j} (x_i - x_j)^2 \cdot \prod_{i<j} (y_i - y_j)^2 \cdot \prod_{i,j} (x_i - y_j)^2 = d(f) d(g) r(f, g)^2.$$

Вторая формула доказывается аналогично. □

Теорема 10.3. *Пусть* f *– вещественный многочлен степени* n, *не имеющий кратных корней. Если* f *не имеет вещественных корней (заметим, что в этом случае* n *четное), то* $\operatorname{sign} d(f) = (-1)^{n/2}$.

[5] Мы используем латинские буквы для обозначения коэффициентов и корней, а также не уточняем кольцо коэффициентов, потому что на самом деле все происходит над универсальным кольцом $\mathbb{Z}[a_n, x_1, \dots, x_n]$, из которого результаты могут быть спроектированы в кольцо многочленов с любым кольцом коэффициентов.

Доказательство. Так как старший коэффициент входит в выражение для дискриминанта в четной степени, можно считать, что он равен 1. Пусть a и $\bar{a}$ – пара сопряженных корней многочлена f, т. е. $f(t) = (t-a)(t-\bar{a})g(t)$. Тогда по предыдущей теореме

$$d(f) = d(g)r(t-a, t-\bar{a})^2 r(t-\bar{a}, g)^2 r(g, t-a)^2 = \\ d(g)(\bar{a}-a)^2 g(a)^2 g(\bar{a})^2 = -d(g)(\operatorname{Im} a)^2 |g(a)|^4.$$

Следовательно, $\operatorname{sign} d(f) = -\operatorname{sign} d(g)$, и доказательство заканчивается индукцией по n. □

Упражнение 10.4. Если многочлен, не имеющий кратных корней, имеет ровно m пар комплексно сопряженных корней, то $\operatorname{sign} d(f) = (-1)^m$.

11. Конечнопорожденные алгебры над полем

Наша следующая цель – построить взаимно однозначное соответствие между радикальными идеалами кольца многочленов $F[t_1, \ldots, t_n]$ над алгебраически замкнутым полем F и алгебраическими множествами в аффинном пространстве, что является основой классической алгебраической геометрии. Для начала нам потребуется несколько сведений о целых расширениях колец и полей.

Пусть $A \subseteq B$ – пара колец. Элемент $\beta \in B$ называется целым над A, если он является корнем унитального многочлена с коэффициентами из A (в частности, целые алгебраические числа – это комплексные числа целые над $\mathbb{Z}$). Кольцо B называется целым над A (или целым расширением A), если все элементы являются целыми над A. Целое расширение поля называется алгебраическим расширением. Если B – конечнопорожденный A-модуль, то B называется конечным расширением A. Легко видеть, что элемент $\beta \in B$ является целым над A тогда и только тогда, когда $A[\beta]$ является конечным расширением A. Любое конечное расширение является целым, однако для произвольных коммутативных колец доказательство этого утверждения не просто. Мы докажем его только для нетеровых колец.

Лемма 11.1. *Пусть A – нетерово кольцо. Конечное расширение B кольца A является целым.*

Доказательство. Пусть $\beta \in B$. Если B является конечнопорожденным A-модулем, то по предложению 11.1 главы 4 $A[\beta]$ также является конечно порожденным A-модулем. Если он порожден элементами $p_1(\beta), \ldots, p_m(\beta)$ для некоторых $p_1, \ldots, p_m \in A[t]$, а $d > \max(\deg p_i)$, то существуют элементы $a_1, \ldots, a_m \in A$ такие, что $\beta^d = \sum_i a_i p_i$, что и означает, что β является корнем унитального многочлена. □

Лемма 11.2. *Последовательность конечных расширений является конечным расширением.*

Доказательство. Пусть $A \subseteq B \subseteq C$ – кольца, причем $\beta_1, \ldots, \beta_n$ порождают B, как A-модуль, а $\gamma_1, \ldots, \gamma_m$ порождают C, как B-модуль. Тогда любой элемент $\gamma \in C$ представляется в виде $\gamma = \sum\limits_i \gamma_i \delta_i = \sum\limits_{i,j} \gamma_i \beta_j \varepsilon_{ij}$, где $\delta_i \in B$, а $\varepsilon_{ij} \in A$. Таким образом, C порождено элементами $\gamma_i \beta_j$, $i = 1, \ldots, n$, $j = 1, \ldots, m$ как A-модуль. Для более длинных последовательностей конечных расширений работает индукция по длине последовательности. □

Лемма 11.3. *Пусть B является целым над A. Если B – поле, то и A – поле.*

Доказательство. Пусть $\alpha \in A \smallsetminus \{0\}$. Так как B – поле, то $1/\alpha \in B$ является целым над A. Следовательно, $\frac{1}{\alpha^n} + \frac{\gamma_{n-1}}{\alpha^{n-1}} + \cdots + \gamma_0 = 0$ для некоторых $\gamma_i \in A$. Умножая на α^n, получаем $1 + \gamma_{n-1}\alpha + \cdots + \gamma_0 \alpha^n = 0$, откуда $1/\alpha = -\gamma_{n-1} - \cdots - \gamma_0 \alpha^{n-1} \in A$. □

В оставшейся части параграфа речь идет о кольце многочленов от нескольких переменных. Мы продолжаем использовать обозначение $T = (t_1, \ldots, t_n)$.

Лемма 11.4. *Пусть $f \in F[T]$. Существует автоморфизм φ F-алгебры $F[T]$ такой, что*

$$\varphi(f) = \alpha t_n^l + g_{l-1}t_n^{l-1} + \cdots + g_0, \text{ где } \alpha \in F^*,\ g_0, \ldots, g_{l-1} \in F[t_1, \ldots, t_{n-1}].$$

Доказательство.[6] Пусть $f = \sum\limits_{a \in X} \beta_a T^a$, где X – конечное подмножество в $\mathbb{N}_0^n$, а $\beta_a \in F^*$. Выберем $k \in \mathbb{N}$ большее максимума из a_i по всем $a \in X$ и всем $i = 1, \ldots, n$. Положим

$$\varphi(t_i) = t_i + t_n^{k^{n-i}} \text{ при } 1 \leqslant i \leqslant n-1, \quad \text{а } \varphi(t_n) = t_n.$$

Образ одночлена T^a под действием φ равен $t_n^{a_n} \prod_{i=1}^{n-1}(t_i + t_n^{k^{n-i}})^{a_i}$, а его старший член, как многочлена от t_n равен

$$LM_{t_n}\big(\varphi(T^a)\big) = t_n^{a_n} \prod_{i=1}^{n-1} t_n^{k^{n-i}a_i} = t_n^{m}(a), \text{ где } m(a) = \sum_{i=1}^{n} k^{n-i}a_i.$$

Так как $a_i < k$ для всех ненулевых мономов многочлена f, то для них $m(a)$ является k-ичной записью числа $a_n a_{n-1} \ldots a_0$. Таким образом, если $a \neq b \in X$, то $m(a) \neq m(b)$. Следовательно, существует $c \in X$ такое, что $m(c) > m(a)$ для любого $a \in X \setminus \{c\}$.

Ясно, что $LM_{t_n}\big(\varphi(f)\big) = LM_{t_n}\big(\varphi(\beta_c T^c)\big) = \beta_c t_n^{m(c)}$, т. е. коэффициент при старшей степени t_n является ненулевой константой. □

Теорема 11.5 (лемма Нетер о нормализации)**.** *Любая конечнопорожденная алгебра над полем является конечным расширением кольца многочленов.*

Доказательство. Любая конечнопорожденная алгебра над полем F изоморфна $A = F[t_1, \ldots, t_n]/I$ для некоторого идеала I. Проведем индукцию по n. При $n = 0$ (в этой теореме не исключается случай кольца многочленов от нуля переменных) наша алгебра равна F и доказывать нечего.

Пусть $n > 0$. Случай $I = 0$ очевиден. Если же $I \neq 0$, то, воспользовавшись леммой 11.4 и заменив I на $\varphi(I)$, можно считать, что I содержит многочлен f, унитальный по переменной t_n. Пусть $\bar{\cdot} : F[t_1, \ldots, t_n] \to A$ – гомоморфизм редукции. Положим $R = \overline{F[t_1, \ldots, t_{n-1}]}$, а $x = \overline{t_n}$. Ясно, что $R \cong F[t_1, \ldots, t_{n-1}]/J$ для некоторого идеала J, а $A = R[x]$. Очевидно также, что $\overline{g}(x) = 0$, где g – это f, рассмотренный как многочлен от t_n с коэффициентами из $F[t_1, \ldots, t_{n-1}]$. Так как g – унитальный многочлен, то x – целый над R и $A = R[x]$ является конечным расширением R. По индукционному предположению R является конечным расширением кольца многочленов. Наконец по лемме 11.2 исходная алгебра также является конечным расширением того же кольца многочленов, что и R. □

Теорема 11.6 (лемма Зарискoго)**.** *Пусть A – конечнопорожденная алгебра над полем F, а I – ее максимальный идеал. Тогда $K = A/I$ – конечное расширение поля F. В частности, если A – поле, то $K = A$ – конечное расширение F.*

Доказательство. По лемме Нетер K является конечным расширением кольца многочленов $F[t_1, \ldots, t_d]$, а так как I – максимален, то K – поле. По лемме 11.3 $F[t_1, \ldots, t_d]$ также является полем, т. е. $d = 0$. □

12. Теорема Гильберта о нулях

Рассмотрим идеал кольца $F[T]$ порожденный элементами $t_1 - \alpha_1, \ldots, t_n - \alpha_n$. По теореме о делении с остатком 4.1 для любого $f \in F[T]$ существуют $q_1, \ldots, q_n, r \in F[T]$ такие, что $f = (t_1 - \alpha_1)q_1 + \cdots + (t_n - \alpha_n)q_n + r$, причем ни один из членов полинома r не делится ни на один из t_i. Это означает, что $r \in F$. Подставляя $t_i = \alpha_i$ получим $r = f(\alpha_1, \ldots, \alpha_n)$ (теорема Безу для многих переменных). Из этого сразу вытекает следующий факт.

[6] Спасибо Федору Ушакову за упрощение этого доказательства.

Лемма 12.1. *Идеал кольца $F[T]$ порожденный элементами $t_1-\alpha_1,\dots,t_n-\alpha_n$ является максимальным.*

Доказательство. Обозначим этот идеал через M. Любой многочлен из M имеет корень $(\alpha_1,\dots,\alpha_n)$, поэтому $M\neq F[T]$. С другой стороны, если $f\notin M$, то $f-(t_1-\alpha_1)q_1-\dots-(t_n-\alpha_n)q_n=f(\alpha_1,\dots,\alpha_n)\in F^*$, следовательно, $M+fF[t]=F[T]$. □

Единичным идеалом кольца R называется идеал, содержащий 1, т. е. само кольцо, как идеал над собой. Смысл введения этого термина: выражение "порождает единичный идеал" уже уточняет, что кольцо порождается как идеал, а не как алгебра.

Сейчас мы сформулируем в одном утверждении все вариации слабой теоремы Гильберта о нулях.

Теорема 12.2 (Слабая теорема Гильберта о нулях)**.** *Пусть F – поле. Следующие условия эквивалентны.*

(1) *F алгебраически замкнуто.*
(2) *$F[t_1,\dots,t_n]/I\cong F$ для любого максимального идеала I кольца многочленов $F[t_1,\dots,t_n]$.*
(3) *Любой максимальный идеал кольца $F[t_1,\dots,t_n]$ порожден элементами $t_1-\alpha_1,\dots,t_n-\alpha_n$ для некоторых $\alpha_1,\dots,\alpha_n\in F$.*
(4) *Набор многочленов $P\subseteq F[t_1,\dots,t_n]$ не имеет общих корней тогда и только тогда, когда P порождает единичный идеал.*

Доказательство. Вспоминаем обозначение $T=(t_1,\dots,t_n)$.

(1) $\Longrightarrow$ (2). По теореме 11.6 $F[T]/I$ является конечным, а значит и целым расширением поля F. Но замкнутое поле не имеет нетривиальных целых расширений, потому что любой многочлен из $F[T]$ раскладывается на линейные множители уже в $F[T]$.

(2) $\Longrightarrow$ (3). Пусть $\rho: F[T]\to F$ – гомоморфизм F-алгебр с ядром I. Положим $\alpha_i=\rho(t_i)$. Тогда $t_i-\alpha_i\in I$ для любого $i=1,\dots,n$. Следовательно, I содержит идеал M, порожденный всеми $t_i-\alpha_i$, который по предыдущей лемме является максимальным. Поэтому $I=M$.

(3) $\Longrightarrow$ (4). Если P имеет общий корень $\alpha\in F^n$, то $\sum_{p\in P}f_p(\alpha)p(\alpha)=0$ для любых $f_p\in F[t_1,\dots,t_n]$. Поэтому равенство $\sum_{p\in P}f_pp=1$ невозможно (эта часть доказательства не зависит от других условий).

Обратно, если P не порождает единичный идеал, то оно содержится в некотором максимальном идеале $I\leqslant F[t_1,\dots,t_n]$. По пункту (3) многочлены из I имеют общий корень $\alpha\in F^n$, следовательно, и $p(\alpha)=0$ для любого $p\in P$.

(4) $\Longrightarrow$ (1). Если F не замкнуто, то существует многочлен $p\in F[t_1]$, не имеющий корней. Ясно, что этот многочлен не порождает единичный идеал в кольце $F[T]$. □

Наконец, докажем полную версию теоремы Гильберта о нулях, которая по другому называется Nullstellensatz. Для подмножества $P\subseteq F[T]$ обозначим через $V(P)$ множество общих нулей всех многочленов из P.

Теорема 12.3 (Теорема Гильберта о нулях)**.** *Пусть F – алгебраически замкнутое поле, $P\subseteq F[t_1,\dots,t_n]$, а $f\in F[t_1,\dots,t_n]$. Для того чтобы $f\big(V(P)\big)=0$ необходимо и достаточно, чтобы многочлен f^d принадлежал идеалу, порожденному P, для некоторого натурального d.*

Доказательство. Если f^d принадлежит идеалу, порожденному P, то $f^d\big(V(P)\big)=0$, откуда $f\big(V(P)\big)=0$. Обратную импликацию выведем из слабой теоремы о нулях при помощи *трюка Рабиновича*.

В кольце многочленов $F[t_0,\dots,t_n]$ рассмотрим подмножество $P\cup\{1+t_0f\}$. Для $a\in F^n$, если $P(a)=0$, то $1+t_0f(a)=1$, поэтому $V(P\cup\{1+t_0f\})=\varnothing$. По слабой теореме о нулях множество $P\cup\{1+t_0f\}$ порождает единичный идеал, т. е. существуют $g,g_1,\dots,g_m\in F[t_0,\dots,t_n]$

и $p_1, \ldots, p_m \in P$ такие, что

$$(1 + t_0 f)g + \sum_{k=1}^{m} g_k p_k = 1.$$

Существует единственный гомоморфизм $F[t_1, \ldots, t_n]$-алгебр из $F[t_0, \ldots, t_n]$ в главную локализацию $F[t_1, \ldots, t_n]_f$, посылающий t_0 в $-1/f$. Применяя его к выделенной формуле, получаем $\sum h_k p_k = 1$, где $h_k = g_k(-1/f, t_1, \ldots, t_n)$ – образ g_k при указанном гомоморфизме. Любой элемент локализации $F[t_1, \ldots, t_n]_f$ записывается в виде q/f^d для $q \in F[t_1, \ldots, t_n]$ и $d \in \mathbb{N}_0$. Записав все h_k в виде $h_k = q_k/f^d$ для $q_k \in F[t_1, \ldots, t_n]$ и достаточно большого d и домножив на знаменатель, получим $f^d = \sum q_k p_k$, что и требовалось. □

13. Начала алгебраической геометрии

Глава 12

Теория представлений

И снова теория групп, но теперь немного в другом разрезе. Если в главе 8 мы изучали действия групп на множествах, что соответствует гомоморфизму группы в симметрическую группу, то теперь мы рассмотрим *линейное* действие групп на векторных пространствах, что будет соответствовать гомоморфизмам в полную линейную группу. Методы и результаты теории представлений используются не только в самой теории групп, но, например, и в теоретической физике.

Так как наш курс ознакомительный, то мы будем изучать только *конечномерные* представления *конечных* групп над *алгебраически замкнутым* полем *характеристики* 0. Однако, все определения и утверждения, которые не требуют этого предположения будут сформулированы в общем случае.

В отличие от предыдущих глав, в этой главе кольца не предполагаются по умолчанию коммутативными.

1. Основные определения

Пусть V – векторное пространство над полем F, а G – группа. Напомним, что $\mathrm{GL}(V) = \mathrm{Aut}(V)$ – это группа всех обратимых линейных операторов на V.

Определение 1.1. Линейным представлением G на V называется гомоморфизм $\pi : G \to \mathrm{GL}(V)$. Размерность пространства V называется степенью представления и обоначается $\deg \pi$.

Базис пространства V задает изоморфизм $\mathrm{GL}(V) \cong \mathrm{GL}_n(F)$, где $n = \dim V$: оператору сопоставляется его матрица в выбранном базисе. Гомоморфизм $\pi : G \to \mathrm{GL}_n(F)$ называется матричным представлением группы G. Говоря про линейное или матричное представление π обычно пишут π_g вместо $\pi(g)$, потому что выражение $\pi_g(v)$ выглядит приятнее для глаза, чем $\pi(g)(v)$. Матричное представление задает n^2 функций $\pi_{ij} : G \to F$ по правилу $\pi_{ij}(g) = (\pi_g)_{ij}$. Говорят что G действует линейно на V, если задано действие $G \curvearrowright V$, $(g, v) \mapsto gv$, удовлетворяющее равенствам

$$g(u + v) = gu + gv \text{ и } g(\alpha v) = \alpha g v \ \forall g \in G,\ v \in V,\ \alpha \in F.$$

Линейное представление определяет линейное действие $G \curvearrowright V$ по правилу $gv = \pi_g(v)$. Линейное действие G на V очевидным образом продолжается до действия групповой алгебры FG на V:

$$(\sum_{g \in G} \alpha_g g)v := \sum_{g \in G} \alpha_g (gv),\ \text{где } \alpha_g \in F,\ v \in V.$$

Таким образом, V превращается в левый FG-модуль. Аналогично, говоря про линейное или матричное представление $\pi : G \to \mathrm{GL}(V)$, мы продолжим его по линейности до гомоморфизма алгебр $FG \to \mathrm{End}(V)$, которое будем обозначать той же буквой π.

Мы связали с линейным представлением группы еще 3 объекта. Класс объектов каждого типа естественным образом превращается в категорию так, чтобы четыре получившиеся категории были эквивалентны. Определим морфизмы в каждой из категорий.

Определение 1.2. Морфизмом представлений $\pi : G \to \mathrm{GL}(U)$ в $\rho : G \to \mathrm{GL}(V)$ называется линейное отображение $\varphi : U \to V$ удовлетворяющее условию $\varphi\big(\pi_g(u)\big) = \rho_g\big(\varphi(u)\big)$

для любых $g \in G$ и $u \in U$. Морфизм представлений называется еще гомоморфизмом представлений или сплетающим отображением.

Определение 1.3. Морфизмом матричных представлений $\pi : G \to \mathrm{GL}_n(F)$ в $\rho : G \to \mathrm{GL}_m(F)$ называется матрица $A \in \mathrm{M}_{m,n}(F)$ такая, что $A\pi_g = \rho_g A$ при всех $g \in G$.

Определение 1.4. Морфизмом линейных действий $G \curvearrowright U$ в $G \curvearrowright V$ называется G-эквивариантное линейное отображение $\varphi : U \to V$ (напомним, что эквивариантность означает: $\varphi(gu) = g\varphi(u)$ для любых $g \in G$ и $u \in U$).

Таким образом, мы определили категории линейных представлений, матричных представлений и линейных действий данной группы G, а что такое категория левых FG-модулей мы уже знаем. Доказательство следующего утверждения является рутинным.

Предложение 1.5. *Категории линейных представлений, матричных представлений и линейных действий данной группы G, а также категория левых FG-модулей эквивалентны между собой (для категории матричных представлений необходима аксиома выбора для классов).*

Сформулируем теперь несколько определений на языке линейных представлений и модулей. Пусть $\pi : G \to \mathrm{GL}(V)$ – линейное представление группы G на векторном пространстве V над полем F.

- Два представления называются эквивалентными, если они изоморфны в категории линейных представлений. Другими словами, представления $\pi : G \to \mathrm{GL}(V)$ и $\pi' : G \to \mathrm{GL}(V')$ эквивалентны, если существует изоморфизм $\varphi : V \to V'$ такой, что $\varphi\pi_g = \pi'_g\varphi$ для любого $g \in G$. Последнее равенство можно переписать в виде $\pi_g = \varphi^{-1}\pi'_g\varphi$. Ясно, что из эквивалентности представлений следует $\dim V = \dim V'$. В случае матричных представлений $V = V' = F^n$. Матричные представления $\pi : G \to \mathrm{GL}_n(F)$ и $\pi' : G \to \mathrm{GL}_m(F)$ эквивалентны тогда и только тогда, когда $m = n$ и существует $C \in \mathrm{GL}_n(F)$ такая, что $\pi_g = C^{-1}\pi'_g C$. Таким образом, если C интерпретировать как матрицу замены базиса, то получится, что эквивалентные матричные представления отличаются друг от друга только выбором базиса в F^n. Модули эквивалентных представлений называются изоморфными.
- Подпространство $U \leqslant V$ называется G-инвариантным, если оно инвариантно относительно всех операторов π_g, $g \in G$. Ясно, что в этом случае U является подмодулем FG-модуля V.
- Если U – G-инвариантное подпространство в V, то индуцированные гомоморфизмы $G \to \mathrm{GL}(U)$ и $G \to \mathrm{GL}(V/U)$ называются подпреставлением и факторпредставлением соответственно. На языке модулей этому соответствует подмодуль U и фактормодуль V/U FG-модуля V. На матричном языке G-инвариантному подпространству соответствует представление π', эквивалентное π, у которого все матрицы π'_g имеют клеточно треугольный вид $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$ с диагональными клетками размера $\dim U$ и $\dim V - \dim U$. Тогда клетка в верхнем левом углу является подпредставлением, а в нижнем правом – факторпредставлением.
- Представление называется неразложимым, если V не раскладывается в прямую сумму двух ненулевых G-инвариантных подпространств. В этом случае FG-модуль V также называется неразложимым. На матричном языке это означает, что не существует представления, π' эквивалентного π такого, что все π'_g имеют клеточно диагональный вид $\left(\begin{smallmatrix} * & 0 \\ 0 & * \end{smallmatrix}\right)$.
- Представление называется неприводимым, если в V ровно 2 G-инвариантных подпространства: $\{0\} \neq V$. В этом случае FG-модуль V называется простым.

- Представление называется вполне приводимым, если V раскладывается в прямую сумму конечного числа неприводимых G-инвариантных подпространств. В этом случае FG-модуль V называется полупростым.
- Представление $\pi : G \to \mathrm{GL}(V)$ называется точным, если π инъективно. Из этого *не следует*, что FG-модуль модуль V точен (R-модуль M называется точным, если $rM \neq 0$ для любого $r \in R \setminus \{0\}$).

2. Строение артиновых колец

Артиновы кольца. Сначала мы изучим представления на языке модулей. Для этого мы классифицируем все артиновы кольца, в частности, групповые кольца конечных групп.

Определение 2.1. Кольцо R называется *артиновым слева*, если любое линейно упорядоченное (по включению) множество левых идеалов содержит наименьший элемент.

Это условие равносильно условию обрыва убывающих цепей левых идеалов или, коротко, DCC (descending chain condition). Так как идеал групповой алгебры в первую очередь является подпространством, то групповая алгебра конечной группы над полем всегда является артиновой (и слева, и справа) из соображений размерности.

Любое факторкольцо артинова кольца артиново, потому что прообраз под действием эпиморфизма цепочки строго вложенных идеалов является цепочкой строго вложенных идеалов. Однако, это не верно для подколец: $\mathbb{Z}$ не артиново подкольцо артинова кольца $\mathbb{Q}$.

Радикал Джекобсона. Для того чтобы сформулировать определение радикала Джекобсона, докажем следующее утверждение. Кольцо R, как левый (правый) модуль над собой, называется *регулярным левым* (*правым*) R*-модулем*.

Лемма 2.2. *Пусть r – элемент произвольного (ассоциативного с 1) кольца R. Следующие условия эквивалентны.*

1L. r содержится в любом максимальном левом идеале кольца R.
1R. r содержится в любом максимальном правом идеале кольца R.
2L. r можно исключить из любой системы образующих регулярного левого R-модуля.
2R. r можно исключить из любой системы образующих регулярного правого R-модуля.
3L. Для любого $x \in R$ элемент $1 + xr$ обратим слева.
3R. Для любого $x \in R$ элемент $1 + rx$ обратим справа.
4L. Для любого $x \in R$ элемент $1 + xr$ двусторонне обратим.
4R. Для любого $x \in R$ элемент $1 + rx$ двусторонне обратим.

Определение 2.3. Множество элементов, удовлетворяющих условиям предыдущей леммы, называется радикалом Джекобсона кольца.

Радикал Джекобсона кольца R обозначается через $\operatorname{Rad} R$ или $\operatorname{JRad} R$. Он равен пересечению максимальных левых (правых) идеалов, и, следовательно, является двусторонним идеалом. Если $\operatorname{Rad} R = \{0\}$, то кольцо R называется полупростым.

Предложение 2.4. *$R/\operatorname{Rad} R$ – полупростое кольцо.*

Идеал $I \lhd R$ называется нильпотентным, если существует натуральное n такое, что $I^n = 0$. По определению степени идеала это означает, что $r_1 \cdot \ldots \cdot r_n = 0$ для любых $r_1 \ldots r_n \in I$. Из пункта (3L) леммы 2.2 вытекает следующее утверждение.

Предложение 2.5. *Любой нильпотентный идеал содержится в радикале Джекобсона.*

Доказательство. Для любого r, принадлежащего нильпотентному левому идеалу, и любого $x \in R$ выполнено равенство $(xr)^n = 0$ при каком-то $n \in \mathbb{N}$. Тогда $(1+xr)(1-xr+(xr)^2-\cdots\pm(xr)^{n-1}) = 1-(xr)^n = 1$. Следовательно, $1+xr$ обратим для любого $x \in R$, откуда $r \in \operatorname{Rad} R$. □

В артиновом кольце верно и обратное утверждение.

Предложение 2.6. *В артиновом (слева) кольце существует наибольший двусторонний нильпотентный идеал, фактор по которому не содержит ненулевых нильпотентных идеалов. Этот идеал совпадает с радикалом Джекобсона.*

Артиново (слева) кольцо с нулевым радикалом Джекобсона называется *классически полупростым*.

Теорема 2.7 (Веддербарна–Артина)**.** *Классически полупростое кольцо изоморфно прямой сумме конечного числа полных матричных колец над телами.*

Как следует из теоремы Веддербарна–Артина, артиново слева кольцо с нулевым радикалом Джекобсона является также артиновым справа. Однако в общем случае это не так. Простейший пример: подкольцо в $\mathrm{M}_2(\mathbb{R})$ состоящее из матриц A, у которых $a_{21} = 0$, a_{11} и a_{12} любые, а $a_{22} \in \mathbb{Q}$. В этом кольце конечное число левых идеалов, в то время как множество $\{\left(\begin{smallmatrix} 0 & a \\ 0 & 0 \end{smallmatrix}\right) \mid a \in A\}$ является правым идеалом для любого $\mathbb{Q}$-подпространства в $\mathbb{R}$. Конечно, такие подпространства не удовлетворяют никаким условиям обрыва.

Естественно, в теории представлений *конечных* групп все групповые кольца артиновы с любой стороны. Так что этот пример не имеет отношения к содержанию настоящей главы.

Казалось бы, что артиновость в каком-то смысле двойственна к нетеровости, однако это совсем не так. Самые простые нетеровы кольца: $\mathbb{Z}$, кольцо многочленов над полем и т. п. не являются артиновыми, но:

Теорема 2.8. *Любое артиново слева (справа) кольцо является нетеровым слева (соотв. справа).*

Этот факт следует из двух только что сформулированных результатов. Как и теорема Веддербарна–Артина последняя теорема в нашем курсе идет без доказательства.

Модули над классически полупростым кольцом. Пусть $R = R_1 \oplus \cdots \oplus R_k$ – прямая сумма колец, а M – R-модуль. Легко видеть, что как абелева группа M равен прямой сумме подмодулей R_iM. Для $n = 2$: $m = (1,0)m + (0,1)m$, а если $m = (r_1,0)m' = (0,r_2)m''$, то $m = (1,1)m = (1,0)(0,r_2)m'' + (0,1)(r_1,0)m' = 0$. Таким образом, неприводимый R-модуль – это неприводимый R_i-модуль, на котором остальные R_j, $j \neq i$, действуют тривиально. Можно показать, что конечнопорожденный левый модуль над кольцом $M_n(D)$ (где D – тело) равен прямой сумме нескольких модулей D^n.

Теорема 2.9. *Пусть R – полупростое артиново кольцо. Тогда любой левый R-модуль полупрост, а простые левые модули изоморфны $M_n(D)$-модулям D^n для некоторого тела D.*

Если R – конечномерная алгебра над замкнутым полем F, то все тела D изоморфны F.

Пример неполупростой групповой алгебры. Пусть C_p – циклическая группа порядка p, а F – поле характеристики p. Тогда групповая алгебра $FC_p \cong F[x]/(x^p - 1) \cong F[t]/(t^p)$ – это кольцо усеченных многочленов. Действительно, зададим эпиморфизм $F[x] \to FC_p$, переводящий x в образующую группы C_p. По определению групповой алгебры образ ненулевого многочлена степени меньше, чем p, не равен нулю, а $x^p - 1$ переходит в 0. Поэтому ядро порождено $x^p - 1$. В характеристике p имеем $x^p - 1 = (x-1)^p$. Делая замену переменных $t = x - 1$, получаем второй изоморфизм. Отождествим x с образующей C_p. Радикал FC_p – это главный идеал, порожденный $x - 1$, потому что $x - 1$ нильпотентнен, а $FC_p/(x-1) \cong F$ полупросто.

Представление $\pi : C_p \to \mathrm{GL}_n(F)$ однозначно определяется матрицей π_x. $(\pi_x - E)^p = \pi_x^p - E = 0$, откуда все собственные числа π_x равны 1. Если $\pi \neq 1$, то π – точное, потому что C_p простая группа. Следовательно π_x имеет порядок p. Если π неразложимое, то π_x – жорданова клетка порядка $\leqslant p$ с 1 по диагонали.

Кольцо представлений. Скелет категории конечномерных линейных представлений группы G над полем F (или эквивалентной ей категории FG-модулей) является малой категорией, потому что уже категория матричных представлений малая (объект – гомоморфизм $G \to \mathrm{GL}_n(F)$; имеем право образовать множество всех таких гомоморфизмов).

На объектах этого скелета мы хотим определить структуру полукольца. Проще сделать это, если стартовать с категории FG-модулей. Итак: элементы полукольца $\mathrm{Rep}_F(G)$ – классы изоморфизмов FG-модулей, конечномерных над F.

$$[M] + [N] := [M \oplus N]; \qquad [M] \cdot [N] := [M \otimes N].$$

Ясно, что результат не зависит от представителя класса изоморфизма. Свойства полукольца проверены в предложении 1.1 главы 10. Из этого полукольца можно универсальным образом сделать кольцо $R_F(G)$, подействовав на него левым сопряженным к функтору вложения категории колец в категорию полуколец (везде есть 1, которую сохраняют морфизмы). Конструктивно: $R_F(G) := \{[M] - [N]\}$, где M и N пробегают F-конечномерные FG-модули, с операциями, продолженными по линейности с полукольца $\mathrm{Rep}_F(G)$.

3. Лемма Шура

Пусть сначала R – произвольное кольцо.

Лемма 3.1 (Лемма Шура)**.** *Пусть U и V простые R-модули. Тогда*

(1) *Если $U \ncong V$, то* $\mathrm{Hom}(U, V) = 0$.
(2) $\mathrm{End}(U)$ – *тело.*

Доказательство. Пусть $\varphi : U \to V$ – гомоморфизм R-модулей. Образ φ – подмодуль в V, а ядро – подмодуль в U. Так как U и V простые, есть два варианта:

(1) $\mathrm{Ker}\,\varphi = U$ и $\mathrm{Im}\,\varphi = 0$, т. е. $\varphi = 0$.
(2) $\mathrm{Ker}\,\varphi = 0 \neq U$ и $\mathrm{Im}\,\varphi = V$, т. е. φ – изоморфизм.

□

Следствие 3.2. *Пусть F – алгебраически замкнутое поле, R является F-алгеброй, а U – простой R-модуль конечномерный над F. Тогда* $\mathrm{End}_R(U) \cong F$. *Другими словами, любой эндоморфизм простого конечномерного R-модуля – это умножение на константу (гомотетия).*

Доказательство. Любой R-линейный эндоморфизм является F-линейным, поэтому $\mathrm{End}_R(U) \subseteq \mathrm{End}_F(U) \cong M_n(F)$, где $n = \dim_F U$. С другой стороны, если $\varphi \in \mathrm{End}_R(U)$, то и $\alpha\varphi \in \mathrm{End}_R(U)$ для любого $\alpha \in F$. Поэтому $\mathrm{End}_R(U)$ является подпространством в n^2-мерном пространстве $\mathrm{End}_F(U)$. Доказательство заканчивает следующая лемма. □

Тело D, являющееся F-алгеброй, называется алгеброй с делением над F.

Лемма 3.3. *Любая конечномерная алгебра D с делением над алгебраически замкнутым полем F изоморфна F.*

Доказательство. Для $\alpha \in D$ рассмотрим гомоморфизм F-алгебр $F[t] \to D$, $t \mapsto \alpha$. Его образ $F[\alpha]$ – подкольцо в D. Его ядро – главный идеал $pF[t]$, так что $F[\alpha] \cong F[t]/(p)$. Если $\deg p > 1$, то он раскладывается на нетривиальные множители, так как F замкнуто. Но тогда в $F[\alpha]$ есть делители нуля, чего не может быть. Следовательно, $\deg p = 1$, т. е. $p = t - \alpha$, откуда $\alpha \in F$. □

Неприводимые представления абелевой группы.

Следствие 3.4. *Каждое неприводимое конечномерное представление абелевой группы над алгебраически замкнутым полем одномерно.*

Доказательство. Пусть V – простой конечномерный FG-модуль, а F алгебраически замкнуто. Для любых $h, g \in G$ имеем $hgv = ghv$, так что умножение на h является автоморфизмом V. По следствию 3.2 любой автоморфизм V – это умножение на скаляр. Таким образом, любое F-линейное подпространство V является FG-подмодулем. Для того чтобы быть простым V должно не иметь нетривиальных подпространств, т. е. быть одномерным. □

Обратное утверждение: "*Если все неприводимые представления группы G одномерны, то группа G абелева.*" тоже верно. Оно будет доказано ниже.

Для поля, не являющегося алгебраически замкнутым, утверждение неверно. Циклическая группа $C_4 = \langle g \rangle$ имеет над $\mathbb{R}$ двумерное представление $g \mapsto \pi_g = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$. Так как $\chi_{\pi_g}(t) = t^2 + 1$ неприводим над $\mathbb{R}$, матрица π_g не приводится к (блочно) треугольному виду, следовательно представление неприводимо.

4. Полная приводимость

Однозначность разложения на неприводимые. Доказательство следующего результата совпадает с доказательством следствия 9.3 главы 8.

Теорема 4.1 (Жордана–Гёльдера). *Пусть R – ассоциативное кольцо с 1. Если R-модуль обладает композиционным рядом, то набор факторов этого ряда определен однозначно с точностью до изоморфизма.*

"Набор" в этом контексте означает мультимножество, т. е. множество с учетом кратности, с формальной точки зрения – множество вместе с функцией из него в натуральные числа.

Следствие 4.2 (теорема Ремака–Крулля–Шмидта). *Если линейное представление вполне приводимо, то набор неприводимых прямых слагаемых определен однозначно с точностью до изоморфизма.*

Полная приводимость. В отличие от примера в конце параграфа 2, если порядок группы не делится на характеристику поля, то любое конечномерное представление вполне приводимо.

Теорема 4.3 (Машке). *Если характеристика поля F не делит порядок конечной группы G, то любое конечномерное линейное представление $\pi : G \to \mathrm{GL}(V)$ над F вполне приводимо.*

Для доказательства теоремы Машке достаточно проверить, что любой подмодуль $U \leqslant V$ выделяется прямым слагаемым. Это условие равносильно наличию ретракции $V \twoheadrightarrow U$. Доказательство следующего утверждения аналогично доказательству предложения 7.1 главы 8, но проще за счет коммутативности операции сложения.

Лемма 4.4. *Пусть R – ассоциативное кольцо с 1, а $U \leqslant V$ – R-модули. Следующие условия эквивалентны.*

(1) *Существует подмодуль $W \leqslant V$ такой, что $V = U \oplus W$.*
(2) *Существует ретракция $V \twoheadrightarrow U$.*

Доказательство теоремы Машке. Пусть U – FG-подмодуль V. Так как любой F-модуль имеет прямое дополнение, существует F-линейная ретракция $\rho : V \twoheadrightarrow U$. Определим F-линейное отображение $\rho' : V \to U$ формулой

$$\rho'(v) = \frac{1}{|G|} \sum_{g \in G} g\rho(g^{-1}v) \in \frac{1}{|G|} \sum_{g \in G} g\rho(V) = \frac{1}{|G|} \sum_{g \in G} gU \subseteq U.$$

(по условию $|G|$ обратимо в F). Проверим, что это отображение FG-линейно. Для любого $h \in G$ имеем

$$\rho'(hv) = \frac{1}{|G|}\sum_{g\in G} g\rho\big(g^{-1}hv\big) = h\left(\frac{1}{|G|}\sum_{g\in G} h^{-1}g\rho\big((h^{-1}g)^{-1}v\big)\right) = h\rho'(v)$$

Далее, $u \in U \implies g^{-1}u \in U \implies g\rho(g^{-1}u) = u$, откуда $\rho'(u) = u$. Таким образом, ρ' является ретракцией FG-модулей, что по предыдущей лемме влечет, что U выделяется прямым слагаемым. Доказательство заканчивается индукцией по размерности V над F. □

Следствие. Предположим, что характеристика поля F не делит порядок конечной группы G. Если все неприводимые представления группы G над полем F одномерны, то G абелева.

Доказательство. По теореме Машке любое представление является прямой суммой неприводимых. Если все неприводимые представления одномерны, то, в частности, регулярное представление является прямой суммой одномерных. В матричном виде это означает, что любой элемент группы отображается в диагональную матрицу. Регулярное представление точно, так что группа вкладывается в группу диагональных матриц, которая абелева. □

Унитаризуемость. Трюк усреднения отображения по группе, использованный для доказательства теоремы Машке, помогает доказать, что над $\mathbb{C}$ любую группу можно вложить не в полную линейную, а в унитарную группу. Пусть $B : V \times V \to \mathbb{C}$ – эрмитово скалярное произведение на комплексном векторном пространстве V. Напомним, что (классическая) унитарная группа $\mathrm{U}(B) = \mathrm{U}(V, B)$ – это подгруппа в $\mathrm{GL}(V)$, состоящая из операторов φ удовлетворяющих равенству

$$B\big(\varphi(x), \varphi(y)\big) = B\big(x, y\big) \text{ для любых } x, y \in V.$$

Теорема 4.5. *Пусть G – конечная группа, а V – конечномерное векторное пространство над $\mathbb{C}$. Для любого представления $\pi : G \to \mathrm{GL}(V)$ существует эрмитово скалярное произведение $B : V \times V \to \mathbb{C}$ такое, что $\operatorname{Im}\pi \subseteq \mathrm{U}(B)$.*

Доказательство. Пусть B' – произвольное скалярное произведение на V. Тогда

$$B\big(x, y\big) := \frac{1}{|G|}\sum_{g\in G} B'\big(\pi_g(x), \pi_g(y)\big)$$

является G-инвариантным скалярным произведением. □

Доказательство теоремы Машке над $\mathbb{C}$. Если $\operatorname{Im}\pi \subseteq \mathrm{U}(B)$, а U – G-инвариантное подпространство, то $U^{\perp_B}$ также G-инвариантно, а $V = U \oplus U^{\perp_B}$, см. предложение 4.7 и лемму 5.8 главы 7. □

Еще один бонус унитаризуемости состоит в том, что унитарная группа компактна в топологии, индуцированной (любой) нормой на $\mathrm{End}(V)$, в отличие от $\mathrm{GL}(V)$.

5. Характеры представления

До конца настоящей главы мы рассматриваем только конечномерные представления конечных групп. Пусть $\rho : G \to \mathrm{GL}(V)$ – такое представление.

Определение 5.1. Характером представления ρ называется функция $\chi_\rho : G \to F$ сопоставляющая каждому элементу $g \in G$ след оператора ρ_g:

$$\chi_\rho(g) = \operatorname{Tr}\rho_g.$$

Напомним свойства следа:

(1) $\operatorname{Tr}(A + B) = \operatorname{Tr}\big(\operatorname{diag}(A, B)\big) = \operatorname{Tr} A + \operatorname{Tr} B$;
(2) $\operatorname{Tr} AB = \operatorname{Tr} BA$ как только произведения существуют;

(3) $\operatorname{Tr} C^{-1}AC = \operatorname{Tr} A$;
(4) След матрицы равен сумме ее собственных чисел (над замыканием поля) с учетом алгебраической кратности.

Простейшие свойства характеров. Первые 2 пункта следующего утверждения очевидны.

Предложение 5.2 (простейшие свойства характеров)**.** *Пусть π и ρ – представления группы G, а $\chi = \chi_\rho$.*

(1) $\chi(1) = n = \deg \rho$.
(2) *Характеры эквивалентных представлений равны.*
(3) $\chi(hgh^{-1}) = \chi(g)$. *Последнее свойство обычно выражается так: характер представления является* центральной функцией *на группе или* функцией классов *(class function). Иными словами, значение характера на элементе $g \in G$ зависит не от самого этого элемента, а только от его класса сопряженности.*
(4) $\chi_{\pi\oplus\rho} = \chi_\pi + \chi_\rho$.
(5) $\chi_{\pi\otimes\rho} = \chi_\pi \cdot \chi_\rho$.
(6) *Если $F \subseteq \mathbb{C}$, то $\chi(g^{-1}) = \overline{\chi(g)}$.*
(7) *Если $F \subseteq \mathbb{C}$, то $|\chi(g)| \leqslant \deg \rho$.*

Доказательство. (4). Можно считать, что π и ρ – матричные представления. Тогда матрица $(\pi \oplus \rho)_g$ равна $\operatorname{diag}(\pi_g, \rho_g)$, а ее след равен сумме следов.

(5). При доказательстве этого пункта удобно говорить о линейных представлениях $\pi : G \to \mathrm{GL}(U)$ и $\rho : G \to \mathrm{GL}(V)$. Пусть u и v – базисы пространств U и V соответственно Рассмотрим матрицу оператора $\pi_g \otimes \rho_g$ в базисе $w = \{u_i \otimes v_j\}$.

$$\pi_g \otimes \rho_g(u_i \otimes v_j) = \pi_g(u_i) \otimes \rho_g(v_j) = \left(\sum_k u_k \pi_{ki}(g)\right) \otimes \left(\sum_m v_m \rho_{mj}(g)\right) = \sum_{k,m} u_k \otimes v_m \pi_{ki}(g)\rho_{mj}(g)$$

Таким образом, диагональные элементы матрицы $(\pi_g \otimes \rho_g)_w$ равны $\pi_{ii}(g)\rho_{jj}(g)$, а их сумма:

$$\sum_{i,j} \pi_{ii}(g)\rho_{jj}(g) = \left(\sum_i \pi_{ii}(g)\right)\left(\sum_j \rho_{jj}(g)\right) = (\operatorname{Tr} \pi_g)(\operatorname{Tr} \rho_g).$$

(6). Каждый элемент группы имеет конечный порядок, поэтому его собственные числа λ_i – корни из 1. В частности, $|\lambda_i| = 1$, откуда $\lambda_i^{-1} = \overline{\lambda_i}$. Таким образом,

$$\chi(g^{-1}) = \operatorname{Tr} \rho_{g^{-1}} = \sum_i \lambda_i^{-1} = \sum_i \overline{\lambda_i} = \overline{\chi(g)}.$$

(7). $|\chi(g)| = |\sum \lambda_i| \leqslant \sum |\lambda_i| = 1 + \cdots + 1 = \deg \rho$. □

Двойственное представление и его характер. Как мы знаем, переход от пространства к сопряженному является контравариантным функтором. Поэтому для *левого* линейного представления $\pi : G \to \mathrm{GL}(V)$ имеем *правое* линейное представление (антигомоморфизм) $\pi' : G \to \mathrm{GL}(V^*)$, заданное формулой $\pi'_g(f)(x) := f\big(\pi_g(x)\big)$. Так как мы хотим, чтобы *двойственное представление* $\pi^* : G \to \mathrm{GL}(V^*)$ было бы левым, определим его формулой

$$\pi^*_g(f)(x) := f\big(\pi_g^{-1}(x)\big).$$

Нетрудно видеть, что матрицы операторов π_g и π'_g в двойственных базисах пространств V и V^* транспонированы друг к другу. Следовательно, $(\pi^*_g)_{e^*} = (\pi_g)_e^{-\mathsf{T}}$, где $a^{-\mathsf{T}}$ – обратная транспонированная (т. е. контрградиентная) к a.

Предложение 5.3. *Пусть $\pi : G \to \mathrm{GL}(V)$ конечномерное комплексное представление группы G. Тогда $\chi_{\pi^*} = \overline{\chi_\pi}$.*

6. Соотношения ортогональности

Пусть G – конечная группа, а F – поле, характеристика которого не делит порядок G. На пространстве функций F^G определим симметричную билинейную форму

$$B(\eta,\theta) := \frac{1}{|G|}\sum_{g\in G}\eta(g)\theta(g^{-1}).$$

Ортогональность таких функций по умолчанию будет означать ортогональность относительно формы B. Рассмотрим два линейных представления $\pi : G \to \mathrm{GL}(U)$ и $\rho : G \to \mathrm{GL}(V)$ и произвольное линейное отображение $\varphi : U \to V$. Назовем усреднением этого отображения относительно группы G отображение

$$\varphi_0 = \frac{1}{|G|}\sum_{g\in G}\rho_g \circ \varphi \circ \pi_g^{-1}.$$

В доказательстве теоремы Машке мы проверили, что это отображение G-эквивариантно, т. е. является морфизмом представлений. По лемме Шура, если представления π и ρ неприводимы и неэквивалентны, то $\varphi_0 = 0$.

Лемма 6.1 (матричная форма леммы Шура). *Пусть $\pi : G \to \mathrm{GL}_n(F)$ и $\rho : G \to \mathrm{GL}_m(F)$ – неприводимые матричные представления группы G, а $A \in \mathrm{M}_{m,n}(F)$. Положим $A_0 = \frac{1}{|G|}\sum_{g\in G}\rho_g A\pi_g^{-1}$.*

(1) *Если π и ρ не эквивалентны, то $A_0 = 0$.*
(2) *Если $\pi = \rho$, а поле F алгебраически замкнуто, то $A_0 = \frac{\operatorname{Tr} A}{\deg\pi}E$. В частности, $\deg\pi \neq 0$ в F.*

Доказательство. Заметим, что оператор умножения на матрицу A_0 является G-эквивариантным, т. е. A_0 является морфизмом матричных представлений. Если π и ρ не эквивалентны, то по лемме Шура $A_0 = 0$.

В случае (2) по лемме Шура $A_0 = \lambda E$ для некоторого $\lambda \in F$.

$$\lambda n = \operatorname{Tr} A_0 = \frac{1}{|G|}\sum_{g\in G}\operatorname{Tr}(\rho_g A\pi_g^{-1}) = \frac{1}{|G|}\sum_{g\in G}\operatorname{Tr} A = \operatorname{Tr} A.$$

□

Теорема 6.2 (соотношения ортогональности Шура). *Пусть π, ρ – неприводимые матричные представления группы G над полем F, характеристика которого не делит $|G|$.*

(1) *Если π и ρ не эквивалентны, то для любых $i, j \in [\deg\pi]$ и $k, h \in [\deg\rho]$ функции π_{ij} и ρ_{kh} ортогональны.*
(2) *Если поле F алгебраически замкнуто, то*

$$B(\pi_{ij},\pi_{kh}) = \frac{1}{\deg\pi}\delta_{jk}\delta_{ih}.$$

Доказательство. Сразу следует из предыдущей леммы с $A = E_{jk}$. В случае (2) с $j = k$ равенство $A_0 = \frac{1}{\deg\pi}E$ влечет

$$\frac{1}{|G|}\sum_{g\in G}\pi_{ij}(g)\pi_{jh}(g^{-1}) = \frac{1}{\deg\pi}\delta_{ih}.$$

□

Если $F = \mathbb{C}$, то мы можем изучать только унитарные представления. Для унитарных представлений $\pi_{ij}(g^{-1}) = (\pi_g^{-1})_{ij} = \overline{\pi_{ji}(g)}$. Определим эрмитово скалярное произведение C на

множестве функций $\mathbb{C}^G$ по формуле

$$C(\eta,\theta) := \frac{1}{|G|}\sum_{g\in G}\eta(g)\overline{\theta(g)}$$

Тогда из соотношений ортогональности вытекает следующее утверждение.

Следствие 6.3. *Пусть $\pi^{(i)}$ – все (с точностью до эквивалентности) неприводимые унитарные матричные представления группы G над $\mathbb{C}$, Тогда набор функций $\sqrt{\deg\pi^{(i)}}\pi^{(i)}_{jk}$ ортонормирован относительно формы C.*

Возвращаемся к общей ситуации алгебраически замкнутого поля F, характеристика которого не делит порядок группы G.

Следствие 6.4. *Пусть $\pi^{(i)}$ – все (с точностью до эквивалентности) неприводимые представления группы G над F, а $\chi_i = \chi_{\pi^{(i)}}$. Тогда набор χ_i является ортонормированным набором функций классов.*

Доказательство.

$$B(\chi_i,\chi_k) = B\left(\sum_j \pi^{(i)}_{jj}, \sum_h \pi^{(k)}_{hh}\right) = \sum_{jh} B(\pi^{(i)}_{jj},\pi^{(k)}_{hh}) = \delta_{ik}\sum_j B(\pi^{(i)}_{jj},\pi^{(i)}_{jj}).$$

Таким образом, характеры разных представлений ортогональны, а $B(\chi_i,\chi_i)$ является суммой n_i слагаемых, равных $1/n_i$, где $n_i = \deg\pi^{(i)}$. □

По умолчанию, везде далее F – алгебраически замкнутое поле характеристики 0.

Заметим, что для характеров комплексных представлений χ и θ имеем $B(\chi,\theta) = C(\chi,\theta)$. Из полной приводимости и первого соотношения ортогональности сразу следует, что на множестве характеров обе формы принимают целые неотрицательные значения.

Пусть π – неприводимое, а ρ – любое представление группы G. По теореме Машке ρ раскладывается в прямую сумму неприводимых.

Следствие 6.5. *Количество раз, которое π встречается в разложении ρ на неприводимые, равно $B(\chi_\pi,\chi_\rho)$.*

Для доказательства достаточно вспомнить, что характер прямой суммы равен сумме характеров прямых слагаемых.

Заметим, что из этого следует теорема Ремака–Крулля–Шмидта для частного случая замкнутых полей характеристики 0.

Следствие 6.6. $\pi\sim\rho \iff \chi_\pi = \chi_\rho$.

Это утверждение совершенно неверно при $\operatorname{char}F = p \neq 0$. В качестве контрпримера достаточно взять прямую сумму любого представления в количестве $p+1$ штук, его характер такой же, как у самого представления.

Следствие 6.7. $B(\chi_\pi,\chi_\pi) = 1 \iff \pi$ *неприводимо.*

Неприводимый характер – это характер неприводимого преставления.

Следствие 6.8 (ортогональность характеру тривиального представления)**.** *Для любого неприводимого характера $\chi \neq 1$*

$$\sum_{g\in G}\chi(g) = 0.$$

7. Разложение регулярного представления

Пусть G действует на множестве X. Это действие индуцирует линейное действие G на свободном модуле $\langle X\rangle_F$. Обозначим через χ характер этого действия.

Лемма 7.1. $\chi(g) = |\operatorname{Fix}_X(g)|$.

Доказательство. Пусть χ – характер представления π. Ясно, что для любого $g \in G$ каждый столбец матрицы π_g в базисе X имеет ровно одну единицу и остальные нули. Столбцы, в которых эти единицы стоят на диагонали соответствуют $x \in X$, для которых $gx = x$. □

Обозначим через reg – регулярное представление.

Следствие 7.2. $\chi_{\mathrm{reg}}(g) = \begin{cases} |G|, & \text{если } g = 1, \\ 0, & \text{иначе.} \end{cases}$

Пусть $\chi_1, \ldots, \chi_s$ – характеры всех неприводимых представлений, а $n_i = \chi_i(1)$ – их степени.

Теорема 7.3 (Веддербарна). *Каждое неприводимое представление входит в регулярное с кратностью, равной степени.*

Доказательство. Вычислим $B(\chi_{\mathrm{reg}}, \chi_i)$:

$$B(\chi_{\mathrm{reg}}, \chi_i) = \frac{1}{|G|}\sum_{g\in G} \chi_{\mathrm{reg}}(g)\chi_i(g^{-1})$$

так как все слагаемые, кроме слагаемых с $g = 1$, равны 0, а $\chi_{\mathrm{reg}}(1) = |G|$, то эта сумма равна

$$\frac{1}{|G|}|G|\chi_i(1) = \chi_i(1) = n_i.$$

□

Следствие 7.4. $n_1^2 + \ldots + n_s^2 = |G|$

Теорема 7.5 (Петера–Вейля). *Матричные элементы π_{ij}, $1 \leqslant i, j \leqslant \dim(\pi)$, всех неэквивалентных неприводимых представлений π, образуют базис пространства $FG = F^G$ всех функций на G.*

В случае $F = \mathbb{C}$, а все π унитарны, этот базис является ортогональным относительно формы C.

Доказательство. Пусть $\pi^{(1)}, \ldots, \pi^{(s)}$ – набор всех неэквивалентных между собой неприводимых представлений группы G. По предыдущему следствию количество функций $\pi_{ij}^{(k)}$ равно $|G| = \dim FG$. Пусть $\sum_{k,i,j} \alpha_{kij}\pi_{ij}^{(k)} = 0$. Тогда по теореме 6.2

$$\alpha_{hqp} = \dim \pi^h \sum_{k,i,j} \alpha_{kij} B(\pi_{ij}^{(k)}, \pi_{pq}^{(h)}) = \dim \pi^h B(\sum \alpha_{kij}\pi_{ij}^{(k)}, \pi_{pq}^{(h)}) = 0.$$

Поэтому функции π_{ij} линейно независимы. Второе утверждение вытекает из следствия 6.3. □

8. Количество неприводимых представлений

Пусть $\varphi : G \to F$ – функция классов, а $\rho : G \to \mathrm{GL}(V)$ – линейное представление. Следующая формула определяет усреднение с весом φ:

$$\rho_\varphi := \frac{1}{|G|}\sum_{g\in G} \varphi(g^{-1})\rho_g \in \mathrm{End}(V).$$

Заметим, что $\mathrm{Tr}(\rho_\varphi) = B(\varphi, \chi_\rho)$.

Лемма 8.1. *Для любого поля F оператор ρ_φ является морфизмом представлений.*

Если ρ неприводимо степени n, а F – алгебраически замкнуто, то ρ_φ гомотетия с константой $B(\varphi, \chi_\rho)/n$.

Доказательство. Для любого $h \in G$ имеем

$$\rho_h \rho_\varphi = \frac{1}{|G|} \sum_{g \in G} \varphi(g^{-1}) \rho_h \rho_g \rho_{h^{-1}} \rho_h = \frac{1}{|G|} \sum_{g \in G} \varphi({}^h(g)^{-1}) \rho_{{}^h g} \rho_h = \rho_\varphi \rho_h.$$

Тот факт, что ρ_φ – умножение на константу $\lambda \in F$, следует из первого утверждения и леммы Шура над замкнутым полем. Тогда $\lambda n = \operatorname{Tr} \rho_\varphi = B(\varphi, \chi_\rho)$, что и требовалось. □

Теорема 8.2. *Пусть F – алгебраически замкнутое поле характеристики 0. Неприводимые характеры $\chi_1, \dots, \chi_s$ образуют ортонормированный базис пространства функций классов на группе G.*

Доказательство. Мы уже знаем, что $\chi_1, \dots, \chi_s$ ортонормированная система относительно формы B. Обозначим через Y пространство функций классов, а через X – подпространство, порожденное неприводимыми характерами. Так как B невырождена на X, то по лемме 4.7 главы 7 $Y = X \oplus X^{\perp_B}$. Поэтому достаточно показать, что любая функция φ, ортогональная ко всем $\chi_1, \dots, \chi_s$ равна 0. По предыдущей лемме ρ_φ равно 0 для всех неприводимых представлений ρ.

По теореме Машке любое представление π вполне приводимо. На матричном языке это означает, что $\pi_g = \operatorname{diag}(\rho_g^{(1)}, \dots, \rho_g^{(k)})$ для некоторых неприводимых $\rho^{(1)}, \dots, \rho^{(k)}$. Из того, что $\rho_\varphi^{(i)} = 0$ для всех прямых слагаемых сразу следует, что $\pi_\varphi = 0$. В частности, для регулярного представления $\operatorname{reg}_\varphi = 0$. Применим это равенство к $1_G \in FG$:

$$0 = \operatorname{reg}_\varphi(1_G) = \frac{1}{|G|} \sum_{g \in G} \varphi(g^{-1}) \operatorname{reg}_g(1_G) = \frac{1}{|G|} \sum_{g \in G} \varphi(g^{-1}) g.$$

Так как G – базис групповой алгебры, то $\varphi(g^{-1}) = 0$ для всех $g \in G$, а это и значит, что $\varphi = 0$. □

Следствие 8.3. *Количество неприводимых представлений равно количеству классов сопряженных элементов.*

Доказательство. И то, и другое равно размерности пространства функций классов. □

Предложение 8.4. *Для любого поля F центр групповой алгебры $FG = F^G$ состоит из функций классов.*

Доказательство. Пусть $f = \sum_{g \in G} \alpha_g g \in FG$. Элемент f лежит в центре $\iff$ $f = h^{-1} f h$ $\forall h \in G$. Последнее условие равносильно тому, что

$$\sum_{g \in G} \alpha_g g = h^{-1} \left(\sum_{g \in G} \alpha_g g \right) h = \sum_{g \in G} \alpha_g g^h \iff \alpha_g = \alpha_{g^h} \ \forall g, h \in G$$

Таким образом, f лежит в центре тогда и только тогда, когда коэффициенты при сопряженных элементах равны, что и означает, что f соответствует функции классов. □

Теорема 8.5 (Второе соотношение ортогональности)**.** *Пусть $\chi_1 = 1, \chi_2, \dots, \chi_s$ – все различные неприводимые характеры. Если $h, g \in G$, то*

$$\sum_{i=1}^{s} \chi_i(h) \chi_i(g^{-1}) = \begin{cases} |C_G(h)|, & h \text{ сопряжен с } g, \\ 0, & \text{иначе.} \end{cases}$$

Доказательство. Пусть $g_1 = 1, g_2, \ldots, g_s$ – представители различных классов сопряженности группы G (по следствию 8.3 их количество равно s). Положив $m_i = |g_i^G|$, первое соотношение ортогональности можно переписать в виде

$$\frac{1}{|G|}\sum_{k=1}^{s} m_k \chi_i(g_k)\chi_j(g_k^{-1}) = \delta_{ij}.$$

Рассмотрим матрицы $A, D \in \mathrm{M}_s(F)$ с элементами $a_{ik} = \chi_i(g_k)$ и $d_{pq} = m_p\chi_q(g_p^{-1})$. Тогда $\frac{1}{|G|}AD = E$.

Так как A и D квадратные, то и $\frac{1}{|G|}DA = E$. Записав последнее равенство поэлементно, получим

$$\frac{1}{|G|}\sum_{k=1}^{s} m_i \chi_k(g_i^{-1})\chi_k(g_j) = \delta_{ij}$$

Заметим, что $m_i = |G : C_G(g_i)|$, откуда $|C_G(g_i)| = |G|/m_i$. Таким образом,

$$\sum_{k=1}^{s} \chi_k(g_j)\chi_k(g_i^{-1}) = \delta_{ij}|C_G(g_i)|.$$

Так как характеры являются функциями классов, а порядок централизатора одинаков для всех элементов данного класса сопряженных, то можно заменить g_j на g, g_i на h, а δ_{ij} на δ_{g^G,h^G}. □

9. Таблицы характеров

В строках таблицы – все неприводимые характеры ($\chi_i := \chi_{\pi_i}$), в столбцах – классы сопряженных элементов, а в клетках – значение характера на представителе g_i класса. Из следствия 8.3 вытекает, что эта таблица – квадратная ($s \times s$). В первой строке стоят единицы – характер единичного представления. В первом столбце – $\chi_i(1) = \deg \pi_i$. По следствию 7.4 $\sum_{i=1}^{s} \chi_i(1)^2 = |G|$. Из соотношения ортогональности характеров следует, что строки ортонормированы относительно формы B:

$$B(\chi, \theta) = \frac{1}{|G|}\sum_{g\in G} \chi(g)\theta(g^{-1}) = \frac{1}{|G|}\sum_{i=1}^{s} |g_i^G|\chi(g_i)\theta(g_i^{-1}).$$

Над $\mathbb{C}$ для любого характера $\theta(g^{-1}) = \overline{\theta(g)}$, так что строки ортонормированы относительно эрмитова скалярного произведения с матрицей $\frac{1}{|G|}\operatorname{diag}(|g_1^G|, \ldots, |g_s^G|)$. По второму соотношению ортогональности столбцы также ортогональны, а их скалярные квадраты равны $|C_G(g_i)| = |G|/|g_i^G|$.

Если G – абелева, то все неприводимые представления одномерны, т. е. это гомоморфизмы в F^*. Характеры совпадают с представлениями.

$G = \mathbb{Z}/2$

	0	1
χ_1	1	1
χ_2	1	-1

$G = \mathbb{Z}/3$

	0	1	2
η_1	1	1	1
η_2	1	ω	ω^2
η_3	1	ω^2	ω

$G = \mathbb{Z}/2 \times \mathbb{Z}/2$

	(0,0)	(0,1)	(1,0)	(1,1)
θ_1	1	1	1	1
θ_2	1	-1	1	-1
θ_3	1	1	-1	-1
θ_4	1	-1	-1	1

Неприводимые характеры прямого произведения равны произведению неприводимых характеров сомножителей, как, например, $\theta_k(x, y) = \chi_i(x)\chi_j(y)$.

Одномерные представления неабелевой группы G соответствуют представлениям $G^{\mathrm{ab}} = G/[G, G]$.

Нарисуем таблицу характеров для S_3.

g_i^G	e	$(1\,2)$	$(1\,2\,3)$
$\lvert g_i^G\rvert$	1	3	2
χ_1	1	1	1
χ_2	1	-1	1
χ_3	2	0	-1

Приведем регулярное представление S_3 к блочно диагональному виду. Достаточно привести в такому виду образующие. В качестве образующих возьмем $(1\,2)$ и $(1\,2\,3)$.

$$\mathrm{reg}(1\,2) = \begin{pmatrix} & & & 0 & 0 & 1 \\ & & & 0 & 1 & 0 \\ & & & 1 & 0 & 0 \\ 0 & 0 & 1 & & & \\ 0 & 1 & 0 & & & \\ 1 & 0 & 0 & & & \end{pmatrix} \qquad \mathrm{reg}(1\,2\,3) = \begin{pmatrix} 0 & 0 & 1 & & & \\ 1 & 0 & 0 & & & \\ 0 & 1 & 0 & & & \\ & & & 0 & 0 & 1 \\ & & & 1 & 0 & 0 \\ & & & 0 & 1 & 0 \end{pmatrix}$$

Базис u:

$$\begin{pmatrix}1\\1\\1\\1\\1\\1\end{pmatrix} \begin{pmatrix}1\\1\\1\\-1\\-1\\-1\end{pmatrix} \begin{pmatrix}1\\\omega\\\omega^2\\1\\\omega\\\omega^2\end{pmatrix} \begin{pmatrix}\omega^2\\\omega\\1\\\omega^2\\\omega\\1\end{pmatrix} \begin{pmatrix}1\\-1\\0\\-1\\1\\0\end{pmatrix} \begin{pmatrix}0\\1\\-1\\0\\-1\\1\end{pmatrix}$$

$$\mathrm{reg}(1\,2)_u = \begin{pmatrix} 1 & & & & & \\ & -1 & & & & \\ & & 0 & 1 & & \\ & & 1 & 0 & & \\ & & & & 0 & 1 \\ & & & & 1 & 0 \end{pmatrix} \qquad \mathrm{reg}(1\,2\,3)_u = \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & \omega^2 & 0 & & \\ & & 0 & \omega & & \\ & & & & 0 & -1 \\ & & & & 1 & -1 \end{pmatrix}$$

Нарисуем таблицу характеров для неабелевой группы порядка 8.

Центр p-группы нетривиален, а факторгруппа по центру не может быть циклической.

Следовательно, центр неабелевой группы порядка 8 имеет порядок 2.

Фактор по центру изоморфен $\mathbb{Z}/2 \times \mathbb{Z}/2$.

Так как фактор по центру абелев, то центр содержит коммутант, а так как группа неабелева, то коммутант не равен 1.

Следовательно коммутант равен центру.

g_i^G	e	c	g_3	g_4	g_5
$\lvert g_i^G\rvert$	1	1	2	2	2
χ_1	1	1	1	1	1
χ_2	1	1	1	-1	-1
χ_3	1	1	-1	1	-1
χ_4	1	1	-1	-1	1
χ_5	2	-2	0	0	0

10. Таблица характеров как матрица перехода

В этом параграфе F – алгебраически замкнутое поле, характеристика которого не делит порядок группы G. По теореме Веддербарна–Артина групповое кольцо FG изоморфно прямой сумме матричных колец $M_{n_i}(F)$, где i меняется от 1 до s. Очевидно, что центр прямой суммы равен прямой сумме ценров слагаемых, а центр матричного кольца состоит из скалярных матриц. Поэтому центр FG порожден центральными идемпотентами $e_i = (0, \ldots, 0, E, 0, \ldots, 0)$ (единичная матрица стоит на i-м месте).

С другой стороны, мы знаем, что центр алгебры FG порожден элементами $c_j = \sum\limits_{g\in C_i} g$, где C_i – i-й класс сопряженных элементов, а $i = 1, \dots, s$ (это то же самое s, что и выше, потому что это – размерность центра алгебры FG). Для удобства будем считать, что $C_1 = \{1_G\}$.

Пусть $c_i = \sum e_j \lambda_{ji}$, так что матрица из λ_{ji} является матрицей перехода между базисами e и c. Заметим, что проекция прямой суммы матричных колец на j-ю компоненту – это и есть j-е неприводимое представление алгебры FG или, что то же самое, группы G. Таким образом, формулу из начала абзаца можно переписать в виде

$$\pi_j(c_i) = \lambda_{ji} E_{n_j}.$$

Эту формулу легко доказать при помощи леммы Шура, без использования (недоказанной) теоремы Веддербарна–Артина. Пусть V_j – это простой FG-модуль, соответствующий j-му неприводимому представлению. Если a лежит в центре алгебры FG, то умножение на a является эндоморфизмом (любого) FG-модуля. По лемме Шура любой эндоморфизм простого модуля над замкнутым полем – это умножение на элемент поля. Таким образом, $\pi_j(c_i) = \lambda_{ji} E_{n_j}$, где n_j – степень представления π_j

Посчитаем след левой и правой части последнего равенства:

$$\operatorname{Tr}\bigl(\pi_j(c_i)\bigr) = \operatorname{Tr}\bigl(\sum_{g\in C_i} \pi_j(g)\bigr) = \sum_{g\in C_i} \chi_j(g) = |C_i|\chi_j(g_i) = \lambda_{ji} \deg \pi_j,$$

где $g_i \in C_i$, а χ_j – характер π_j. Таким образом, мы доказали следующее утверждение.

Лемма 10.1. *В обозначениях, введенных выше, $\pi_j(c_i) = \lambda_{ji} E_{n_j}$, причем $|C_i|\chi_j(g_i) = \lambda_{ji} \deg \pi_j$.*

11. Степени неприводимых представлений

Обозначим через $\mathbb{A}$ кольцо алгебраических чисел, то есть целое замыкание $\mathbb{Z}$ в $\mathbb{C}$ (см. параграф 20 главы 4). В этом параграфе мы докажем, что степень любого неприводимого представления π делит порядок группы. Для этого мы покажем, что $|G|/\deg \pi \in \mathbb{A}$, то есть это число является корнем унитального многочлена с целыми коэффициентами. Так как $|G|/\deg \pi$ очевидно рационально, то по теореме о рациональных корнях многочлена (теорема 20.6 главы 4) оно должно быть целым.

Лемма 11.1. *Числа λ_{ij} из предыдущей леммы лежат в $\mathbb{A}$.*

Доказательство. Из предложения 8.4 следует, что элементы $c_i = \sum_{g\in C_i} g$ образуют базис центра групповой алгебры FG. Так как $c_i c_j$ также лежит в центре FG, то

$$c_i c_j = \sum_{k=1}^{s} \alpha_{ijk} c_k$$

для некоторых $\alpha_{ijk} \in \mathbb{N}_0$ (число α_{ijk} равно количеству решений уравнения $xy = z$, где $x \in C_i$, $y \in C_j$, а z – фиксированный элемент из C_k). Применяя π_r к вынесенной формуле, получаем $\pi_r(c_i)\pi_r(c_j) = \sum_{k=1}^{s} \alpha_{ijk}\pi_r(c_k)$, откуда по лемме 10.1 $\lambda_{ri}\lambda_{rj} = \sum_{k=1}^{s} \alpha_{ijk}\lambda_{rk}$. Последнее равенство переписывается в виде $\sum_{k=1}^{s}(\alpha_{ijk} - \delta_{ik}\lambda_{rj})\lambda_{rk} = 0$. Зафиксируем r и j. При $i = 1, \dots, s$ получим однородную систему линейных уравнений с матрицей $A - \lambda_{rj}E$, где $a_{ik} = \alpha_{ijk}$. Так как $\chi_r(1_G) \neq 0$, то $\lambda_{r1} \neq 0$, поэтому система имеет ненулевое решение. Следовательно, $\det(A - \lambda_{rj}E) = 0$. Таким образом, λ_{rj} является корнем характеристического многочлена целочисленной матрицы, в частности, целым алгебраическим числом. □

Теорема 11.2. *Степень неприводимого представления делит порядок группы.*

Доказательство. Так как порядок любого элемента конечной группы делит порядок группы, то все собственные числа матрицы $\pi(g)$ являются корнями из 1 для любого представления π и любого $g \in G$. Поэтому $\chi(g)$ является суммой корней из 1 степени $|G|$ для

любого характера χ и, следовательно, лежит в $\mathbb{A}$. Первое соотношение ортогональности 6.4 показывает, что в обозначениях леммы 10.1

$$\sum_{p=1}^{s} |C_p| \chi_q(g_p) \chi_q(g_p^{-1}) = |G|.$$

Подставляя сюда равенство из упомянутой леммы имеем

$$\sum_{p=1}^{s} \deg \pi_q \lambda_{pq} \chi_q(g_p^{-1}) = |G| \iff \frac{|G|}{\deg \pi_q} = \sum_{p=1}^{s} \lambda_{pq} \chi_q(g_p^{-1}) \in \mathbb{A}.$$

Так как $|G|/\deg \pi_q$ очевидно рационально, то по теореме о рациональных корнях многочлена это число целое. □

Этот результат можно уточнять и далее.

Теорема 11.3. *Пусть π – неприводимое представление n группы G, а A – абелева нормальная подгруппа. Тогда $\deg \pi$ делит $|G : A|$.*

Если $A \leqslant G$ абелева подгруппа, то, вообще говоря, неверно, что $\deg \pi$ делит $|G : A|$. Однако в любом случае можно утверждать, что $\deg \pi \leqslant |G : A|$.

Доказательство этих результатов требует более продвинутой техники, поэтому мы его опускаем.

12. Представления прямого произведения

Следующее утверждение не понадобится для нашей основной цели, но дает представление о том, что происходит с декартовым произведением групп. Доказательство оставляется в качестве упражнения.

Предложение 12.1. *$F(H \times G) = FH \otimes_F FG$.*

Пусть $\pi : H \to \mathrm{GL}(U)$ и $\rho : G \to \mathrm{GL}(V)$ – линейные представления. Их тензорным произведением называется представление $\pi \boxtimes \rho : H \times G \to \mathrm{GL}(U \otimes V)$, заданное формулой

$$(\pi \boxtimes \rho)_{(h,g)}(u \otimes v) = \pi_h(u) \otimes \rho_g(v).$$

В частности, $\chi_{\pi \boxtimes \rho}(h, g) = \chi_\pi(h)\chi_\rho(g)$, см. доказательство свойства 5 предложения 5.2. Обозначим через $\mathrm{Irr}(G)$ множество всех неэквивалентных между собой неприводимых представлений группы G. Напомним, что по следствию 8.3 $|\mathrm{Irr}(G)|$ равно количеству смежных классов в G. Теперь у нас все готово, чтобы доказать основной результат настоящего параграфа.

Теорема 12.2.
- *Если π, ρ – неприводимые представления групп H и G. Тогда представление $\pi \boxtimes \rho$ группы $H \times G$ неприводимо.*
- *Каждое неприводимое представление группы $H \times G$ имеет такой вид.*

Доказательство. Пусть π, π' – представления группы H, а ρ, ρ' – представления группы G. Тогда

$$B(\chi_{\pi\boxtimes\rho}, \chi_{\pi'\boxtimes\rho'}) = \frac{1}{|H \times G|} \sum_{(h,g)\in H\times G} \chi_{\pi\boxtimes\rho}(h, g)\chi_{\pi'\boxtimes\rho'}(h^{-1}, g^{-1}) =$$

$$\frac{1}{|H| \cdot |G|} \sum_{h\in H} \sum_{g\in G} \chi_\pi(h)\chi_\rho(g)\chi_{\pi'}(h^{-1})\chi_{\rho'}(g^{-1}) =$$

$$\frac{1}{|H|} \sum_{h\in H} \chi_\pi(h)\chi_{\pi'}(h^{-1}) \cdot \frac{1}{|G|} \sum_{g\in G} \chi_\rho(g)\chi_{\rho'}(g^{-1}) = B(\chi_\pi, \chi_{\pi'})B(\chi_\rho, \chi_{\rho'}).$$

Если π и ρ неприводимы то по следствию 6.7 скалярные квадраты их характеров равны 1. Следовательно, скалярный квадрат храктера $\chi_{\pi\otimes\rho}$ также равен 1, откуда, по тому же следствию, $\pi\boxtimes\rho$ неприводимо.

Если $\pi\not\sim\pi'$ или $\rho\not\sim\rho'$ и все четыре представления неприводимы, то по следствию 6.4 $B(\chi_\pi,\chi_{\pi'})=0$ или $B(\chi_\rho,\chi_{\rho'})=0$. Следовательно, $B(\chi_{\pi\boxtimes\rho},\chi_{\pi'\boxtimes\rho'})=0$, откуда вытекает, что $\pi\boxtimes\rho$ не эквивалентно $\pi'\boxtimes\rho'$. Таким образом, мы построили $|\operatorname{Irr}(H)|\cdot|\operatorname{Irr}(G)|$ неприводимых представлений группы $H\times G$.

Заметим, что класс сопряженных элементов $(h,g)^{H\times G}$ равен (h^H,g^G). Поэтому количество классов сопряженных в $H\times G$ равно количеству классов сопряженных в H умножить на количество классов сопряженных в G. Учитывая, что последнее равно $|\operatorname{Irr}(H)|\cdot|\operatorname{Irr}(G)|$, заключаем, что мы построили все неприводимые представления группы $H\times G$. □

13. Индуцированные представления

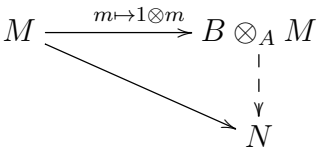
Пусть $\varphi: A\to B$ – гомоморфизм колец (некоммутативных с 1). Если M – левый B-модуль, то его можно рассматривать как левый A-модуль, задав умножение по правилу $a\cdot m:=\varphi(a)m$. Нетрудно видеть, что это сопоставление определяет функтор $\varphi^\#: B\text{-}\mathbf{Mod}\to A\text{-}\mathbf{Mod}$. Этот функтор называется ограничением (в случае, когда φ инъективно, это действительно ограничивает множество “скаляров”). Если задан гомоморфизм групп $G\to H$, то представление группы H индуцирует представление группы G. Мы использовали это в случае $H=G/[G,G]$ при построении таблиц характеров.

Левым сопряженным к функтору ограничения будет функтор тензорного домножения на B. Действительно, φ задает на B структуру правого A-модуля, поэтому левый A-модуль M можно тензорно домножить на B над A. Структура левого B-модуля задается на разложимых тензорах посредством $b'\cdot(b\otimes m):=(b'b)\otimes m$. Этот функтор называется индуцированием и обозначается $\varphi_\#: A\text{-}\mathbf{Mod}\to B\text{-}\mathbf{Mod}$.

До сих пор мы изучали только тензорные произведения над коммутативными кольцами, поэтому сформулируем хотя бы определения для некоммутативного случая. Пусть A – кольцо, M – левый, а N – правый A-модуль. Так как M и N – абелевы группы, то можно образовать абелеву группу $N\otimes_{\mathbb{Z}} M$. Тогда $N\otimes_A M:=N\otimes_{\mathbb{Z}} M/K$, где K – подгруппа, порожденная всеми элементами $na\otimes m-n\otimes am$, $a\in A$, $m\in M$, $n\in N$.

Если N – левый B-модуль, правый A-модуль, и имеет место внешняя ассоциативность: $(bn)a=b(na)$ для любых $b\in B$, $n\in N$ и $a\in A$, то N называется B-A-бимодулем. Если N является B-A-бимодулем, то на $N\otimes_A M$ естественно определена структура левого B-модуля: $b\cdot(n\otimes m):=(bn)\otimes m$. Проверка корректности последнего определения использует, в частности, внешнюю ассоциативность.

Вернемся к ситуации, когда задан гомоморфизм колец $\varphi: A\to B$. Тот факт, что функторы $\varphi_\#$ и $\varphi^\#$ сопряжены, выражается следующим универсальным свойством (см. теорему 5.2 главы 9):



Для любого A-линейного отображения A-модуля M в B-модуль N существует единственное B-линейное отображение $B\otimes_A M\to N$, для которого эта диаграмма коммутативна.

Пусть $H\leqslant G$, а V – (левый) FH-модуль. Функторы ограничения и индуцирования, соответствующие вложению H в G обозначаются res^G_H и ind^G_H. Таким образом, индуцированный модуль определяется как $\operatorname{ind}^G_H(V)=V^G=FG\otimes_{FH}V$. Построим этот модуль явно. Пусть $G=g_1H\sqcup\ldots\sqcup g_nH$ – разбиение G на левые смежные классы по H ($g_1=1$). Каждый элемент групповой алгебры FG однозначно представляется в виде $g_1x_1+\ldots+g_nx_n$, где $x_i\in FH$. Иными словами, FG является свободным правым FH модулем с базисом $g_1,\ldots,g_n$. Таким образом,

$V^G = g_1 \otimes V \oplus \cdots \oplus g_n \otimes V \cong V^n$, как F-модуль. Поэтому $\dim_F V^G = |G : H| \cdot \dim_F V$. Каждый элемент V^G однозначно записывается в виде суммы $g_1 \otimes v_1 + \cdots + g_n \otimes v_n$. Если $u_1, \ldots, u_m$ базис V над F, то $g_i \otimes u_j$ представляет собой F-базис V^G. Посмотрим на индуцированное представление на матричном языке. Пусть $\pi : H \to \mathrm{GL}(V)$ – матричное представление, соответствующее FH-модулю $V = F^m$. Возьмем следующий (упорядоченный) F-базис модуля V^G: $g_1 \otimes e_1, \ldots, g_1 \otimes e_m, \ldots, g_n \otimes e_1, \ldots, g_n \otimes e_m$. Пусть $g \in G$. Запишем в этом базисе матрицу π_g^G умножения на элемент g. Пусть $gg_i = g_k h$ для некоторого $k = k(i) \in [n]$ и $h \in H$. Тогда $h = g_k^{-1} g g_i$. Получим $g \cdot (g_i \otimes e_j) = g_k h \otimes e_j = g_k \otimes h e_j = g_k \otimes (\pi_h)_{*j}$.

Доопределим π до функции $G \to \mathrm{M}_m(F)$, положив $\pi_g = 0$ при всех $g \in G \setminus H$. Тогда в выбранном базисе матрица индуцированного представления $\pi^G = \mathrm{ind}_H^G(\pi) : G \to \mathrm{GL}(V^G)$ имеет блочный вид:

$$\pi_g^G = \begin{pmatrix} \pi_{g_1^{-1} g g_1} & \cdots & \pi_{g_1^{-1} g g_n} \\ \vdots & \cdots & \vdots \\ \pi_{g_n^{-1} g g_1} & \cdots & \pi_{g_n^{-1} g g_n} \end{pmatrix}$$

(в каждом столбце ровно 1 ненулевой блок, а так как матрица обратима, то и в каждой строке тоже). Заметим, что на категории матричных представлений ind_H^G становится функтором только после выбора трансверсали.

Пусть $H \leqslant G$, $\pi : H \to \mathrm{GL}(V)$ – линейное представление, $g \in G$, а $g_1, \ldots, g_n$ – представители левых смежных классов G по H.

Упражнение 13.1. Представление $(1_H)^G$ является перестановочным. Оно соответствует действию G на множестве смежных классов G/H. В частности, $(1_1)^G$ – левое регулярное представление группы G.

Указание. Рассмотрите действие g на базисе $g_i \otimes 1$. □

Упражнение 13.2. Пусть $H \leqslant K \leqslant G$. Тогда $\mathrm{ind}_K^G(\mathrm{ind}_H^K(V)) = \mathrm{ind}_H^G(V)$.

Решение. $FG \otimes_{FK} (FK \otimes_{FH} V) \cong (FG \otimes_{FK} FK) \otimes_{FH} V \cong FG \otimes_{FH} V$. □

Упражнение 13.3. Если представление π точное, то π^G тоже точное.

Решение. Предположим, что g действует тривиально на базисе $g_i \otimes u_j$. Если $gg_i \notin g_i H$, то $g \cdot (g_i \otimes u_j) \notin g_i \otimes V$. Значит $gg_i = g_i h$ для некоторого $h \in H$. $g \cdot (g_i \otimes u_j) = g_i \otimes h u_j = g_i \otimes u_j$ при всех j. Так как H действует точно, то из этого следует $h = 1$, откуда $g = 1$. □

Упражнение 13.4. Пусть U – левый FG-модуль, $U = U_1 \oplus \ldots \oplus U_n$ как векторные пространства над F, причем G переставляет U_i транзитивно. Обозначим через H стабилизатор подпространства U_1. Тогда U_1 является FH-модулем и $U \cong \mathrm{ind}_H^G(U_1)$.

Решение. Зададим гомоморфизм $FG \otimes_{FH} U_1 \to U$ на разложимых тензорах по правилу $g_i \otimes u \mapsto g_i u$ и продолжим по линейности. Ясно, что это – гомоморфизм FG-модулей. Транзитивность действия G на наборе U_i обеспечивает сюръективность отображения. Заметим, что $g_i u = g_j u' \implies g_j^{-1} g_i u = u' \in U_1 \implies g_j^{-1} g_i \in H$, что невозможно при $i \neq j$. Поэтому $\sum_{i=1}^n g_i u^{(i)} = 0 \iff g_i u^{(i)} = 0$ при всех i, откуда все $u^{(i)} = 0$. Из этого следует инъективность. □

Индуцированный характер – это характер индуцированного представления. Пусть $H \leqslant G$, π – представление H с характером χ, а π^G – индуцированное представление G с характером χ^G. Пусть, по-прежнему, $g_1, \ldots, g_n$ – представители левых смежных классов G по H. Из построения матрицы π_g^G видно, что ее след χ_g^G равен сумме следов ненулевых диагональных клеток. Клетка стоит на диагонали, если $g_i^{-1} g g_i \in H$. Положим $I(g) := \{i \in [n] \mid g_i^{-1} g g_i \in H\}$. Тогда

$$\chi^G(g) = \sum_{i \in I(g)} \chi(g_i^{-1} g g_i).$$

Пусть $s = g_i h \in g_i H$. Заметим, что $s^{-1}gs \in H \iff i \in I(g)$, причем в этом случае $s^{-1}gs = (g_i^{-1}gg_i)^h$ для $h \in H$, так что $\chi(g_i^{-1}gg_i) = \chi(s^{-1}gs)$. Поэтому

$$\chi^G(g) = \frac{1}{|H|} \sum_{s:\, s^{-1}gs \in H} \chi(s^{-1}gs).$$

Пусть $H \leqslant G$.

Теорема 13.5 (Закон взаимности Фробениуса)**.** *Для любых характеров χ на H и θ на G имеет место равенство*

$$B(\chi, \mathrm{res}_H^G(\theta))_H = B(\mathrm{ind}_H^G(\chi), \theta)_G.$$

Доказательство.

$$B(\mathrm{ind}_H^G(\chi), \theta)_G = \frac{1}{|G|} \sum_{g \in G} \mathrm{ind}_H^G(\chi)(g)\theta(g^{-1}) = \frac{1}{|H| \cdot |G|} \sum_{\substack{s,g \in G \\ s^{-1}gs \in H}} \chi(s^{-1}gs)\theta(g^{-1}).$$

Положив теперь $s^{-1}gs = h \in H$ мы можем переписать эту сумму в виде

$$\frac{1}{|H| \cdot |G|} \sum_{h \in H,\ s \in G} \chi(h)\theta(sh^{-1}s^{-1}).$$

Так как θ – функция классов на G, то $\theta(sh^{-1}s^{-1}) = \theta(h^{-1})$ не зависит от s, и, значит, эта сумма равна

$$\frac{1}{|H|} \sum_{h \in H} \chi(h)\theta(h^{-1}) = B(\chi, \mathrm{res}_H^G(\theta))_H.$$

□

Следствие 13.6. *Предположим, что представления π группы H и ρ группы G неприводимы. Тогда кратность вхождения π в $res_H^G(\rho)$ равна кратности вхождения ρ в* $\mathrm{ind}_H^G(\pi)$.

Глава 13

Теория Галуа

1. Расширения полей

Пусть $F \subseteq K$ – два поля. Говорят, что K является расширением F и пишут "пусть K/F – расширение полей". Обозначение нехорошее, потому что можно перепутать с факторгруппой аддитивных групп этих полей, но общепринятое. Поле K можно рассматривать как векторное пространство над F. Размерность этого пространства называется степенью расширения K/F и обозначается $[K : F]$. Расширение называется конечным в случае, когда его степень конечна.

Расширение K/F называется простым, если существует элемент $\theta \in K$ такой, что K – наименьшее расширение F, содержащее θ. Полем дробно-рациональных функций $F(t)$ называется поле частных кольца многочленов $F[t]$.

Лемма 1.1. *Простое расширение изоморфно либо полю дробно-рациональных функций $F(t)$, и тогда оно бесконечно, либо факторкольцу $F[t]/(p)$, где $p \in F[t]$ – неприводимый многочлен, и тогда оно конечно, и его степень равна* $\deg p$.

Доказательство. Пусть K – простое расширение F, полученное присоединением θ. Рассмотрим канонический гомоморфизм F-алгебр $\varepsilon : F[t] \to K$, посылающий t в θ. Если ядро равно нулю, то образ изоморфен $F[t]$. По универсальному свойству поля частных существует гомоморфизм $F(t) \to K$, который по определению простого расширения будет эпиморфизмом. Так как любой гомоморфизм полей является мономорфизмом, то в этом случае $F(t) \cong K$.

Пусть $\operatorname{Ker}\varepsilon \neq 0$. Так как $F[t]$ – область главных идеалов, то $\operatorname{Ker}\varepsilon = (p)$ для некоторого многочлена $p \in F[t]$. Так как $\operatorname{Im}\varepsilon$ – область целостности, то (p) – простой идеал, т. е. p неприводим. Ненулевой простой идеал является максимальным, таким образом $\operatorname{Im}\varepsilon$ – поле. По определению простого расширения $K = \operatorname{Im}\varepsilon \cong F[t]/(p)$. □

Образ гомоморфизма ε из предыдущего доказательства обозначается $F[\theta]$. Если $F[\theta]$ изоморфно кольцу многочленов, т. е. θ не является корнем многочлена с коэффициентами из F, то θ называется трансцендентным над F. В противном случае θ называется алгебраическим, а неприводимый многочлен p, корнем которого является θ – его минимальным многочленом.

Следствие 1.2. *Пусть $p \in F[t]$ – неприводимый многочлен. Если p имеет корень θ в расширении K/F, то существует единственное вложение $F[t]/(p) \to K$, отображающее $t + (p)$ в θ.*

Следующее утверждение похоже на лемму 11.2 главы 11.

Лемма 1.3. *Если $L/K/F$ – расширения полей, то $[L : F] = [L : K] \cdot [K : F]$.*

Доказательство. Если X – базис L над K, а Y – базис K над F, то $\{yx \mid y \in Y, x \in X\}$ очевидно порождает L, как векторное пространство над F. Если $\sum_{x\in X, y\in Y} \alpha_{xy} yx = 0$, где почти все коэффициенты равны 0, то $\sum_{x\in X}(\sum_{y\in Y} \alpha_{xy} y)x = 0$. Так как X – базис, то $\sum_{y\in Y} \alpha_{xy} y = 0$ для любого $x \in X$, а так как Y – базис, то все α_{xy} равны нулю. □

Следствие 1.4. *Расширение, порожденное конечным числом алгебраических элементов, является конечным.*

Любой элемент конечного расширения K/F алгебраичен над F.

Любое конечное расширение порождено конечным числом алгебраических элементов.

Доказательство. Первое утверждение является прямым следствием леммы 11.2 главы 11. Оно также сразу вытекает из предыдущей леммы.

Для любого $\alpha \in K$ его степени α^k, $k \in \mathbb{N}_0$, линейно зависимы. Их линейная зависимость – это и есть полиномиальное равенство $p(\alpha) = 0$. На старший коэффициент многочлена p можно сократить.

Присоединяя к F любой элемент $\alpha \in K \setminus F$, получим $[K : F[\alpha]] = [K : F]/[F[\alpha] : F] < [K : F]$. Далее – индукция по $[K : F]$. □

2. Поле разложения

Пусть $p \in F[t]$ – (не обязательно неприводимый) многочлен над F. Мы хотим построить универсальное расширение K/F, в котором многочлен p раскладывается на линейные множители. К сожалению, универсальности добиться не удастся, потому что не удастся соблюсти единственность. Это видно уже на примере $F = \mathbb{R}$ и $p = t^2 + 1$, так что $K = \mathbb{C}$, но у поля $\mathbb{C}$ есть автоморфизм, оставляющий $\mathbb{R}$ на месте – комплексное сопряжение. Несмотря на эту неудачу, такое расширение все же будет единственным с точностью до (неединственного) изоморфизма.

Определение 2.1. Полем разложения многочлена $p \in F[t]$ над полем F называется такое расширение K/F, что многочлен p раскладывается над K на линейные множители, и для любого другого расширения L/F, обладающего тем же свойством, существует гомоморфизм полей $K \to L$ тождественный на F.

Теорема 2.2. *Поле разложение многочлена p над F существует и единственно с точностью до изоморфизма, тождественного на F.*

Доказательство. Проведем доказательство существования индукцией по $\deg p$. База ($\deg p = 1$) очевидна. Пусть $p = fg$, где f неприводим (а g может быть равен 1). Тогда над полем $\tilde{F} = F[t]/(f)$ многочлен f имеет корень и, следовательно, раскладывается на множители $f = (t - \theta)\tilde{f}$. Таким образом, над $\tilde{F}$ имеем $p = (t - \theta)\tilde{p}$, где $\tilde{p} = \tilde{f}g$. По индукционному предположению существует поле разложение K многочлена $\tilde{p}$ над $\tilde{F}$. Это поле и будет полем разложения p над F.

Ясно, что p раскладывается над K на линейные множители. Докажем универсальное свойство поля K. Пусть L/F другое расширение, в котором p раскладывается на линейные множители. По единственности разложения на неприводимые, f должен делиться хотя бы на один линейный сомножитель, т. е. иметь корень в поле L. По следствию 1.2 существует вложение $\tilde{F} \rightarrowtail L$. Отождествляя элементы $\tilde{F}$ с их образами в L, и используя универсальное свойство поля разложения многочлена $\tilde{f}$ над $\tilde{F}$, убеждаемся, что существует вложение $K \rightarrowtail L$, что и доказывает, что K – поле разложения p над F.

По лемме 1.3, только что построенное расширение K/F имеет конечную степень, как последовательное выполнение конечного числа конечных расширений. Пусть $\hat{K}$ – другое поле разложения p над F. По определению поля разложения существуют гомоморфизмы $\varphi : K \to \hat{K}$ и $\hat{\varphi} : \hat{K} \to K$, тождественные на F. Любой гомоморфизм полей инъективен, а при инъективном отображении размерность векторного пространства не падает. Следовательно, $\dim_F K = \dim_F \hat{K}$. Инъективные отображения между векторными пространствами одинаковой размерности являются биективными, поэтому φ и $\hat{\varphi}$ – изоморфизмы (не обязательно взаимно обратные, но это и не требуется). □

Поле разложения многочлена g над полем F будет обозначаться через F_g.

Лемма 2.3. *Пусть E/F – конечное расширение, многочлен $f \in F[t]$ раскладывается в E на линейные множители, и E порождено корнями f над F. Тогда $E = F_f$.*

Доказательство. Пусть K – поле разложения f над F. Так как f раскладывается на линейные множители над E, то по определению поля разложения существует гомоморфизм $\varphi : K \to E$ над F. При этом, так как $\varphi(f) = f$, то корни f в K переходят под действием φ в корни f в E. Так как E порождено этими корнями, то φ сюръективно, следовательно, $K \cong E$. □

3. Существование и единственность конечных полей

Теорема 3.1. *Любое конечное поле содержит p^n элементов, где p – простое число, а $n \in \mathbb{N}$.*

Обратно, для любого простого числа p и натурального n существует поле из p^n элементов, и любые два таких поля изоморфны.

Доказательство. Пусть K – конечное поле, а F – аддитивная подгруппа в K, порожденная 1. Как и любая циклическая группа $F \cong \mathbb{Z}/p\mathbb{Z}$ для некоторого $p \in \mathbb{N}$. Ясно, что F замкнуто относительно умножения, следовательно, является подкольцом в K, изоморфным $\mathbb{Z}/p\mathbb{Z}$. Любое подкольцо области целостности является областью целостности, следовательно, p простое. Поэтому $F \cong \mathbb{F}_p$ – простое поле из p элементов.

Размерность $[K : F] = n$ конечна. По теореме о классификации векторных пространств K изоморфно F^n для некоторого n и, следовательно, $|K| = |F^n| = p^n$.

Пусть, теперь, $q = p^n$, а L – поле разложения многочлена $t^q - t$ над $\mathbb{F}_p$. Заметим, что $\operatorname{char} L = p$, т. е. $p = 0$ в L. Обозначим через K множество корней этого многочлена. Формальная производная $(t^q - t)' = qt^{q-1} - 1 = -1$ в поле L. Следовательно, многочлен $t^q - t$ не имеет кратных корней, поэтому $|K| = q$. Пусть $x, y \in K$. Тогда $(xy)^q = x^q y^q = xy$, то $xy \in K$, аналогично $x^{-1} \in K$ при $x \neq 0$. Так как биномиальные коэффициенты C_p^k делятся на p, то из бинома Ньютона следует, что $(x + y)^p = x^p + y^p$. Применяя эту формулу n раз, получаем $(x + y)^q = x^q + y^q = x + y$, откуда $x + y \in K$. Таким образом, K является подполем в L, содержащим все корни многочлена $t^q - t$. По определению поля разложения L вкладывается в K, откуда $L = K$.

Пусть, наконец, $\tilde{K}$ – произвольное поле из q элементов. Так как его мультипликативная группа состоит из $q - 1$ элемента, то $x^{q-1} = 1$ для любого $x \in \tilde{K}^* = \tilde{K} \setminus \{0\}$. Следовательно, все элементы поля $\tilde{K}$ являются корнями многочлена $t^q - t$. По определению поля разложения L вкладывается в $\tilde{K}$, а в предыдущем абзаце мы показали, что $|L| = q$. Следовательно, $\tilde{K} \cong L$. □

4. Алгебраическое замыкание

Следующее утверждение фактически уже было доказано в главе 4, см. также лемму 3.3 главы 12.

Предложение 4.1. *Пусть K – поле. Следующие условия эквивалентны.*

(1) *Каждый многочлен в $K[t]$ раскладывается на линейные множители.*
(2) *Каждый многочлен положительной степени из $K[t]$ имеет хотя бы один корень в K.*
(3) *Любой неприводимый многочлен в $K[t]$ имеет степень 1.*
(4) *Любое конечное расширение K совпадает с K.*

Поле, удовлетворяющее условиям предложения, называется алгебраически замкнутым.

Пусть F – произвольное поле. Алгебраически замкнутое поле K, содержащее F, каждый элемент которого алгебраичен над F, называется алгебраическим замыканием F. В этом параграфе мы докажем, что алгебраическое замыкание существует и единственно с точностью до (неединственного) изоморфизма.

Лемма 4.2. *Пусть L – расширение поля F. Множество K всех элементов поля L, алгебраических над F, является полем.*

Доказательство. Пусть $\alpha, \beta \in K$. По следствию 1.4 $F[\alpha, \beta]$ – конечное расширение, и, следовательно, любой его элемент алгебраичен над F. В частности, $\alpha + \beta$, $\alpha\beta$ и α/β лежат в K, следовательно, K – поле. □

Лемма 4.3. *Пусть L – расширение поля F, в котором каждый многочлен из $F[t]$ раскладывается на линейные множители. Пусть K – подмножество в L, состоящее из всех элементов L, алгебраических над F. Тогда K является алгебраическим замыканием F.*

Поле K называется алгебраическим замыканием F в L.

Доказательство. Пусть $g(t) = t^n + \alpha_{n-1}t^{n-1} + \cdots + \alpha_0$ – неприводимый многочлен с коэффициентами из K. Положим $\tilde{K} = F[\alpha_0, \ldots, \alpha_{n-1}]$. Это поле является конечным расширением F, поэтому $\tilde{K}[t]/(g)$ также является конечным расширением F. Обозначим через θ корень многочлена g в $\tilde{K}[t]/(g)$. Как и любой элемент конечного расширения, θ алгебраический над F. Пусть $f \in F[t]$ – его минимальный многочлен. Ясно, что f делится на g в $\tilde{K}[t]$. По универсальному свойству поля разложения F_f вкладывается в L. Обозначим его образ через E. Будучи конечным расширением F, поле E содержится в K. Так как g делит f, а f раскладывается на линейные множители в K, то и g раскладывается на линейные множители, следовательно, K алгебраически замкнуто. □

Предложение 4.4. *Пусть $\tilde{F}$ – алгебраическое расширение F, а L – такое расширение F, в котором каждый многочлен из $F[t]$ раскладывается на линейные множители. Тогда существует вложение $\tilde{F}$ в L над F. Если $\tilde{F} = \overline{F}$ – алгебраическое замыкание F, то образ этого вложения совпадает с алгебраическим замыканием F в L.*

Доказательство. Обозначим через K алгебраическое замыкание F в L и предположим, что вложение $\tilde{F}$ в L над F существует. Так как $\tilde{F}$ состоит из алгебраических элементов, то его образ также состоит из алгебраических элементов, т. е. содержится в K. Если $\tilde{F}$ алгебраически замкнуто, то и его образ замкнут, т. е. любой многочлен из $F[t]$ раскладывается на линейные множители в образе. Следовательно, все алгебраические над F элементы лежат в образе.

Пусть $\mathscr{X}$ – множество состоящее из пар (E, φ), где E – подполе в $\tilde{F}$, содержащее F, а φ – вложение $E \to L$. Зададим на $\mathscr{X}$ структуру частично упорядоченного множества, положив $(E, \varphi) < (E', \varphi')$, если $E \subseteq E'$, а $\varphi = \varphi'|_E$. Если $\mathscr{Y}$ – линейно упорядоченное подмножество в $\mathscr{X}$, положим $\hat{E} = \bigcup_{(E,\varphi)\in\mathscr{Y}} E$, и определим $\hat{\varphi} : \hat{E} \to L$ как отображение, продолжающее $\varphi : E \to L$ по всем $(E, \varphi) \in \mathscr{Y}$. Ясно, что $(\hat{E}, \hat{\varphi})$ является верхней гранью $\mathscr{Y}$. По лемме Цорна в $\mathscr{X}$ существует максимальный элемент, который мы обозначим через $\hat{F}$. Если существует $\theta \in \tilde{F} \setminus \hat{F}$, то θ является корнем неприводимого многочлена $g \in \hat{F}[t]$. Так как K алгебраически замкнуто, то $\hat{\varphi}(g)$ имеет корень $\hat{\theta} \in K$. Тогда $\hat{F}[\theta] \cong \hat{F}[t]/(g)$ вкладывается в K так, что на $\hat{F}$ это вложение совпадает с $\hat{\varphi}$, а θ (или $t + (g)$) отображается в $\hat{\theta}$. Это противоречит максимальности $\hat{\varphi}$. Противоречие показывает, что $\hat{F} = \overline{F}$. □

В следующем утверждении предполагается существование алгебраического замыкания, факт, который мы скоро докажем.

Следствие 4.5. *Любое алгебраическое расширение вкладывается в алгебраическое замыкание.*

Следствие показывает, что алгебраическое замыкание – это самое большое алгебраическое расширение данного поля. Интуитивно алгебраическое замыкание так и надо воспринимать.

Для доказательства существования алгебраического замыкания нам понадобится существование F-алгебры, в которую отображается данный (бесконечный) набор F-алгебр. Универсальная конструкция, т. е. копроизведение данного набора, выглядит наиболее естественно, поэтому мы ее и приведем. Существование конечных копроизведений в категории

F-алгебр мы уже знаем – это тензорные произведения. Существование бесконечных можно было бы вывести из этого доказав существование инъективных пределов. Вместо этого мы дадим явную конструкцию.

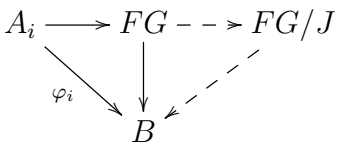
Пусть F – коммутативное кольцо с 1, а A_i – F-алгебры, где i пробегает некоторое множество индексов I. Рассмотрим копроизведение мультипликативных моноидов $G = \coprod\limits_{i\in I} A_i$ в категории коммутативных моноидов. Это – множество функций $x : I \to \cup A_i$, для которых $x(i) \in A_i$ и почти все значения равны 1, с поточечным умножением. При этом отображение $\psi_i : A_i \to G$ строится очевидным образом: $\psi_i(a) = x_a^{(i)}$, где $x_a^{(i)}(i) = a$, а $x_a^{(i)}(k) = 1$ при всех $k \neq i$.

Лемма 4.6. *Моноид G с указанными гомоморфизмами действительно является копроизведением A_i в категории коммутативных моноидов.*

Доказательство. Если $\varphi_i : A_i \to N$ – гомоморфизмы моноидов, то положим $\varphi(x) = \prod_{i\in I} \varphi_i\big(x(i)\big)$ (произведение определено, так как почти все сомножители равны 1). Ясно, что $\varphi_i = \varphi \circ \psi_i$. С другой стороны, равенство $\varphi_i(a) = \varphi(x_a^{(i)})$ однозначно задает значения φ на функциях, отличающихся от 1 только одним значением, а такие функции порождают G. □

В полугрупповой алгебре FG рассмотрим идеал J, порожденный всеми элементами $x_a^{(i)} + x_b^{(i)} - x_{a+b}^{(i)}$ и $x_{\alpha a}^{(i)} - \alpha x_a^{(i)}$, где $i \in I$, $a, b \in A_i$, а $\alpha \in F$. Положим $A = \bigotimes_F A_i := FG/J$ и докажем, что это и есть копроизведение A_i в категории F-алгебр. Нетрудно проверить, что композиции отображений $A_i \xrightarrow{\psi_i} G \to FG \to A$ являются гомоморфизмами F-алгебр. С другой стороны, отображения F-алгебр $\varphi_i : A_i \to B$ являются в частности гомоморфизмами моноидов. Поэтому они единственным образом пропускаются через моноид G. По универсальному свойству полугрупповой алгебры любое отображение моноида G в F-алгебру единственным образом пропускается через FG (полугрупповая алгебра – сопряженный функтор к забывающему из F-алгебр в мультипликативные моноиды).

Далее, из коммутативности левого треугольника



следует, что образующие идеала J лежат в ядре отображения $FG \to B$, поэтому это отображение пропускается единственным образом через FG/J, как показано на диаграмме. Таким образом, мы доказали следующее утверждение.

Предложение 4.7. *Построенная выше алгебра $\bigotimes_F A_i$ не равна нулю и является копроизведением A_i по $i \in I$ в категории F-алгебр.*

Доказательство. То, что эта алгебра является копроизведением, мы уже доказали. Если она равна нулю, то идеал J алгебры FG совпадает со всей алгеброй. Но тогда существует конечное число образующих этого идеала, линейная комбинация которых равна 1. В этой линейной комбинации задействовано только конечное подмножество $I' \subseteq I$. Это означает, что $J \cap F \coprod\limits_{i\in I'} G_i = F \coprod\limits_{i\in I'} G_i$, откуда $\bigotimes\limits_{i\in I'} A_i = 0$. Но конечное тензорное произведение ненулевых векторных пространств не равно нулю. Противоречие показывает, что $FG/J \neq 0$. □

Теорема 4.8. *Для любого поля F существует его алгебраическое замыкание. Два разных алгебраических замыкания изоморфны между собой.*

Доказательство. Единственность сразу следует из предложения 4.4. Действительно, если L – алгебраическое замыкание, то $K = L$ и, следовательно, построенное вложение $\overline{F}$ в L является изоморфизмом.

Для доказательства существования по лемме 4.3 достаточно доказать существование такого расширения поля F, в котором каждый многочлен из $F[t]$ раскладывается на линейные множители. Пусть P – множество всех многочленов из $F[t]$. Положим $R = \bigotimes_{f \in P} F_f$. Пусть M – максимальный идеал в R, тогда R/M – поле. По построению существует гомоморфизм $F_f \to R \to R/M$. Любой гомоморфизм полей инъективен, поэтому мы можем отождествить F_f с его образом в R/M. Так как f раскладывается на линейные множители в F_f он раскладывается и в R/M, что завершает доказательство. □

5. Типы расширений

Любое расширение полей является композицией чисто трансцендентного и алгебраического расширений. Алгебраическое расширение, в свою очередь, является композицией сепарабельного и чисто несепарабельного расширений. Оба этих факта мы сформулируем в этом параграфе с намеком на доказательство. Начнем с основных определений и некоторых свойств трансцендентных и (алгебраических) сепарабельных расширений.

Чисто трансцендентное расширение E/F – это расширение в котором каждый элемент из $E \setminus F$ трансцендентен над F. Такое расширение устроено просто: это поле частных кольца многочленов над исходным полем. Элементы $\alpha_1, \ldots, \alpha_n \in E$ называются алгебраически зависимыми над F, если существует ненулевой многочлен $f \in F[t_1, \ldots, t_n]$ такой, что $f(\alpha_1, \ldots, \alpha_n) = 0$. Максимальное алгебраически независимое над F подмножество в E называется базисом трансцендентности, а его мощность (которая не зависит от выбора базиса) – степенью трансцендентности E над F. Доказательство существования базиса похоже на доказательство аналогичного факта для векторных пространств.

Теорема 5.1. *Пусть E/F – расширение полей, а $S \subseteq E$ – базис трансцендентности E над F. Тогда E является алгебраическим расширением поля $F(S)$, где $F(S)$ обозначает поле рациональных функций, т. е. поле частных кольца многочленов $F[S]$.*

Заметим, что не только базис трансцендентности S, но и поле $F(S)$ определено неоднозначно. Однако, мощности всех базисов трансцендентности данного расширения одинаковы. Эта мощность называется степенью трансцендентности. Доказательство последнего факта выходит за рамки нашего курса.

Многочлен называется сепарабельным, если он не имеет кратных корней над замыканием исходного поля. Пусть E/F – алгебраическое расширение полей. Элемент $\theta \in E$ называется сепарабельным над F, если его минимальный многочлен сепарабелен.

Расширение E/F называется *сепарабельным*, если все элементы E сепрабельны над F.

Пример. Возведение в степень p всегда сохраняет (коммутативное) умножение, а в характеристике p оно сохраняет еще и сложение. Таким образом, если $\operatorname{char} K = p$, то отображение $\varphi : K \to K$, $\varphi(\alpha) = \alpha^p$ является эндоморфизмом поля K. Он называется эндоморфизмом Фробениуса.

Пусть $E = \mathbb{F}_p(x)$, а $F = \varphi(E) = \mathbb{F}_p(x^p)$ – подполе в E. Многочлен $(t-x)^p = t^p - x^p \in F[t]$ имеет кратный корень в E и неприводим над F. Действительно, любой нетривиальный сомножитель этого многочлена имеет вид $(t-x)^k$, где $0 < k < p$, но даже свободный член этого сомножителя x^k не лежит в F.

Предложение 5.2. *Пусть $F \subseteq E$ – поля характеристики p. Следующие условия эквивалентны*

(1) *Расширение E/F несепарабельно.*
(2) *Существует неприводимый несепарабельный многочлен из $F[t]$, имеющий корень в E.*
(3) *Существует неприводимый многочлен из $F[t]$, имеющий корень в E, производная которого равна нулю.*
(4) *Существует многочлен $g \in F[t]$ такой, что $g(t^p)$ неприводим и имеет корень в E.*

(5) *Существует неприводимый многочлен из $F[t]$ имеющий кратный корень в E.*

Доказательство. (1) $\Longrightarrow$ (2). Если $\alpha \in E$ несепарабельный элемент, то его минимальный многочлен неприводим, имеет корень в E и является несепарабельным.

(2) $\Longrightarrow$ (3). Пусть f – неприводимый несепарабельный многочлен. Так как он имеет кратные корни в алгебраическом замыкании $\overline{F}$ поля F, то $\gcd(f, f') \neq 1$ в $\overline{F}[t]$. Но наибольший общий делитель можно вычислить при помощи алгоритма Евклида, который не выводит за пределы исходного поля F. Значит и над F эти многочлены не взаимно просты. Так как f неприводим, то $\gcd(f, f') = f$, а так как $\deg f' < \deg f$, то это возможно лишь в случае $f' = 0$.

(3) $\Longrightarrow$ (4). Если производная многочлена равна нулю, то степени всех его одночленов должны делиться на характеристику поля.

(4) $\Longrightarrow$ (5). Пусть $\theta \in E$ – корень неприводимого над F многочлена $h(t) = g(t^p)$. Тогда $h(\theta) = g(\theta^p) = 0$, откуда $g(y)$ делится на $y - \theta^p$, а $g(t^p)$ делится на $t^p - \theta^p = (t - \theta)^p$. Таким образом, $g(t^p)$ неприводим над F и имеет кратный корень в E.

Импликация (5) $\Longrightarrow$ (1) очевидна. □

Поле называется совершенным, если любое его алгебраическое расширение сепарабельно. Из предыдущего предложения вытекает, что поле F совершенно тогда и только тогда, когда любой неприводимый многочлен из $F[t]$ является сепарабельным.

Ясно, что алгебраически замкнутое поле совершенно. Из предыдущего предложения следует, что любое поле характеристики 0 совершенно. Немного сложнее понять, что любое конечное поле совершенно. Пусть $F = \mathbb{F}_q$, а E – его алгебраическое расширение. Пусть $f \in F[t]$ – минимальный многочлен элемента $\theta \in E$. Поле $F[\theta]$ конечно, скажем $|F[\theta]| = q^n$. Тогда θ является корнем многочлена $t^{q^n} - t$, который не имеет кратных корней, так как его производная равна -1. Минимальный многочлен делит любой многочлен, аннулирующий θ. Поэтому f делит $t^{q^n} - t$ и, следовательно, не имеет кратных корней.

Приведем без доказательства несколько свойств сепарабельных расширений. Один из возможных способов доказательства этих свойств – подсчет числа вложений E над F в алгебраическое замыкание поля F, см. упражнение 6.2.

Предложение 5.3. *Сепарабельное расширение сепарабельного расширения сепарабельно.*

Присоединение сепарабельного элемента – сепарабельное расширение. Другими словами, если $f \in F[t]$ – неприводимый сепарабельный многочлен, то $F[t]/(f)$ – сепарабельное расширение F.

Из последнего утверждения вытекает, что множество всех элементов поля E, сепарабельных над подполем $F \subseteq E$, является подполем в E. Назовем это подполе сепарабельным замыканием F в E. Сепарабельным замыканием F (без указания, в каком поле) называется сепарабельное замыкание F в его алгебраическом замыкании. Оно обозначается через F^{sep}. Поле называется сепарабельно замкнутым, если оно совпадает со своим сепарабельным замыканием.

Нетрудно доказать что:

Предложение 5.4.
- *Поле E является сепарабельным замыканием F, если все элементы E алгебраические и сепарабельные над F, и любой сепарабельный многочлен из $F[t]$ раскладывается над E на линейные множители;*
- *любое алгебраическое сепарабельное расширение вкладывается в сепарабельное замыкание;*
- *если любой сепарабельный многочлен из $F[t]$ раскладывается на линейные множители в E, то сепарабельное замыкание F вкладывается в E;*
- *сепарабельное замыкание существует и единственно с точностью до изоморфизма.*
- *сепарабельное замыкание сепарабельно замкнуто.*

Алгебраическое расширение E/F называется чисто несепарабельным, если любой элемент $E \setminus F$ несепарабелен над F.

Теорема 5.5. *Любое алгебраическое расширение является композицией сепарабельного и чисто несепарабельного.*

Точнее, пусть E/F – алгебраическое расширение, а $\hat{F}$ – сепарабельное замыкание F в E. Тогда $E/\hat{F}$ является чисто несепарабельным.

Расширение E/F называется нормальным, если любой неприводимый многочлен из $F[t]$, имеющий корень в E, раскладывается в $E[t]$ на линейные множители. Конечное нормальное сепарабельное расширение называется расширением Галуа. Другими словами, конечное расширение является расширением Галуа, если минимальный многочлен любого элемента из E раскладывается в $E[t]$ на *различные* линейные множители. Обозначим через $\mathrm{Aut}(E/F)$ группу автоморфизмов поля E, оставляющих на месте все элементы F. Если E/F – расширение Галуа, то $\mathrm{Aut}(E/F)$ называется группой Галуа E над F и обозначается через $\mathrm{Gal}(E/F)$. Группой Галуа сепарабельного многочлена называется группа Галуа его поля разложения. Группой Галуа элемента называется группа Галуа его минимального многочлена.

Изначально, мотивировкой для изучения группы Галуа был тот факт (который мы докажем чуть позже), что разрешимость уравнения $p(t) = 0$ в радикалах равносильна разрешимости группы Галуа этого многочлена.

6. Группы автоморфизмов

Теорема 6.1. *Пусть $\varphi : F \to \tilde{F}$ – изоморфизм полей, $f \in F[t]$, $\tilde{f} = \varphi(f)$, E порождено (не обязательно всеми) корнями f над F, а K – расширение $\tilde{F}$, над которым $\tilde{f}$ раскладывается на линейные множители. Тогда количество гомоморфизмов $E \to K$, продолжающих φ, больше нуля, и не превосходит $[E : F]$.*

Предположим, что f не делится на квадрат неприводимого многочлена из $F[t]$, а любой неприводимый сомножитель f имеет корень в E. Тогда количество гомоморфизмов равно $[E : F]$ $\iff$ f сепарабелен.

Доказательство. Пусть $f = \prod f_i^{k_i}$, где $f_i \in F[t]$ неприводимы. Ясно, что условия теоремы не поменяются, если заменить f на $\prod f_i$, где каждый f_i имеет хотя бы один корень в E. Так что можно с самого начала считать, что f свободно от квадратов и от "лишних" неприводимых сомножителей.

Проведем индукцию по $[E : F]$. Если $[E : F] = 1$, то $E = F$ и все очевидно. Если f сепарабелен, то все f_i сепарабельны. Так как $E \neq F$, то можно считать, что $\deg f_1 > 1$. В противном случае, один из f_i несепарабелен. Для определенности будем считать, что это f_1 (и в этом случае также $\deg f_1 > 1$). Так как f_1 делит f, то $\tilde{f}_1 := \varphi(f_1)$ делит $\tilde{f}$, и поэтому $\tilde{f}_1$ раскладывается над K на линейные множители. Пусть α – корень f_1 в E. По следствию 1.2 для каждого корня $\tilde{\alpha}$ многочлена $\tilde{f}_1$ в K существует единственное вложение $\varphi_1 : F[\alpha] \cong F[t]/(f_1) \cong \tilde{F}[t]/(\tilde{f}_1) \to K$, продолжающее φ и посылающее α_1 в $\tilde{\alpha}$. Поэтому количество различных вложений $F[\alpha]$ в K равно количеству различных корней многочлена $\tilde{f}_1$ в K, которое не превосходит $\deg \tilde{f}_1 = \deg f_1 = [F[\alpha_1] : F]$, и равно ей тогда и только тогда, когда f_1 сепарабелен. Обозначим это количество через m_1.

Пусть $L = F[\alpha]$, а $\tilde{L} := \varphi_1(L) = \tilde{F}[\tilde{\alpha}]$. Заметим, что все условия теоремы выполнены с заменой F на L. Так как $[E : L] = [E : F]/[L : F] < [E : F]$, то можно воспользоваться индукционным предположением, которое утверждает, в частности, что количество вложений $E \to K$, продолжающих φ_1, не превосходит $[E : L]$, а если f сепарабелен, то равно $[E : L]$. Обозначим это количество через k.

Таким образом, для каждого из m_1 вложений $\varphi_1 : L \to K$ над φ существует k вложений E в K над φ_1. В результате, количество вложений E в K над φ равно $m_1 k \leqslant [K : F] \cdot [E : K] = [E : F]$. Если f сепарабелен, то имеет место равенство, в противном случае, $m_1 < [K : F]$ и $m_1 k < [K : F] \cdot [E : K] = [E : F]$. □

Упражнение 6.2. Докажите предложение 5.3.

Следствие 6.3. *Пусть E, K – расширения поля F, причем $[E : F]$ конечно. Тогда количество вложений E в K над F не превосходит $[E : F]$, и существует расширение L/K такое, что E вкладывается в L над F. Если K/F – нормальное сепарабельное расширение, то количество вложений E в K над F равно $[E : F]$ или нулю.*

Доказательство. Пусть $E = F[\alpha_1, \dots, \alpha_n]$, $f_i \in F[t]$ – минимальные многочлены элементов α_i, f – произведение различных f_i, а L – поле разложения f над K. По предыдущей теореме существует хотя бы одно, но не более $[E : F]$ вложений E в L над F. Любое вложение E в K можно рассматривать, как вложение E в L, так что их количество также не превосходит $[E : F]$.

Если E вкладывается в K, то все многочлены f_i имеют корни в K. Если K/F – нормально, то все f_i, а значит и f, раскладываются в K на линейные множители. Если же K/F сепарабельно, то многочлены f_i сепарабельны, а значит сепарабелен и f (различные из многочленов f_i попарно взаимно просты, так как они неприводимы, и, следовательно, не имеют общих корней). Снова по предыдущей теореме в этом случае количество вложений E в K над F равно $[E : F]$. □

Следствие 6.4. *Пусть E – поле разложения многочлена $f \in F[t]$. Тогда $|\operatorname{Aut}(E/F)| \leqslant [E : F]$, причем, если f сепарабелен, то достигается равенство.*

Пусть G – подгруппа группы автоморфизмов поля E. Обозначим множество неподвижных элементов под действием G через E^G. Легко проверить, что E^G – поле.

Теорема 6.5 (Лемма Артина)**.** *Пусть G – конечная подгруппа группы автоморфизмов поля E, а $F = E^G$. Тогда $[E : F] \leqslant |G|$.*

Доказательство. Пусть $G = \{\varphi_1 = \mathrm{id}, \dots, \varphi_n\}$, а $\alpha_1, \dots, \alpha_m \in E$. Докажем, что при $m > n$ эти элементы линейно зависимы над F. Однородная система линейных уравнений $\sum_{i=1}^m \varphi_j(\alpha_i)x_i = 0$, $j = 1, \dots, n$ имеет в E ненулевое решение, так как количество неизвестных больше количества уравнений (теорема о размерности ядра и образа). Пусть $(\beta_1, \dots, \beta_m)$ – ненулевое решение с наименьшим количеством ненулевых элементов. Перенумеровывая и умножая на константу можно считать, что $\beta_1 = 1$. Если $\beta_2, \dots, \beta_m \in F$, то при $j = 1$ получаем линейную зависимость элементов α_i. В противном случае $\beta_k \notin F$, т. е. $\varphi_\ell(\beta_k) \neq \beta_k$ для некоторых k, ℓ. Применяя φ_ℓ ко всем равенствам системы получим: $\sum_{i=1}^m \varphi_\ell \circ \varphi_j(\alpha_i)\varphi_\ell(\beta_i) = 0$, или, после перенумерации уравнений, $\sum_{i=1}^m \varphi_j(\alpha_i)\varphi_\ell(\beta_i) = 0$. Таким образом, $(\varphi_\ell(\beta_1) = 1, \dots, \varphi_\ell(\beta_m))$ – является решением системы. Но тогда и разность $(0, \varphi_\ell(\beta_2) - \beta_2, \dots, \varphi_\ell(\beta_m) - \beta_m)$ также является решением, причем количество нулевых элементов увеличилось, а так как $\varphi_\ell(\beta_k) - \beta_k \neq 0$, то оно нетривиально. □

Следствие 6.6. *Пусть G – конечная группа автоморфизмов поля E. Тогда $G = \operatorname{Aut}(E/E^G)$.*

Доказательство. $[E : E^G] \leqslant |G| \leqslant |\operatorname{Aut}(E/E^G)| \leqslant [E : E^G]$. Здесь первое равенство следует из леммы Артина, второе из того, что $G \leqslant \operatorname{Aut}(E/E^G)$, а третье – из следствия 6.3. □

Теорема 6.7. *Следующие условия на расширение E/F эквивалентны.*

(1) *E/F – расширение Галуа.*
(2) *E – поле разложения некоторого сепарабельного многочлена $f \in F[t]$.*
(3) *E конечно над F и $F = E^{\operatorname{Aut}(E/F)}$.*
(4) *$F = E^G$ для некоторой конечной группы $G \leqslant \operatorname{Aut}(E)$.*

Доказательство. (1) $\Longrightarrow$ (2). Так как E/F конечно, то $E = F[\alpha_1 \dots, \alpha_n]$. Пусть f_i – минимальный многочлен α_i над F. Так как расширение нормально, то f_i раскладывается в E на линейные множители. Следовательно, произведение f различных f_i также раскладывается на линейные множители, а, значит, E является полем разложения f. Так как расширение сепарабельно, то f сепарабелен.

(2) $\Longrightarrow$ (3). Ясно, что расширение конечно, а $F \subseteq \tilde{F} := E^{\mathrm{Aut}(E/F)}$. Следовательно, E является также полем разложения f над $\tilde{F}$. Любой автоморфизм E над F оставляет на месте $\tilde{F}$ и наоборот, т. е. $\mathrm{Aut}(E/F) = \mathrm{Aut}(E/\tilde{F})$. Так как f сепарабелен, то по следствию 6.4 $[E : F] = |\mathrm{Aut}(E/F)| = |\mathrm{Aut}(E/\tilde{F})| = [E : \tilde{F}]$, откуда $[\tilde{F} : F] = 1$ и $\tilde{F} = F$.

(3) $\Longrightarrow$ (4). По следствию 6.3 $[E : F] < \infty \Longrightarrow |\mathrm{Aut}(E/F)| < \infty$, и можно взять $G = \mathrm{Aut}(E/F)$.

(4) $\Longrightarrow$ (1). По лемме Артина $[E : F] \leqslant |G| < \infty$. Пусть $\alpha \in E$, а f – минимальный многочлен этого элемента. Осталось показать, что f раскладывается в E на различные линейные множители. Пусть $\{\alpha_1 = \alpha, \alpha_2, \ldots, \alpha_m\} = G\alpha$ – орбита α под действием G. Положим $g = \prod_{i=1}^{m}(t - \alpha_i) = \sum_{k=0}^{m} \beta_k t^k$. Коэффициенты $\beta_k = \pm s_k(\alpha_1, \ldots, \alpha_m)$ (здесь s_k – основные симметрические многочлены). Так как элементы группы G переставляют α_i, то $\sigma(\beta_k) = \beta_k$, следовательно, $\beta_k \in F$, т. е. $g \in F[t]$. По определению минимального многочлена f делит g и, следовательно, раскладывается в $E[t]$ на различные линейные множители. □

Следствие 6.8. *Любое конечное расширение, порожденное сепарабельными элементами, вкладывается в расширение Галуа.*

Доказательство. Пусть $E = F[\alpha_1, \ldots, \alpha_n]$, и минимальные многочлены f_i всех элементов α_i сепарабельны. Так как минимальные многочлены неприводимы, то любые 2 различных минимальных многочлена взаимно просты, следовательно, они не имеют общих корней ни над каким расширением. Тогда, сепарабельно и произведение f всех различных многочленов f_i. По теореме 6.7 поле разложение многочлена f является нормальным расширением F, а по теореме 6.1 E вкладывается в поле разложения f. □

Пусть E/F – расширение полей. Если $F \subseteq K \subseteq E$, то K/F называется подрасширением в E/F.

Следствие 6.9. *Пусть E/F – расширение Галуа, а K/F подрасширение в E/F. Тогда E/K – расширение Галуа.*

Доказательство. По теореме 6.7 E является полем разложения некоторого сепарабельного многочлена $f \in F[t] \subseteq K[t]$. Но тогда поле разложения f над K тоже равно E. □

7. Соответствие Галуа

Теорема 7.1 (основная теорема теории Галуа)**.** *Пусть E/F – расширение Галуа с группой $G = \mathrm{Gal}(E/F)$. Тогда отображения $H \mapsto E^H$ и $K \mapsto \mathrm{Gal}(E/K)$ являются взаимно обратными биекциями между множествами подгрупп $H \leqslant G$ и подрасширений K/F в E/F. Более того:*

(1) $H_1 \leqslant H_2 \iff E^{H_1} \supseteq E^{H_2}$;
(2) $|H_2 : H_1| = [E^{H_1} : E^{H_2}]$;
(3) *для $\sigma \in G$ подгруппа $\sigma H \sigma^{-1}$ соответствует подполю $\sigma(E^H)$;*
(4) *$H \triangleleft G$ тогда и только тогда, когда E^H/F является расширением Галуа; и в этом случае* $\mathrm{Gal}(E^H/F) \cong G/H$.

Доказательство. По следствию 6.9 E/K – расширение Галуа для любого промежуточного поля K. По пункту (3) теоремы 6.7 $K = E^{\mathrm{Gal}(E/K)}$. Композиция в обратном порядке тождественна по следствию 6.6.

(1). Очевидно, что
$H_1 \leqslant H_2 \Longrightarrow E^{H_1} \supseteq E^{H_2} \Longrightarrow \mathrm{Gal}(E/E^{H_1}) \leqslant \mathrm{Gal}(E/E^{H_2}) \iff H_1 \leqslant H_2$.

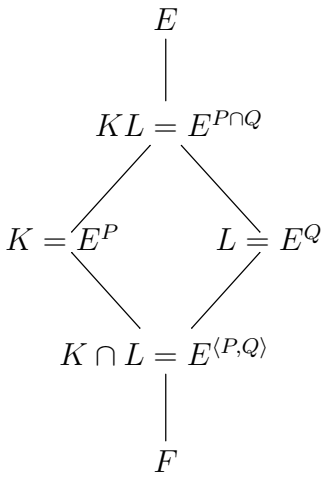
(2). $|H_2 : H_1| = \dfrac{|H_2|}{|H_1|} = \dfrac{[E : E^{H_2}]}{[E : E^{H_1}]} = [E^{H_1} : E^{H_2}]$ (здесь среднее равенство – это следствие 6.4).

(3). Для $\tau \in G$ и $\alpha \in E$ имеем $\tau(\alpha) = \alpha \iff \sigma\tau\sigma^{-1}(\sigma\alpha) = \sigma\alpha$. Поэтому $\alpha \in E^H \iff \sigma\alpha \in E^{\sigma H \sigma^{-1}}$.

(4). Пусть H нормальна в G. По пункту (3) $\sigma(E^H) = E^{\sigma H \sigma^{-1}} = E^H$ для любого $\sigma \in G$. Имеем гомоморфизм $G \to \mathrm{Aut}(E^H/F)$, $\sigma \mapsto \sigma|_{E^H}$. Ядро этого гомоморфизма равно $\mathrm{Gal}(E/E^H) = H$, поэтому G/H отождествляется с подгруппой в $\mathrm{Aut}(E^H/F)$. С другой стороны, $(E^H)^{G/H} = E^G = F$. По теореме 6.7 расширение E^H/F нормально и $G/H = \mathrm{Gal}(E^H/F)$.

Обратно, пусть E^H – нормальное расширение H, а $\alpha \in E^H$. Тогда $\sigma(\alpha)$ является корнем минимального многочлена элемента α над F и, следовательно, лежит в E^H. Таким образом, для любого $\sigma \in G$ имеем $\sigma(E^H) = E^H$, откуда по пункту (3) $E^{\sigma H \sigma^{-1}} = E^H$. Так как соответствие $H \mapsto E^H$ взаимно однозначно, получаем $\sigma H \sigma^{-1} = H$, т. е. H нормальна. □

Если K и L содержатся в поле E, то наименьшее подполе в E, содержащее K и L, обозначается KL и называется композитом полей K и L. Пусть E/F – расширение Галуа, а $P, Q \leqslant G = \mathrm{Gal}(E/F)$. Как видно из теоремы 7.1, соответствие Галуа является антиизоморфизмом решеток[1] подгрупп в G и промежуточных подполей между F и E. При этом соответствии наименьшая верхняя грань переходит в наибольшую нижнюю и наоборот. Поэтому можно рассмотреть башню полей:



Предложение 7.2. *Если расширение $K/K \cap L$ является расширением Галуа, то и KL/L – Галуа, причем* $\mathrm{Gal}(K/K \cap L) \cong \mathrm{Gal}(KL/L)$.

Доказательство. Если $K/K \cap L$ является расширением Галуа, то P нормальна в $\langle P, Q\rangle$, в частности, Q нормализует P. Следовательно, $\langle P, Q\rangle = PQ$. Кроме того, $P \cap Q$ нормальна в Q, а это равносильно тому, что KL/L – расширение Галуа. Наконец, по второй теореме о гомоморфизме

$$\mathrm{Gal}(KL/L) = \mathrm{Gal}(E^{P\cap Q}/E^Q) \cong \frac{Q}{P \cap Q} \cong \frac{PQ}{P} \cong \mathrm{Gal}(E^P/E^{PQ}) = \mathrm{Gal}(K/K \cap L).$$

□

8. Приложения теории Галуа

Доказательство основной теоремы алгебры. Любое конечное расширение $K/\mathbb{C}$ является конечным над $\mathbb{R}$. Поэтому оно вкладывается в расширение Галуа $E/\mathbb{R}$ (в характеристике 0 все сепарабельно). Пусть $G = \mathrm{Gal}(E/\mathbb{R})$, P – силовская 2-подгруппа в G, а $F = E^P$. Тогда степень расширения $F/\mathbb{R}$ нечетна, следовательно, минимальный многочлен любого элемента $\alpha \in F$ имеет нечетную степень. Но неприводимый многочлен над $\mathbb{R}$ не может иметь нечетную степень больше 1. Поэтому $F = \mathbb{R}$, а $P = G$, т. е. G является 2-группой.

[1] Решетка, это частично упорядоченное множество, в котором у любых двух элементов есть супремум и инфимум.

Тогда и $\tilde{G} = \mathrm{Gal}(E/\mathbb{C})$ является 2-группой. Предположим, что $\tilde{G} \neq 1$. Любая 2-группа разрешима, поэтому $\tilde{G} \neq [\tilde{G}, \tilde{G}]$. По теореме о классификации конечнопорожденных абелевых групп, в группе $\tilde{G}/[\tilde{G}, \tilde{G}]$ существует подгруппа индекса 2. Ее прообраз $H \leqslant \tilde{G}$ также имеет индекс 2. Следовательно, $E^H/\mathbb{C}$ – расширение степени 2, но таких не бывает, потому что любой квадратный многочлен имеет корень в $\mathbb{C}$. □

Пусть E/F – расширение Галуа, а $G = \mathrm{Gal}(E/F)$. Действие элементов группы G на поле E задает отображение $\pi : G \to \mathrm{End}_F(E)$. Здесь $\mathrm{End}_F(E)$ обозначает множество эндоморфизмов E, как векторного пространства над F, которое является векторным пространством над E относительно поточечных операций. Если $[E : F] = n$, то нетрудно видеть, что $\dim_F \mathrm{End}_F(E) = n^2$, а $\dim_E \mathrm{End}_F(E) = n$. Действительно, для любого $\varphi \in \mathrm{End}_F(E)$ подпространство $E\varphi$ одномерно над E и n-мерно над F, поэтому размерность над F в n раз больше размерности над E. По следствию 6.3 $|G| = n$, так что $\pi(G)$ может быть базисом $\mathrm{End}_F(E)$ над E. На самом деле мы докажем сейчас гораздо больше, а именно линейную независимость любого набора гомоморфизмов $H \to E^*$, а потом применим это к набору гомоморфизмов $\pi(g) : E^* \to E^*$, где g пробегает G.

Заметим, что такие отображения являются одномерными характерами группы H над E. Напомним, что для конечной группы H при некотором ограничении на характеристику E мы доказывали ортогональность неприводимых характеров.

Теорема 8.1 (Дирихле). *Пусть E – поле, $S = \{\sigma_1, \dots, \sigma_n\}$ – набор различных гомоморфизмов группы H в E^* (одномерных представлений группы H). Тогда S линейно независимо над E.*

Доказательство. Индукция по n. База $n = 1$ очевидна. Пусть $\sum\limits_{i=1}^{n} \alpha_i \sigma_i(h) = 0$ для любого $h \in H$, а $g \in H$ таково, что $\sigma_1(g) \neq \sigma_2(g)$. Домножая на $\sigma_1(g)$ или подставляя gh вместо h получаем:

$$\sum_{i=1}^{n} \alpha_i \sigma_1(g) \sigma_i(h) = 0$$

$$\sum_{i=1}^{n} \alpha_i \sigma_i(g) \sigma_i(h) = 0$$

Вычитая второе равенство из первого, получаем $\sum\limits_{i=2}^{n} \alpha_i \big(\sigma_1(g) - \sigma_i(g)\big)\sigma_i(h) = 0$. По индукционному предположению $\alpha_i\big(\sigma_1(g) - \sigma_i(g)\big) = 0$ при всех $i \geqslant 2$. Так как $\sigma_1(g) \neq \sigma_2(g)$, то $\alpha_2 = 0$. Подставляя это в исходное равенство получим $\sum\limits_{i \neq 2} \alpha_i \sigma_i = 0$, откуда снова по предположению индукции получим $\alpha_i = 0$ при всех i. □

Расширение Галуа E/F называется разрешимым, абелевым, циклическим и т. п., если этим свойством обладает его группа Галуа.

Предположим, что поле F содержит примитивный корень n-й степени из 1, который мы обозначим через ζ (отсюда следует, что $p = \mathrm{char}\, F$ не делит n). Пусть $E = F[\alpha]$, где α – корень многочлена $t^n - \beta$, а $\beta \in F$. Так как все корни этого многочлена имеют вид $\alpha\zeta^i$, то E – поле разложения сепарабельного многочлена $t^n - \beta$, так что E/F – расширение Галуа.

Предложение 8.2. *Пусть $\zeta \in F$ – примитивный корень n-й степени из 1. Расширение E/F является расширением Галуа с циклической группой Галуа порядка n тогда и только тогда, когда $E = F[\alpha]$, где $\alpha^n \in F$, и $\alpha^k \notin F$ при $k < n$.*

Доказательство. ($\Leftarrow$). Мы уже доказали, что E/F расширение Галуа. Положим $G = \mathrm{Gal}(E/F)$. Зададим отображение $\pi : G \to E^*$, $\sigma \mapsto \frac{\sigma(\alpha)}{\alpha}$. Так как $\sigma(\alpha)^n = \sigma(\alpha^n) = \alpha^n$,

то $\sigma(\alpha) = \alpha\zeta^i$, т. е. образ π лежит в циклической группе μ_n корней n-й степени из 1. Далее, $\pi(\tau\sigma) = \frac{\tau\sigma(\alpha)}{\sigma(\alpha)} \cdot \frac{\sigma(\alpha)}{\alpha} = \frac{\tau(\zeta^i\alpha)}{\zeta^i\alpha}\pi(\sigma) = \pi(\tau)\pi(\sigma)$, т. е. π – гомоморфизм. Если $\operatorname{Im}\pi \neq \mu_n$, то он является циклической группой меньшего порядка. Следовательно, $\sigma(\alpha^k) = \sigma(\alpha)^k = \alpha^k$ для любого $\sigma \in G$ и некоторого $k < n$. Но это значит, что $\alpha^k \in F$, что противоречит условию.

Если же $\frac{\sigma(\alpha)}{\alpha} = \frac{\tau(\alpha)}{\alpha}$ для некоторых $\sigma, \tau \in G$, то $\sigma = \tau$, так как α порождает E над F. Таким образом, π – изоморфизм.

($\Rightarrow$). Пусть $G = \{1, \sigma, \sigma^2, \dots, \sigma^{n-1}\}$.

По теореме Дирихле 8.1 $\sum\limits_{i=0}^{n-1} \zeta^i\sigma^i(\beta) = \alpha \neq 0$ для некоторого $\beta \in E$. Тогда

$$\sigma^j(\alpha) = \sum_{i=0}^{n-1} \zeta^i\sigma^{i+j}(\beta) = \sum_{k=0}^{n-1} \zeta^{k-j}\sigma^k(\beta) = \zeta^{-j}\alpha$$

откуда $\sigma^j(\alpha^n) = \sigma^j(\alpha)^n = \alpha^n$. Таким образом, элемент α^n неподвижен под действием группы Галуа и, следовательно, лежит в F.

Предположим, что $\sum\limits_{i=0}^{n-1} \alpha^i x_i = 0$ для некоторых $x_i \in F$. Применяя к этому равенству σ^k получаем $\sum\limits_{i=0}^{n-1} \zeta^{-ik}\alpha^i x_i = 0$. Так как определитель Вандермонда $\det(\zeta^{-ik})_{i,k=0}^{n-1} \neq 0$, то $\alpha^i x_i = 0$ при всех i. Таким образом, $1, \alpha, \dots, \alpha^{n-1}$ линейно независимы, а так как $\dim_F E = n$, то они порождают E над F. Окончательно, $E = F[\alpha]$, что и требовалось. □

Предложение 8.3. *Пусть $n \in \mathbb{N}$, а F – поле, характеристика которого не делит n. Обозначим через E поле разложения многочлена $t^n - 1$ над F. Тогда*

(1) *$E = F[\zeta]$, где ζ – первообразный корень из 1 степени n;*
(2) *Группа Галуа E/F вкладывается в $(\mathbb{Z}/n\mathbb{Z})^*$, в частности, E/F абелево расширение.*

Доказательство. Множество корней многочлена $t^n - 1$ образуют конечную подгруппу в E^*. Любая конечная подгруппа мультипликативной группы поля циклическая. Если ζ – образующая этой группы, то ζ – первообразный корень из 1 (элемент порядка n в E^*), и все корни многочлена $t^n - 1$ имеют вид ζ^i, $i = 0, \dots, n-1$. Таким образом, $F[\zeta]$ содержит все корни многочлена $t^n - 1$ и порождено ими, следовательно, является его полем разложения.

Из условия на характеристику F следует, что $(t^n - 1)' = nt^{n-1}$ взаимно прост с $t^n - 1$ и, следовательно, сепарабелен. Значит E/F – расширение Галуа. Так как ζ порождает E над F, то любой элемент $\sigma \in G := \operatorname{Gal}(E/F)$ однозначно определяется своим действием на ζ. Очевидно, что $\sigma(\zeta)$ является корнем многочлена $t^n - 1$, т. е. $\sigma(\zeta) = \zeta^i$, где $i \in \mathbb{Z}/n\mathbb{Z}$. Обозначим этот i через $\pi(\sigma)$. Тогда π является отображением $G \to \mathbb{Z}/n\mathbb{Z}$. Нетрудно видеть, что $\pi(\sigma\tau) = \pi(\sigma)\pi(\tau)$, т. е. π – гомоморфизм моноидов. Под действием гомоморфизма обратимые элементы должны переходить в обратимые, поэтому, сужая область значений, можно считать, что $\pi : G \to (\mathbb{Z}/n\mathbb{Z})^*$ – гомоморфизм групп. Инъективность π следует из того, что σ однозначно определяется числом $\sigma(\zeta)$. □

Пусть $f \in F[t]$. Группой Галуа сепарабельного многочлена f над полем F называется группа Галуа поля разложения F_f над F. Если F' – расширение поля F, то $f \in F'[t]$ и можно рассмотреть группу Галуа f над F'.

Лемма 8.4. *Группа Галуа f над F' вкладывается в группу Галуа f над F.*

Доказательство. По определению поля разложения многочлен f раскладывается в F'_f на линейные множители, и можно считать, что $F_f \subseteq F'_f$, причем все корни f лежат в F_f. По лемме 2.3 F'_f порождено корнями f, поэтому $F'_f = F' \cdot F_f$. По предложению 7.2 $\operatorname{Gal}(F'_f/F') \cong \operatorname{Gal}(F_f/F_f \cap F')$, а последняя группа является подгруппой в $\operatorname{Gal}(F_f/F)$ (это и есть соответствие Галуа 7.1). □

Уравнение $f(t)=0$ называется разрешимым в радикалах над F, если все корни этого уравнения выражаются через элементы F при помощи арифметических операций и извлечения корня. Другими словами, $f(t)=0$ разрешимо в радикалах, если существует последовательность полей $F=F_0\subseteq F_1\subseteq\dots\subseteq F_n$ такая, что F_n содержит все корни многочлена f, а $F_{i+1}\cong F_i[\alpha_i]$ для всех i и некоторых $\alpha_i\in F_{i+1}$ таких, что какая-то степень α_i лежит в F_i. Если F не указано, то качестве него берется наименьшее поле, содержащее все коэффициенты многочлена f.

Теорема 8.5. *Пусть F – поле характеристики 0, а $f\in F[t]$. Уравнение $f(t)=0$ разрешимо в радикалах над F тогда и только тогда, когда группа Галуа многочлена f над F разрешима.*

Доказательство. ($\Leftarrow$). Добавим к F первообразный корень ζ из 1 достаточно большой степени m (подойдет $m=(\deg f)!$). Пусть $F'=F[\zeta]$. Так как характеристика равна 0, то все расширения сепарабельны. Группа Галуа G многочлена f над F' является подгруппой разрешимой группы, поэтому сама разрешима. Композиционный ряд группы G – это субнормальный ряд $G=G_0\supset G_1\supset\dots\supset G_m=\{1\}$ с простыми абелевыми факторами (т. е. циклическими факторами простого порядка). Положим $F_i=(F'_f)^{G_i}$. Тогда F_i/F_{i-1} циклические расширения и $F\subseteq F'=F_0\subset F_1\subset\dots\subset F_m=F'_f$. По предложению 8.2 $F_i=F_{i-1}[\alpha_i]$, где некоторая степень α_i лежит в F_{i-1}, т. е. уравнение $f(t)=0$ разрешимо в радикалах.

($\Rightarrow$). По условию существует башня полей $F\subset F_1\subset\dots\subset F_m$, где f раскладывается на линейные множители над F_m, а $F_i=F_{i-1}[\alpha_i]$, где $\alpha_i^{k_i}\in F_{i-1}$. Пусть F_{m+1} – поле разложения многочлена t^k-1 над F_m. Заметим, что $F_{m+1}=F[\zeta,\alpha_1,\dots,\alpha_m]$, где ζ – первообразный корень из 1 степени k. Пусть E – нормальное расширение F, содержащее F_{m+1}, например поле разложения минимального многочлена, аннулирующего все элементы $\zeta,\alpha_1,\dots,\alpha_m$. Тогда поле $\tilde{E}$, порожденное над F множеством $S=\{\zeta,\,\sigma(\alpha_i)\mid\sigma\in\mathrm{Gal}(E/F),\,1\leqslant i\leqslant m\}$ также является нормальным расширением F, так как оно инвариантно относительно $\mathrm{Gal}(E/F)$. Занумеровав элементы s_i множества S так, чтобы $s_0=\zeta$, а $\sigma(\alpha_i)$ предшествовало бы $\sigma(\alpha_{i+1})$, получим башню полей

$$F\subset F[\zeta]\subset F[\zeta,s_1]\subset\dots\subset F[\zeta,s_1,\dots,s_j]\subset\dots\subset\tilde{E}.$$

По предложению 8.3 $F[\zeta]/F$ – абелево, а по предложению 8.2 все остальные этажи башни – циклические расширения при условии, что k делится на степени всех этих расширений. Следовательно, этой башне полей соответствует субнормальный ряд группы $\mathrm{Gal}(\tilde{E}/F)$ с абелевыми факторами, т. е. $\mathrm{Gal}(\tilde{E}/F)$ разрешима.

Поле разложения F_f является нормальным расширением F и вкладывается в $\tilde{E}$. Поэтому $\mathrm{Gal}(F_f/F)$ является факторгруппой разрешимой группы $\mathrm{Gal}(\tilde{E}/F)$, а значит и сама разрешима. □

9. Норма и след

Пусть E – конечномерная F-алгебра. Умножение на элемент $\alpha\in E$ является линейным оператором на векторном пространстве E над F. Легко проверить, что заданное таким образом отображение $E\to\mathrm{End}_F E$ является мономорфизмом алгебр. В частности, минимальный многочлен α совпадает с минимальным многочленом умножения на α. Характеристический многочлен, определитель и след оператора умножения на α называются характеристическим многочленом, нормой и следом элемента α. В случае, когда E/F – конечное расширение полей, они обозначаются $\chi_{\alpha,E/F}$, $\mathrm{N}_{E/F}(\alpha)$ и $\mathrm{Tr}_{E/F}(\alpha)$ соответственно. Ясно, что $\mathrm{N}_{E/F}$ является гомоморфизмом мультипликативных групп $E^*\to F^*$, а $\mathrm{Tr}_{E/F}$ – гомоморфизмом аддитивных групп $E\to F$.

Как всегда характеристический и минимальный многочлены имеют один и то же набор корней. В нашем случае выполнено более сильное соотношение между ними.

Предложение 9.1. *Пусть* $f(t) = t^n + \varepsilon_{n-1}t^{n-1} + \dots + \varepsilon_0 \in F[t]$ *– минимальный многочлен элемента* $\alpha \in E$, *а* $m = [E : F[\alpha]]$. *Тогда* $\chi_{\alpha,E/F} = (-1)^{mn}f^m$, $\mathrm{N}_{E/F}(\alpha) = (-1)^{mn}\varepsilon_0^m$, *и* $\mathrm{Tr}_{E/F}(\alpha) = -m\varepsilon_{n-1}$.

Доказательство. Набор $1, \alpha, \dots, \alpha^{n-1}$ является базисом $F[\alpha]$ над F. Если $\gamma_1, \dots, \gamma_m$ – базис E над $F[\alpha]$, то $\gamma_1, \alpha\gamma_1, \dots, \alpha^{n-1}\gamma_1, \dots, \gamma_m, \alpha\gamma_m, \dots, \alpha^{n-1}\gamma_m$ является базисом E над F. Так как $\alpha(\alpha_i\gamma_j) = \alpha^{i+1}\gamma_j$ при $i < n-1$, а $\alpha(\alpha^{n-1}\gamma_j) = \sum_{i=0}^{n-1} -\varepsilon_i(\alpha^i\gamma_j)$, то матрица оператора умножения на α в выбранном базисе имеет клеточно диагональный вид с фробениусовой клеткой по диагонали. Утверждение следует теперь из вычисления характеристического многочлена, определителя и следа фробениусовой клетки, соответствующей многочлену f (характеристический многочлен с точностью до знака равен минимальному хотя бы потому, что их степени равны). □

Следствие 9.2. *Пусть* K *– поле, содержащее* E, *в котором* f *раскладывается на линейные множители, а* $\lambda_1, \dots, \lambda_n \in K$ *корни* f *с учетом кратности. Тогда* $\mathrm{N}_{E/F}(\alpha) = (\prod \lambda_i)^m$, *а* $\mathrm{Tr}_{E/F}(\alpha) = m\sum \lambda_i$.

Следствие 9.3. *Предположим, что* E/F *сепарабельно,* K *– нормальное расширение* F, *содержащее* E, *в котором* f *раскладывается на линейные множители, а* $\sigma_1, \dots, \sigma_k$ *– все различные вложения* E *в* K. *Тогда* $N_{E/F}(\alpha) = \prod_{i=1}^k \sigma_i(\alpha)$, *а* $\mathrm{Tr}_{E/F}(\alpha) = \sum_{i=1}^k \sigma_i(\alpha)$.

Доказательство. При $E = F[\alpha] \cong F[t]/(f)$ вложение E в K однозначно определяется образом α, который может отобразиться в любой корень f. Таким образом, $\sigma_1(\alpha), \dots, \sigma_k(\alpha)$ – это и есть все корни многочлена f в $\overline{E}$, и утверждение вытекает из предыдущего следствия.

В общем случае, пусть $E = F[\alpha, \alpha_1, \dots, \alpha_k]$, а $g \in F[t]$ – минимальный многочлен, аннулирующий все α_i. Так как E/F сепарабельно, то g – сепарабельный многочлен. Так как K/F нормально и K содержит корни всех неприводимых сомножителей многочлена g, то g раскладывается в K на линейные множители. По теореме 6.1, количество вложений E в K, продолжающих данное вложение $F[\alpha]$ в K, равно $\big[E : F[\alpha]\big]$, и утверждение следует из предложения 9.1. □